



3D-SHI index for quantifying 3D spatial heterogeneity in big Earth data

Peijie Sun, Hengcai Zhang, Peixiao Wang & Jianing Yu

To cite this article: Peijie Sun, Hengcai Zhang, Peixiao Wang & Jianing Yu (2025) 3D-SHI index for quantifying 3D spatial heterogeneity in big Earth data, International Journal of Digital Earth, 18:1, 2512058, DOI: [10.1080/17538947.2025.2512058](https://doi.org/10.1080/17538947.2025.2512058)

To link to this article: <https://doi.org/10.1080/17538947.2025.2512058>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 02 Jun 2025.



Submit your article to this journal [↗](#)



Article views: 14



View related articles [↗](#)



View Crossmark data [↗](#)



3D-SHI index for quantifying 3D spatial heterogeneity in big Earth data

Peijie Sun^{a,b}, Hengcai Zhang^{a,b}, Peixiao Wang ^{a,b} and Jianing Yu^{a,b}

^aState Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, People's Republic of China;

^bCollege of Resources and Environment, University of Chinese Academy of Sciences, Beijing, People's Republic of China

ABSTRACT

Investigating spatial heterogeneity is a crucial aspect of spatial analysis. Existing studies have primarily focused on measuring 2D geographic distribution, which cannot be directly applied to extensive 3D big earth datasets. To address this gap, we propose a novel 3D spatial heterogeneity index (3D-SHI) to explore 3D spatial heterogeneity in big earth data. First, we introduce some new concepts including the Earth cube model and connected components to effectively capture 3D spatial heterogeneity. Second, we derive the formulation for 3D-SHI and propose a connected component model (CCM) to facilitate near real-time computation of the index. To elucidate the geographic significance of this index, we analyzed its characterization in comparison with three traditional methods. In our experiments, we simulated 3D datasets to demonstrate the efficiency and capacity of the proposed 3D-SHI. Additionally, we used a real-world sea temperature dataset from the North Atlantic were conducted to quantify high dimensional spatial heterogeneity. The experimental results illustrated that 3D-SHI can effectively capture changes in spatial heterogeneity and reveals patterns that exist across both temporal and spatial dimensions.

ARTICLE HISTORY

Received 20 November 2024

Accepted 17 May 2025

KEYWORDS

Spatial heterogeneity; three dimensional; spatial statistics; voxel

1. Introduction

Spatial heterogeneity, as an important spatial characteristic in the field of geographical information system, aimed to characterize the uneven geographic distribution of elements, and can be divided into local heterogeneity and stratified heterogeneity (Goodchild 2004; Tobler 2004). Local heterogeneity refers to the variation in attribute values of a point compared to its surrounding areas (Fotheringham and Sachdeva 2022), whereas stratified heterogeneity signifies the differences between various regions (Wang and Xu

CONTACT Hengcai Zhang zhanghc@lreis.ac.cn State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, People's Republic of China College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, 100049, People's Republic of China

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

2017). Spatial heterogeneity has been extensively employed in spatial data analysis, such as geographic zoning, which delineates dissimilar units and creates heterogeneous boundaries (Figure 1(a)) (Zhang et al. 2024). Climate zones, land use maps, and geomorphology maps are all typical manifestations of spatial heterogeneity (Huang et al. 2023; Potapov et al. 2020; Taghizadeh-Mehrjardi et al. 2014; Zhu et al. 2022). However, existing studies on spatial heterogeneity primarily focus on two-dimensional geographic analyses, including land use change analysis (Potapov et al. 2020) and the identification of El Niño – Southern Oscillation (ENSO) through sea surface temperature (SST) anomaly patterns (Cai et al. 2018).

Recent advances in Earth observation technology have significantly expanded sensing capabilities across air, space, land, and ocean, enabling acquisition of vast amounts of true three-dimensional (3D) data (Wang et al. 2022). The availability of extensive Earth data underscores the importance of spatial heterogeneity within real-world 3D Contexts. Quantifying heterogeneity in true 3D data is essential for accurately representing geographic phenomena and processes in geographic modeling (Shen et al. 2024; Wu et al. 2017). Research on the heterogeneity of three-dimensional Earth features has been extensively conducted. In the marine domain, efforts have concentrated on quantifying the spatial heterogeneity of physical variables, such as temperature and salinity, and establishing their relationships with oceanic phenomena (Mahadevan 2003; Qin, Zhang, and Yin 2015). In atmospheric science, the simulation of heterogeneity in pollutant concentrations and its correlation with human activities has emerged as a prominent research focus (Fan, Xin, and Xiangzheng 2021; Gao et al. 2018). In soil science, detailed investigations have been conducted on the relationship between the heterogeneity of soil salinity and moisture content, along with the mechanisms of water transmission and their implications for landscape ecology (Yang and Yao 2007). To characterize the heterogeneity of Earth's features in three-dimensional space, a common and intuitive approach involves applying existing spatial heterogeneity results along specific dimensions in a sliced manner. Each slice is analyzed as a two-dimensional plane, facilitating the quantification of spatial heterogeneity through various statistical metrics. These metrics include spatial autocorrelation indices, such as Moran's I (Moran 1950; Tian, Zhang, and Yuan 2014) and Geary's C (Geary

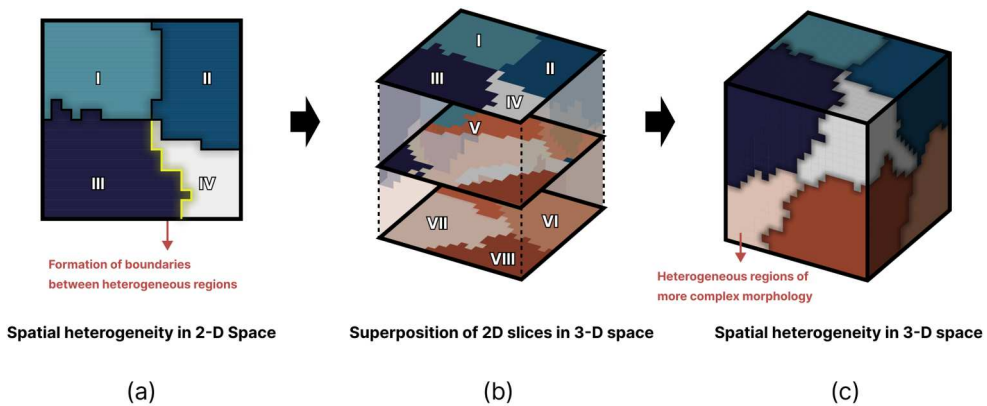


Figure 1. Spatial heterogeneity of geographic elements in different spatial perspectives. (a) in 2D space. (b) in 2.5D space. (c) in 3D space.

1954), which evaluate the degree of spatial correlation. Additionally, measures of spatial variability, including semivariograms and their extensions (Matheron 1963), are employed. Information entropy indices, such as Shannon entropy, offer insights into the randomness and complexity of spatial distributions (Chen and Huang 2018). These metrics are utilized to quantify spatial heterogeneity within each slice. The quantification results for each slice are subsequently analyzed along the chosen dimension to reveal underlying geographic patterns (Qin, Zhang, and Yin 2015; Shen et al. 2024; Sun et al. 2024). The second approach emphasizes inter-slice heterogeneity, building upon the spatial heterogeneity quantification within each slice. This is typically accomplished through statistical methods, such as the geographical detector Q statistic, which quantifies spatial differentiation between slices (Wang and Xu 2017). The results of intra-slice and inter-slice heterogeneity are then integrated to provide a comprehensive representation of three-dimensional spatial heterogeneity (Luo et al. 2023; Sun et al. 2024; Zhao et al. 2020).

However, the distribution of true 3D spatial elements does not align neatly with stratified divisions, as shown in Figure 1(b). True 3D elements typically have more complex distributions; elements with spatial autocorrelation often expand in various directions, forming irregular shapes in space (Figure 1(c)) (Shen et al. 2024; Ying et al. 2021). Thus, a 2D slice perspective cannot fully capture the changes in spatial heterogeneity arising from irregular shapes formed by the expansion of elements in all directions within a 3D space. Hence, how to measure spatial heterogeneity on a 3D scale earth system element, such as soil, water, and atmosphere, remains a pressing research challenge (Borrelli and Relyea 2022; Zhao et al. 2019).

Meanwhile, big Earth data are rapidly evolving toward higher dimensions, higher resolutions, and massive data volumes. For example, marine environmental elements, supported by projects such as the Argo program (Roemmich et al. 2019) and the Global Drifter Program (He et al. 2024; Swenson and Ams 2000), now generate high-quality, long-term, high-resolution gridded numerical products on a large scale. The representative Institute of Atmospheric Physics version 4 dataset (IAPv4), which provides global monthly gridded temperature and heat content data since 1940, has a horizontal resolution of $1^\circ \times 1^\circ$ and covers depths of 1–6000 meters (119 layers) in the ocean (Cheng et al. 2024). Figure 2 presents a set of data from this dataset, clearly revealing an initial observation of the three-dimensional spatial heterogeneity of sea temperature. For atmospheric elements, using carbon dioxide as an example, advancements in ground-based CO₂ observation stations and passive satellite detection technologies have enhanced temporal and spatial resolution. CO₂ concentration data at a global scale can achieve a spatial resolution of $1^\circ \times 1^\circ$, while areas can reach a spatial resolution of up to 30 meters and a vertical resolution of up to 100 meters (He et al. 2022; Liang et al. 2022). Soil elements have also benefited from advances in data collection technologies, achieving a horizontal resolution of $1^\circ \times 1^\circ$ at the global scale and up to 1 km \times 1 km in local areas, with vertical coverage up to 0–289 meters (Li et al. 2022; Orgiazzi et al. 2018; Pelletier et al. 2016). The true 3D distributions of various Earth system elements present highly complex patterns, making sliced spatial heterogeneity solutions inadequate for meeting the practical needs of true 3D Earth system elements.

The increasing dimensions, resolution, and data volume of Big Earth data facilitate the representation of finer-scale and more accurate geographic patterns in three-dimensional space (Guo et al. 2017). The higher dimensionality of Big Earth data requires the

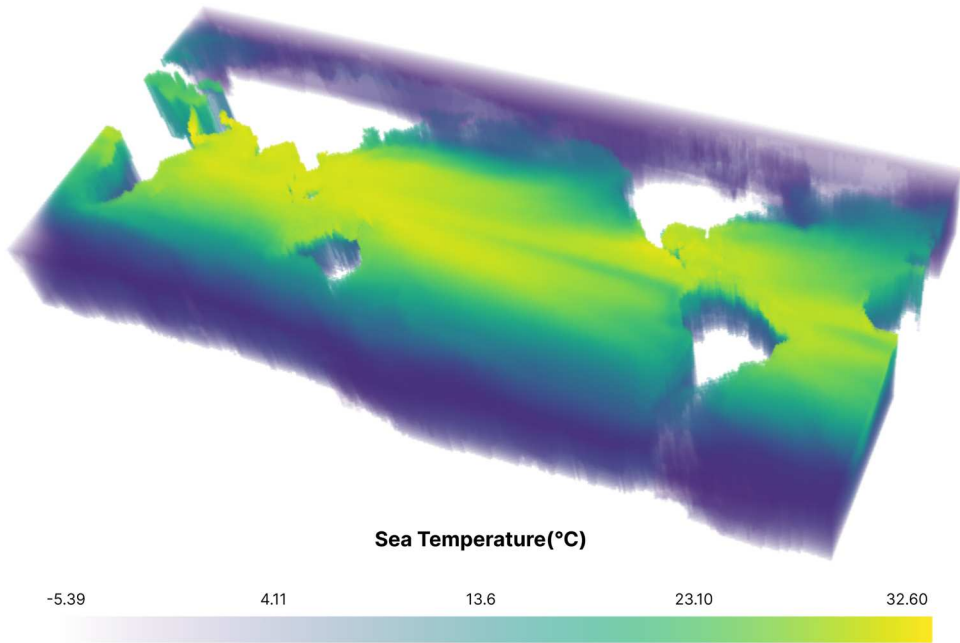


Figure 2. IAPv4 sea temperature gridded dataset for June 2020.

representation of spatial heterogeneity from a genuinely three-dimensional perspective to effectively illustrate the multi-dimensional spatial patterns of geographical features. Enhanced resolution allows the dataset to capture more detailed and realistic complex forms, thereby increasing the demands on methodologies to accurately represent this complexity. Furthermore, the substantial volume of data leads to significant computational loads, imposing elevated requirements on computational resources and performance. However, existing methods necessitate multiple linear computations for each sample point, resulting in considerable time and spatial complexity when calculating heterogeneity within massive datasets. This complicates the accurate representation of heterogeneity in true three-dimensional space (Fortin et al. 2012; Gaudart et al. 2007; Li, Wang, and Wang 1998). Consequently, Big Earth data imposes heightened demands on methodologies aimed at quantifying three-dimensional spatial heterogeneity:

- (1) Higher-dimensional heterogeneity representation requirements: Representing heterogeneity in 3D space requires fully accounting for all dimensions within 3D space.
- (2) Increased demands for computational efficiency: The sheer volume of data places more stringent requirements on the efficiency of representation algorithms.

In this study, we propose a method for characterizing the 3D spatial heterogeneity of big Earth data, called 3D spatial heterogeneity index (3D-SHI). This method quantifies spatial heterogeneity by recording information on voxel-connected components and calculating their distribution scale. The 3D-SHI presents three key advantages. First, it converts regions with complex morphologies into connected components within a voxel data model and preserves connectivity information via a connected component model

(CCM). This model stores connectivity information as a two-dimensional matrix, with rows representing voxel values and columns representing the scale of connected components. This structure makes 3D-SHI both universally applicable and capable of conducting near-real-time analysis on large-scale geospatial data. Second, 3D-SHI quantifies the distribution scale of connected components to accurately measure the spatial heterogeneity of true 3D Earth system data. Unlike methods that address heterogeneity on a sliced basis, it evaluates spatial heterogeneity across all dimensions in 3D space, offering a novel perspective for understanding spatial heterogeneity at a 3D scale. Third, 3D-SHI enables the quantification of 3D spatial heterogeneity within a specific range of interest by setting adjustable weights, thereby providing considerable flexibility.

The rest of this paper is organized as follows. Section 2 presents preliminary concepts, including the big Earth data, voxel data model, connected components, and homogeneous connected components. Section 3 introduces the deviation of 3D-SHI and its algorithm. Section 4 provides a detailed analysis of the characterization capabilities of 3D-SHI. Section 5 demonstrates the use of this method both simulated datasets and a sea temperature dataset of the North Atlantic as case studies. Finally, Section 6 summarizes the conclusions and future research directions.

2. Preliminary

In this section, we will explain some terms, including big Earth data, Earth cube model, connected component and homogeneous connected component. These concepts will serve as the foundation for the subsequent derivation of 3D-SHI.

2.1. Definition 1 big earth data

A scientific big data set with spatial features that records information about Earth elements (Figure 3). It is characterized by its massive volume, heterogeneity, multiple time phases, and multi-scale properties, exhibiting strong spatiotemporal correlations (Guo et al. 2014; Guo, Wang, and Liang 2016). Earth data can encompass a wide range of features with spatial information, including ocean temperature, soil salinity, carbon dioxide concentration, and other relevant element data, all of which constitute big Earth data (Rew and Davis 1990; Wang et al. 2013). For managing such Earth data, netCDF has emerged as a widely adopted file format standard due to its efficient data storage and management capabilities, support for multidimensional data, metadata management, and ability to handle large volumes of array-oriented data. Notably, ERA5, the most representative global climate and weather reanalysis dataset, utilizes netCDF as its file format standard (Hersbach et al. 2020).

2.2. Definition 2 earth cube model

Inspired by netCDF, we propose a data model for big Earth data called Earth Cube Model. It consists of a set of Earth element voxels that store the information of elements within a 3D space of the Earth. Figure 4(b) shows a typical earth cube model $f(x): V \rightarrow T$, where the model $V \subset Z^3$ is composed of a set of voxels $x \in V$, $x = \{x_1, \dots, x_N\}$, with N being the number of voxels.

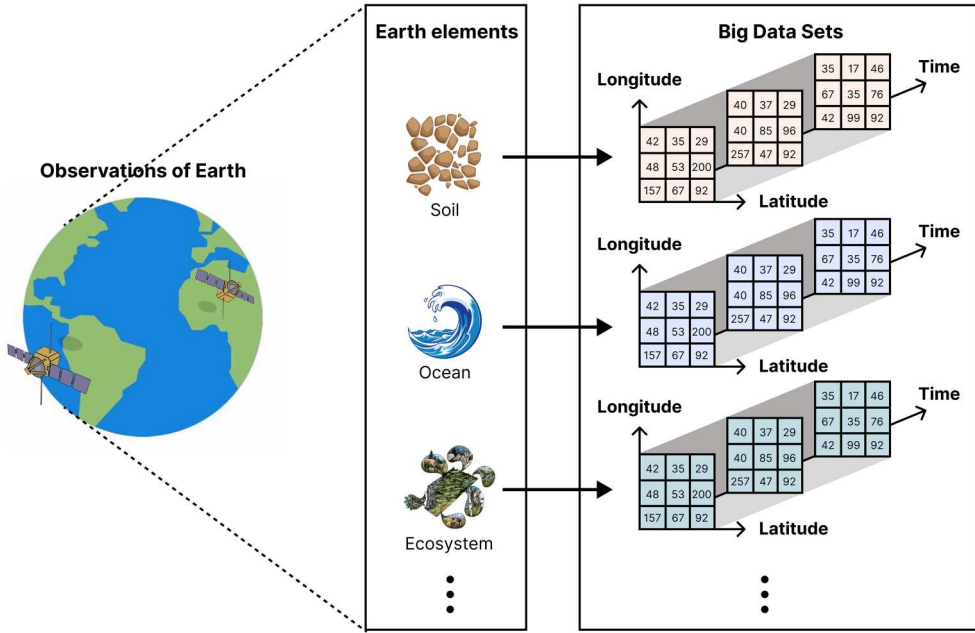


Figure 3. Big Earth data. A scientific big data set with spatial features that records information about Earth elements.

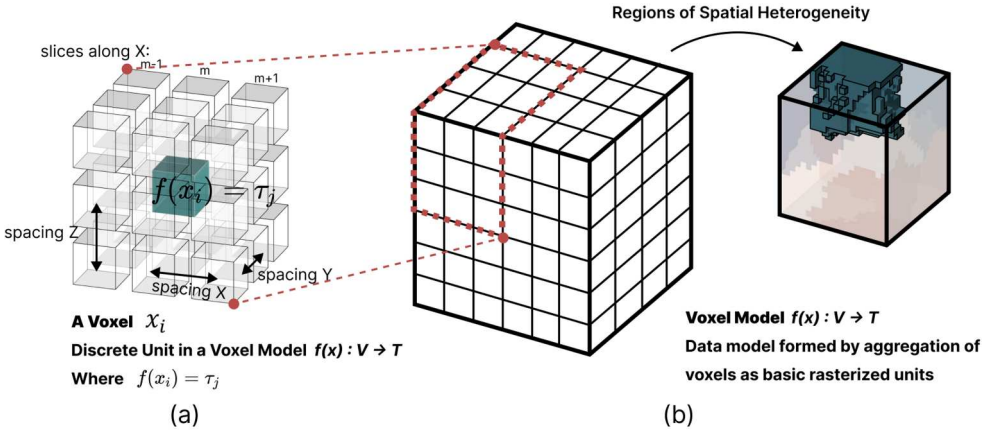


Figure 4. Schematic of Earth cube model and basic voxel unit. (a) Basic voxel unit x_i . (b) Typical Earth cube model $f(x):V \rightarrow T$.

The voxel is the basic unit of this data model. A voxel is a cubic geometric body representing a collection of element features within an actual spatial area, with side lengths corresponding to the dataset's minimum resolution in each direction. In practical scenarios, the center point or corner (i.e. vertex) of the cube can be used to represent the entire voxel (Xu, Tong, and Stilla 2021). Figure 4(a) shows a basic voxel. If two voxels share a face, vertex, or edge, these two voxels are considered adjacent. For every x_k ,

there is a discrete voxel value τ_j , with a range of a closed set $T = \{\tau_1, \tau_2, \dots, \tau_M\}$, and $\Delta\tau = \tau_{i+1} - \tau_i$.

The Earth cube model effectively translates three-dimensional scale information into discrete voxel units, making it an ideal data model for representing elements of the three-dimensional Earth system. By storing feature information – such as gas concentration, soil salinity, ocean temperature, and other parameters – within these voxels, the model enables the representation of various Earth system components, including water, soil, air, and life, thereby facilitating the simulation of diverse geographical phenomena (Meyer et al. 2018; Shen et al. 2006; Shen, Takara, and Liu 2016; Van Lancker et al. 2017). Moreover, utilizing voxels as discrete units allows for an accurate depiction of the morphology and boundaries of elements exhibiting heterogeneous phenomena in three-dimensional space, thus providing a foundational model for investigating heterogeneity within geographical contexts (Xu et al. 2018).

2.3. Definition 3 connected component

As for an Earth cube model $f(x):V \rightarrow T$, if there exists a subset $\Omega \subseteq V$, and for any two voxels in Ω , there exists a sequence of voxels x_i, x_j, \dots, x_k , with all voxels in the sequence having the same value $\tau \in T$. If there is no such that $x \notin \Omega$ and adding x to Ω would still maintain connectivity, then Ω is a connected component in V .

It should be noted that the connected component can vary depending on the voxel adjacency method. The most commonly used is the 26-adjacency, which means that if a voxel is adjacent to any of its 26 neighboring voxels with the same voxel value, then the two voxels are considered connected. Figure 5 shows a typical example of a connected component formed by the 26-adjacency method, where all voxels in Ω have the same value $\tau = 4$, and any two voxels x_i, x_j in Ω can form a continuous sequence of voxels with $\tau = 4$.

2.4. Definition 4 homogeneous connected component

A connected component within the same Earth cube model $f(x):V \rightarrow T$ that has the same voxel value τ and the same scale θ (i.e. the number of voxels is the same). For the voxel value τ and scale θ , the number of homogeneous connected components can be represented as $C_f(\tau, \theta)$.

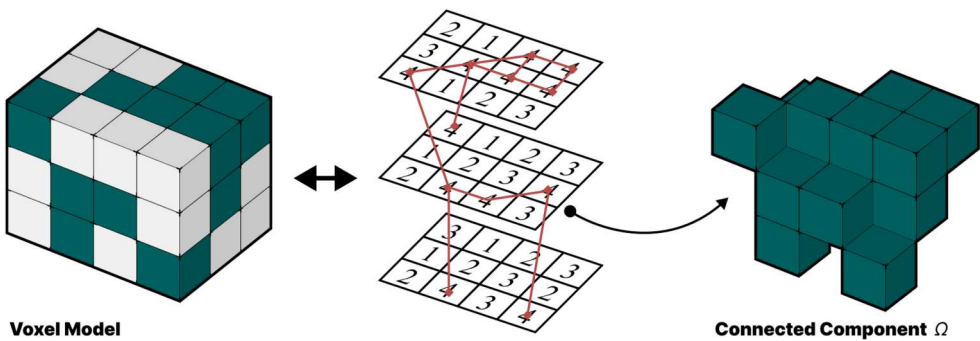


Figure 5. Typical example of a connected component with 26-adjacency. All voxels in Ω have the same value $\tau = 4$.

3. 3D-SHI (3D-spatial heterogeneity index)

In this section, we start by deriving and defining the 3D-SHI from the perspective of connected components. Secondly, we propose the CCM tool, which can store information about the scale of connected components, and demonstrate the algorithm for 3D-SHI using CCM.

3.1. Definition of 3D-SHI

We first consider the size of each connected component (CCS) and assign a scale weight to each connected component (i.e. multiply by itself) to obtain each term. This gives larger-scale connected components a greater weight. Then, we take the square root of the sum of all terms and divide it by the sum of all CCS to obtain the scale of spatial connected component (SSCC).

$$SSCC = \frac{\sqrt{\sum CCS^2}}{\sum CCS} = \frac{\sqrt{\sum_{i=t_1}^{t_m} \sum_{j=1}^{N_s} C_f(i, j) j^2 w_i}}{\sum_{i=t_1}^{t_m} \sum_{j=1}^{N_s} C_f(i, j) j w_i} \quad (1)$$

The sum of all CCS can also be represented by N :

$$N = \sum_{i=t_1}^{t_m} \sum_{j=1}^{N_s} C_f(i, j) j w_i \quad (2)$$

where N is the total number of voxels of interest in the dataset (which is the total number of samples when the weight w is not considered); i is the value of the voxels included in the connected component Ω_k in $f(x): V \rightarrow T$, with $i \in T = \{t_1, t_2, \dots, t_M\}$; j is the scale of Ω_k ; N_s is the maximum scale of the connected component corresponding to the voxel value; $C_f(i, j)$ is the number of homogeneous connected components with voxel value i and volume j ; w_i represents the weight assigned to the voxel value i , typically, the weight for the voxel value of interest k is set to $w_k = 1$, and the weights for other voxel values are set to 0.

The index quantifies the scale at which voxel units form connected components in space. A larger index value indicates that the connected components in space tend to be distributed in the form of large-scale connected components. Note that SSCC is related to the total number of voxels N . To serve as a universally applicable method for characterizing the heterogeneity of three-dimensional space, it is necessary to eliminate the impact of different N on the index. Therefore, we linearly map the value of SSCC to the interval $[0, 1]$ to obtain the three-dimensional spatial heterogeneity index 3D-SHI:

$$3D-SHI = 1 - \frac{SSCC - \frac{1}{N}}{1 - \frac{1}{N}} = \frac{\sum_{i=t_1}^{t_m} \sum_{j=1}^{N_s} C_f(i, j) j w_i - \sqrt{\sum_{i=t_1}^{t_m} \sum_{j=1}^{N_s} C_f(i, j) j^2 w_i}}{\sum_{i=t_1}^{t_m} \sum_{j=1}^{N_s} C_f(i, j) j w_i - 1} \quad (3)$$

where $i, j, C_f(i, j), w_i$ in Equation (3) have the same meanings as in Equation (1).

The 3D-SHI has a value range of $[0, 1]$. When 3D-SHI takes the minimum value of 0, SSCC reaches its maximum value of 1, corresponding to the scenario with the highest

spatial heterogeneity: in the Earth cube model $f(x)$, any voxel x_i has no other voxel x_j with which it can form a connected component, meaning all connected component scales are 1. Conversely, if any two voxels x_i and x_j in $f(x)$ can form a connected component, the scenario exhibits minimal spatial heterogeneity; in this case, SSCC reaches its minimum value of $1/N$, and 3D-SHI reaches its maximum value of 1. A smaller 3D-SHI indicates that the space is more likely to be dominated by larger-scale connected components, hence the spatial heterogeneity is lower.

The weight w_i of a voxel value i typically takes values of 0 or 1, serving a filtering role. By default, all w_i are set to 1. A value of 1 indicates that the contribution of the connected components under this voxel value to the heterogeneity is considered, while a value of 0 means that the contribution of all connected components under this voxel value is ignored in the calculation of 3D-SHI. By setting the weight of the voxel value of interest to 1 and the weights of other voxel values to 0, it is possible to filter and compute the spatial heterogeneity for one or more specific voxel values. In this scenario, it should be noted that a 3D-SHI value of 0 may indicate either the absence of voxels of interest within the Earth cube model or that all voxels with the voxel value of interest form a single connected component. These situations need to be analyzed further. A 3D-SHI value approaching 1 similarly suggests that all voxels with the voxel value of interest tend to be distributed as individual discrete voxels within 3D space.

3.2. Algorithm

From Equation (3), the key to calculating the 3D-SHI is to count the number of homogeneous connected components within the Earth cube model. To improve computational efficiency, we have proposed a new algorithm called CCM (Connected Component Model) to store all homogeneous connected component information. The CCM is denoted as $CCM_f(c_n, a_m)$ and provides a statistical representation by calculating the binary conditional probability density function of the voxel values in $f(x): V \rightarrow T$. Specifically, $CCM_f(c_n, a_m) = C_f(c_n, a_m)$. The CCM generated by f has a fixed number of rows equal to M (the number of elements in set T) and a dynamic number of columns equal to N (the maximum scale of the connected components in f).

Figure 6 illustrates the entire process of calculating 3D-SHI for an example Earth cube model $f(x)$ using CCM: As shown in Figure 6(a), if the original voxel data is numerical, a connected component threshold *width* needs to be set based on the actual scenario to form a voxel value mapping $f \rightarrow g$, generating a connected component Earth cube model $g(x)$. Threshold selection plays a pivotal role in voxel value mapping, where higher values emphasize large-scale spatial trends while lower values resolve fine-scale spatial patterns. Our heuristic determination method leverages the characteristic variance reduction trend observed with increasing threshold values. However, after reaching a certain threshold, the rate of variance reduction diminishes significantly. This juncture indicates the optimal threshold. The specific procedure is outlined as follows:

- (1) Define an equally-spaced candidate threshold list $W = [w_1, w_2, \dots, w_n]$, where $w_1 = mn$ and $w_n = mx$ represent the minimum and maximum values of the threshold range with a constant step size s , such that $w_{i+1} = w_i + s$ for $i = 1, 2, \dots, n - 1$;

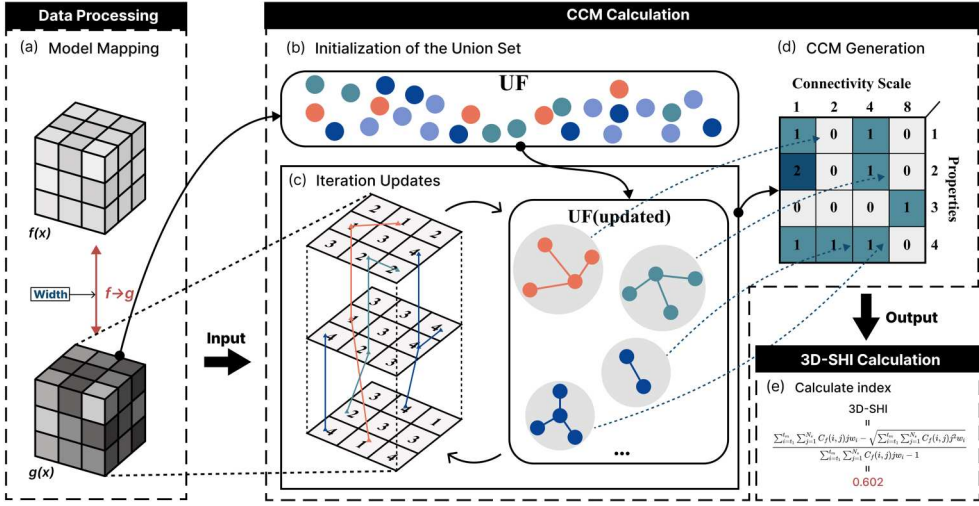


Figure 6. Example of CCM Calculation for 3D-SHI. (a) Model mapping process. (b) Initialization of UF. (c) Recursive updating of the union-find process. (d) Generation of CCM. (e) Calculation of 3D-SHI based on the output CCM.

- (2) Iterate through the list, and for each candidate threshold w_i , perform voxel value mapping on the data;
- (3) Compute the variance v_i of the transformed data and derive the second-order finite difference $\Delta^2 V(i)$ from the variance sequence, as defined in Equation (4).
- (4) Select the threshold corresponding to the maximum $\Delta^2 V$ as the optimal mapping value w_{best} .
- (5) To enable comparable spatial heterogeneity analysis across scenarios, we develop an extremum-based standardization approach: deriving a reference threshold w^* as the mean of maximum w_{max} and minimum w_{min} optimal thresholds from Step (4), which is then uniformly implemented for consistent cross-scenario evaluation.

$$\Delta^2 V(i) = v_{i+1} - 2v_i + v_{i-1}, \quad 1 \leq i \leq n-1 \quad (4)$$

where i indexes the threshold list $W = [w_1, w_2, \dots, w_n]$, v_i is the variance at w_i , and $\Delta^2 V(i)$ (second-order difference at w_i) quantifies the acceleration of change – larger values correspond to sharper transitions.

As shown in Figure 6(b), the $g(x)$ is input into CCM calculation tool, which creates and maintains a union-find (UF) set based on $g(x)$, keeping track of all connected components in the model; Then, as shown in Figure 6(c), the process recursively traverses $V \rightarrow g(x)$, checking if each voxel v has already been assigned to a known connected component. If not, the process starts from this voxel and performs Area expansion to check for connectivity. If connectivity is established, the UF is updated; After the recursion is completed, the matrix CCM shown in Figure 6(d) is generated. Finally, we input the output CCM into Equation (3) to calculate the 3D-SHI value.

The bottleneck of the algorithm arises from the fact that, although the Union-Find (UF) data structure is employed to maintain the connectivity information of all voxels, each find and union operation has a time complexity of $O(n)$ (Galler and Fischer 1964).

Table 1. Algorithm time complexity analysis.

Lines	Complexity	Description
Line 2	$O(n)$	Initialize union-find and generate model process
Lines 3–10	$O(n\alpha(n))$	Path compression and find operations for all voxels; $\alpha(n)$ refers to the inverse Ackermann function, operations that depend on $\alpha(n)$ almost constant time
Lines 11–17	$O(n)$	Generation of CCM based on union-find results
Lines 18–20	$O(n\alpha(n))$	Perform union by rank operations for each union call
Total		$O(n) + O(n\alpha(n)) + O(n) + O(n\alpha(n)) \approx O(n)$

This leads to considerable time consumption, particularly with large datasets. To address this issue, we introduced two key optimizations:

First, we enhanced the Union-Find structure itself through path compression and union by rank (Gao et al. 2008), which reduces the average time complexity of each find and union operation to nearly constant time. Second, to mitigate the problem of redundant voxel lookups, we implemented memoization, caching the lookup results (i.e. the root nodes) for each voxel to avoid repeated lookups. The optimized algorithm is detailed in Algorithm 1.

Additionally, in Table 1, we analyze the time complexity of the algorithm. The results indicate that the optimizations reduce the average time complexity of the Union-Find operations to nearly constant time, thereby bringing the overall time complexity down to $O(n)$ and the space complexity to $O(n^2)$. This enhancement enables the algorithm to manage large-scale datasets and perform near-real-time 3D-SHI computations. Furthermore, the algorithm demonstrates significant potential for parallel computation. The Union-Find structure can utilize local subtree path compression and parallel query operations, necessitating only a global cache to manage and update subtrees through deferred merging and updates.

Algorithm 1: 3D-SHI Calculation through CCM with Optimized Union-Find

Input: Earth cube model $f(x)$, threshold *width*, union-find data structure *UF*.

Output: 3D-SHI value.

```

1:   $g(x) \leftarrow$  Apply threshold width to  $f(x)$ ; // Generate connected components Earth cube model  $g(x)$ 
2:  Initialize UF, Cache; // Initialize union-find and memoization cache
3:  for each voxel  $v$  in  $g(x)$  do
4:      if  $v$  is not assigned to a component then
5:          Start new connected component from  $v$ ; // Perform Area Expansion
6:          Update UF with new connected component;
7:      end if
8:      Perform path compression in UF; // Optimize UF structure
9:       $Cache[v] \leftarrow Find(v)$ ; // Memoize the result of Find
10: end for

11: Initialize CCM with dimensions  $|attributes| \times |connected\_components|$ ;
12: for each voxel  $v$  in  $g(x)$  do
13:      $root \leftarrow Cache[v]$ ; // Get the root from Cache
14:      $voxel\_value \leftarrow f(v)$ ; // Get the value of the voxel
15:     // Increment the CCM matrix entry for the associated attribute and component
16:      $CCM[voxel\_value, root] \leftarrow CCM[voxel\_value, root] + 1$ ; // Add 1 to the existing value
17: end for

18: for each union operation between components do
19:     Union by rank; // Union by rank optimization
20: end for

21:  $3D - SHI \leftarrow$  Compute 3D - SHI using Equation (3) with CCM
22: return 3D - SHI;

```

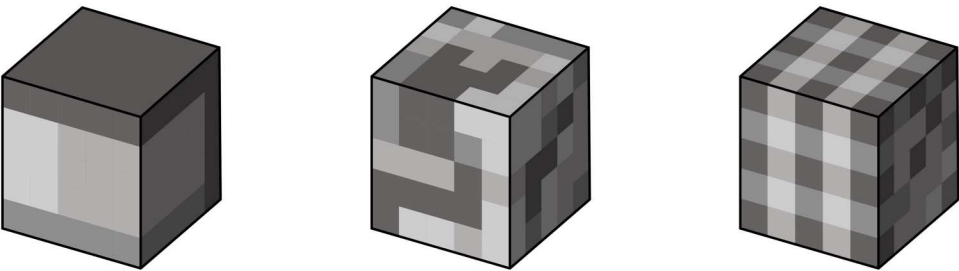








Figure 7. Examples of scenarios with different spatial heterogeneity levels. Noted that the same grayscale areas that are spatially adjacent represent connected components.

4. Characteristics of 3D-SHI

From the perspective of connected components, spatial heterogeneity can be expressed by the distribution characteristics of connected components at different scales. Figure 7 shows three scenarios with connected components of different scales, and thus the degree of spatial heterogeneity in the scenarios also varies significantly. An obvious conclusion is that spaces distributed in the form of large-scale connected components tend to have less heterogeneity. We find that large-scale connected components have a negative contribution to the increase in spatial heterogeneity.

Hence, the 3D-SHI focuses on connected components as the perspective of study. Compared to traditional two-dimensional spatial heterogeneity statistical methods, the 3D-SHI can also reflect differences in results under its different spatial distribution scenarios, making this method an extension of spatial heterogeneity methods. To verify this, we first designed a scenario with different spatial distributions in a two-dimensional plane from the perspective of traditional statistical methods, as shown in Table 2: The six scenarios from left to right represent a change process from completely discrete to completely aggregated spatial distribution, with spatial heterogeneity gradually decreasing. Table 2 shows the calculation results of 3D-SHI and three typical methods (Moran’s I, semivariograms and spatial entropy). We find that (Figure 8):

Table 2. Calculation results of 3D-SHI and three typical spatial heterogeneity statistical methods for scenarios with six different spatial distributions. Noted that the line of the sill ‘–’, represented by a large Nugget value, makes the Sill value of no practical significance.

Scenarios							
		Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
		Discrete					Clustered
Methods		Results					
3D-SHI	1.000	0.886	0.707	0.655	0.500	0.000	
Moran’s I	Index	–1	–0.126	0	0.352	0.924	1
	p-value	0.001	0.001	0.127	0.001	0.001	0.001
Semivariograms	Nugget	2.5	1.47	1.2	0.7	0.2	0
	Sill	–	–	–	4.1	2.0	0
Spatial Entropy		1.38	1.38	1.38	1.38	1.38	0

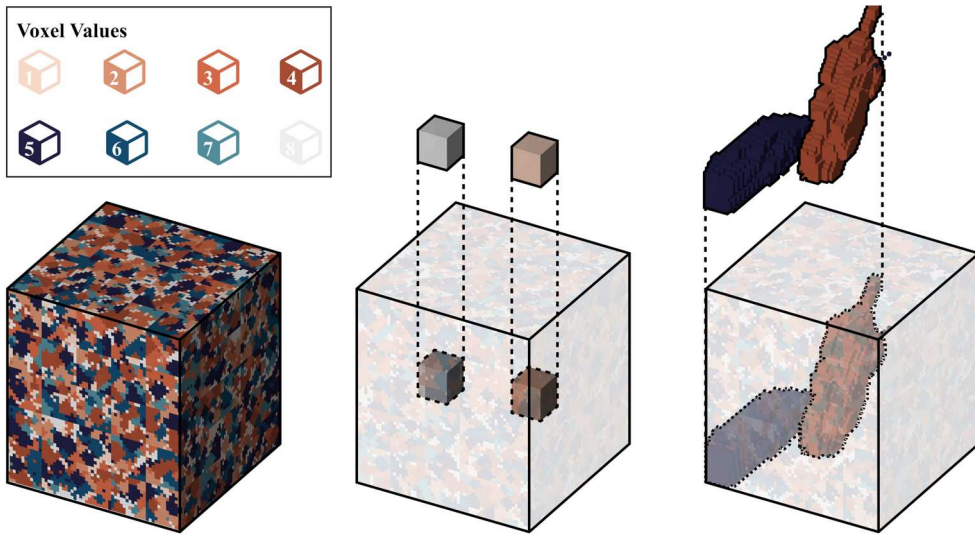


Figure 8. 3D voxel datasets under four different scenarios. (a) Case a, the baseline scenario. (b) Case b. (c) Case c.

3D-SHI can also identify changes in the magnitude of heterogeneity under different spatial distributions. In the extreme scenarios at both ends, the 3D-SHI achieves extreme values opposite to those of Moran's I . The scenario on the far left corresponds to the maximum value situation. Conversely, the minimum value of 0 corresponds to the scenario on the far right. However, the 3D-SHI cannot obtain a determinate value similar to 0 like Moran's I in spatially random distribution, and such distribution is not the focus of our attention in the quantification of spatial heterogeneity issues.

Furthermore, 3D-SHI effectively detects the emergence of local clusters in scenarios characterized by a discrete or scattered overall spatial distribution, yet locally aggregated patterns. In cases of relatively random and discrete spatial distributions, 3D-SHI demonstrates stability in quantifying spatial heterogeneity, even when the semivariogram is unable to determine the Sill value due to pronounced heterogeneity. This capability enables 3D-SHI to identify small-scale connected components. Furthermore, in large-scale datasets, this feature facilitates the quantification of local geographic phenomena across broader areas. Additionally, 3D-SHI incorporates considerations of local spatial autocorrelation, which allows it to monitor changes in spatial heterogeneity, even when voxel values remain unchanged but the spatial distribution varies. Consequently, 3D-SHI is highly generalizable, supporting comparative analyses of heterogeneity across various scales and geographic contexts.

The 3D-SHI assigns greater significance to large-scale connected components, making it particularly effective in characterizing spatial heterogeneity in scenarios where the form and scale of these components fluctuate. These scenarios can be categorized into two types based on the variations in the scale of connected components. The first scenario involves the aggregation and expansion of large-scale connected components, with typical phenomena including the initial aggregation of water bodies that form large-

scale connected components during the development of oceanic water masses (Hu et al. 2017), the construction of large-scale urban green spaces (Willcock et al. 2021), and the emergence of atmospheric phenomena (Resplandy et al. 2024). The second scenario entails the fragmentation of large-scale connected components, with specific phenomena such as the splitting of oceanic water masses (Hu et al. 2017) and (Liao et al. 2013).

5. Case study

5.1. Experimental datasets

5.1.1. Simulation dataset

Considering the emergence of massive amounts of data with vast increases in data volume and spatial resolution (Wang et al. 2022), we design four 3D voxel datasets shown in Figure 9, aimed at simulating the representation capability of the method from a three-dimensional perspective and within large datasets:

Each dataset consists of dimensions measuring (60, 60, 60), resulting in a total of 216000 data samples, with discrete voxel values ranging from 1 to 8. Different colors are used to represent these various voxel values. The distribution of voxels for each of the eight voxel values follows a normal distribution based on the total number of voxels assigned to each value.

- (1) Case a (Figure 9(a)): This represents the baseline scenario, where voxels are randomly assigned to generate connected components. All generated domains are constrained to a maximum volume of 300, making this scenario exhibit the least heterogeneity among the four scenarios examined.

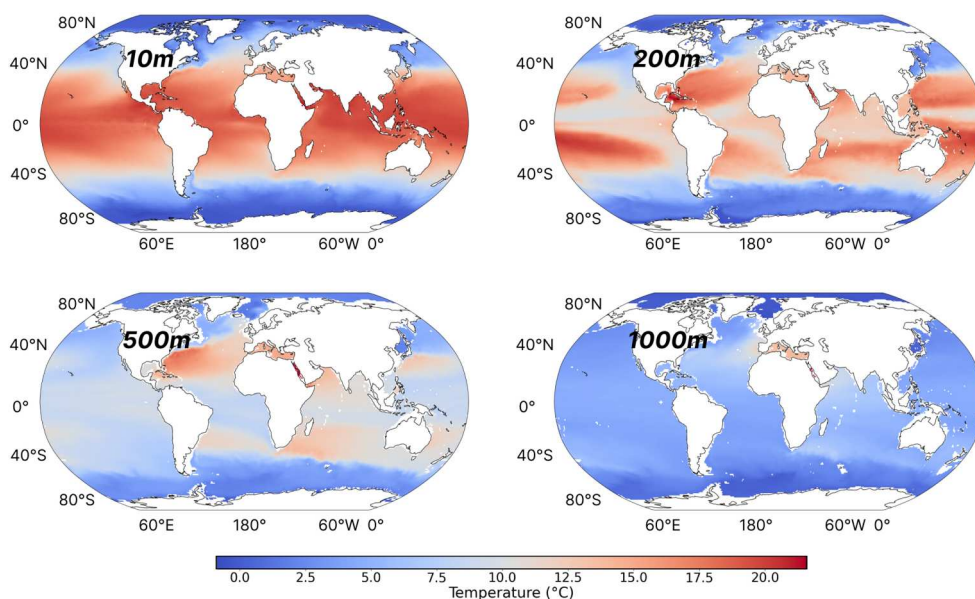


Figure 9. Sea temperature data at different depths for June 2023 in the IAPv4 dataset.

- (2) Case b (Figure 9(b)): Building upon Scenario 1, this case introduces two large connected components with volumes of 1000 and values of 1 and 8. The original position voxels are randomly assigned to regions outside these large connected components, simulating a situation where large-scale connected components appear locally.
- (3) Case c (Figure 9(c)): This scenario also builds on Scenario 1, incorporating two large connected components with volumes of 6000 (value 4) and 5000 (value 5) into the Earth cube model. The original position voxels are again randomly assigned to areas outside the large connected components. Notably, the scale of the two inserted connected components is significantly larger than the total volume of all connected components in the baseline scenario, illustrating a typical case of large-scale connected component aggregation of attribute values within the Earth cube model.

5.1.2. North Atlantic Sea temperature dataset

In the ocean, the spatial heterogeneity of sea temperature reflects the structure of temperature distribution (Wang et al. 2024; Zhang et al. 2023). This distribution structure is strongly correlated with heat transport and exchange in the ocean, as well as the distribution and interaction of various water masses. The heterogeneity of temperature distribution can indirectly indicate differences in heat transport patterns within a region (Cheng et al. 2022; Yang et al. 2020).

To comprehensively validate the capability of 3D-SHI to indicate spatial heterogeneity under different scenarios, we selected the IAP Global Ocean Heat Content 1° grid data set, fourth edition (IAPv4), from the Institute of Atmospheric Physics, Chinese Academy of Sciences (Cheng et al. 2024), Figure 10 presents the sea temperature data at various depths in the dataset for June 2023. The study period spans 12 months from January to December 2023. We identified two distinct areas in the North Atlantic as the primary study regions, as illustrated in Figure 11(a): the Greenland-Iceland Area (designated as Area 1 in this paper), a critical region for the formation of North Atlantic Deep Water (NADW) (Larkin et al. 2022), which extends from 0°W to 15°W longitude

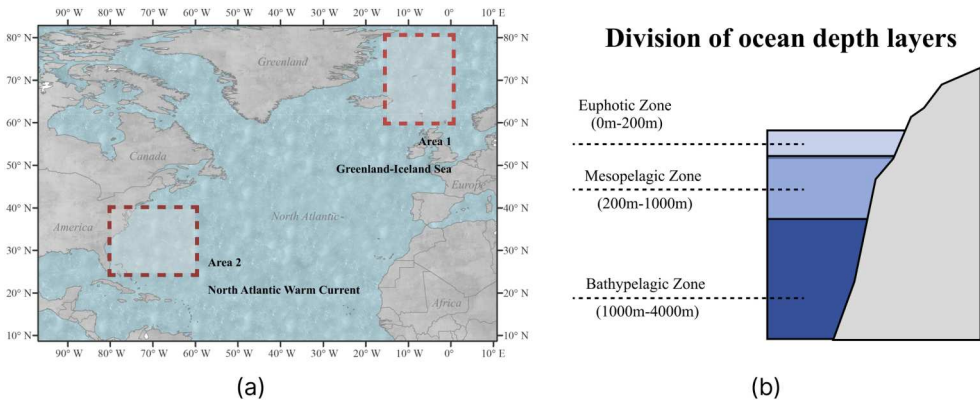


Figure 10. Schematic figure of the study area selection. (a) Horizontal Areal selection: Area1 Greenland-Iceland Area (formation area of North Atlantic Deep Water); Area2 North Atlantic Current Area. (b) Vertical Areal division: Upper layer Euphotic Zone, depth 0–200 m; Middle layer Mesopelagic Zone, depth 200–1000 m; Semi deep sea layer Bathypelagic Zone, depth 1000–4000 m.

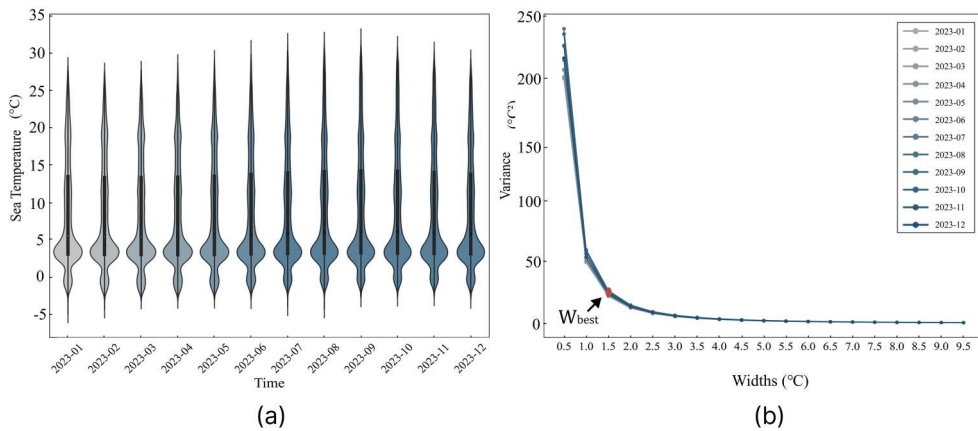


Figure 11. Statistical analysis of experimental results. (a) Distribution of sea temperature across different months. (b) Comparative variance analysis across months using multiple threshold criteria.

and from 60°N to 80°N latitude; and the North Atlantic Current Area (designated as Area 2), characterized by the upwelling of warm water, which ranges from 60°E to 80°E longitude and from 25°N to 40°N latitude (Fröhle, Handmann, and Biastoch 2022). Additionally, the area is subdivided into three depth layers, as shown in Figure 11(b): the upper euphotic zone (0–200 m), the middle mesopelagic zone (200–1000 m), and the bathypelagic zone (1000–4000 m). The horizontal area and depth partitions combine to form six sub-study areas.

Figure 11(a) presents the integrated sea temperature data (January–December 2023) from both study areas. For 3D-SHI computation, we initialized voxel value mapping using a candidate threshold list W ranging from 0.5 to 9.5 with 0.5 intervals. The second-order variance difference analysis based on the algorithm in Section 3.2 (Figure 11(b)) determined 1.5°C to be the optimal threshold. Notably, the lowest temperature in the dataset, -4.38°C , serves as the baseline, and the range from -5°C to -3.5°C is assigned a voxel value of 1, with subsequent voxel values increasing accordingly.

5.2. Simulation experiment results

For the simulation experiment, we first calculated 3D-SHI for the four scenarios, which is shown in Figure 12. We evaluate the performance of 3D-SHI against three established spatial analysis methods (Moran's I, semivariogram, and spatial entropy) applied to the 3D simulation dataset. Quantitative comparisons of computational accuracy and processing time are presented in Table 3. For benchmarking purposes, Table 3 additionally reports the runtime performance of Moran's I and spatial entropy when implemented on 60 equally-spaced 2D slices along the depth axis. The complete 2D slice-based analytical results are graphically displayed in Figure 13. All computations were carried out on an Intel® Xeon® Silver 4216 × 64 workstation. We find that:

Firstly, the 3D-SHI is more sensitive to changes in the scale of large-scale connected components. The 3D-SHI in Scenario 2 decreases compared to Scenario 1, albeit not markedly. This is attributed to the presence of larger-scale connected components for voxel values 1 and 8 within the scenario, without significant changes in the distribution

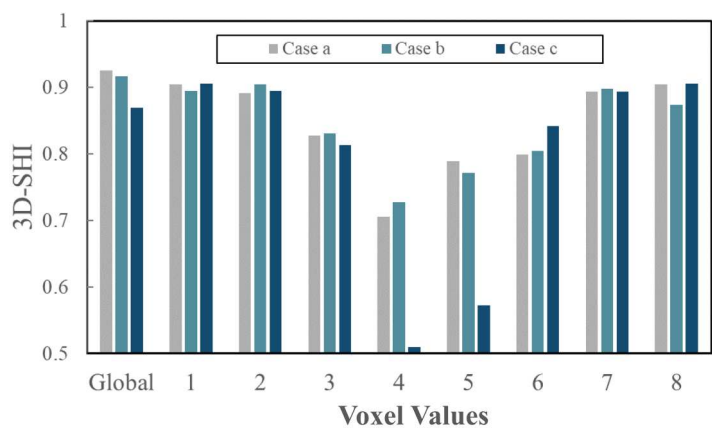


Figure 12. Simulation dataset experiment 3D-SHI calculation results. Noted that ‘Global’ refers to the global calculation results, while the others refer to the voxel-wise calculation results.

Table 3. Calculation results of 3D-SHI and five typical spatial heterogeneity statistical methods for simulation datasets. Noted that the average CPU time includes both data preprocessing and method execution time.

Methods		Calculation Results			Average CPU Time (s)
		Case a	Case b	Case c	
3D-SHI		0.925	0.886	0.707	0.191
Moran’ s I (3D)	Index	0.2175	0.2177	0.228	79.43
	p-value	0.001	0.001	0.001	
Semivariograms (3D)	Nugget	0.02	0.03	0.03	234.47
	Sill	4.8	4.8	4.7	
Spatial Entropy (3D)		2.07	2.07	2.07	0.124
Moran’ s I (2D, 60 slices)		–	–	–	7.792
Spatial Entropy (2D, 60 slices)		–	–	–	0.132

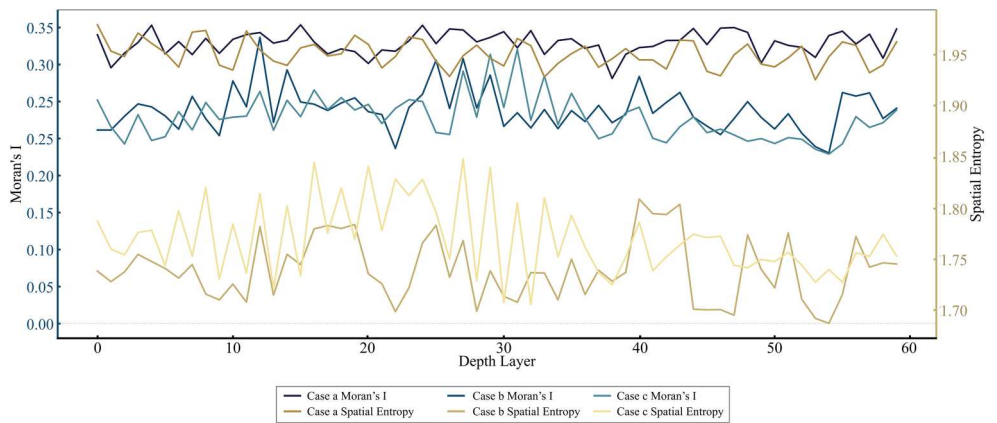


Figure 13. Moran’s I (all $p < 0.05$) and spatial entropy across 60 depth slices.

of the largest connected component volumes. In contrast, the connected components introduced in Scenario 3 are significantly larger than all other connected domains, resulting in a substantial decrease in the 3D-SHI for Scenario 3 compared to the other two scenarios. This observation provides empirical evidence of the greater contribution of large connected components to the 3D-SHI.

Secondly, the calculations of 3D-SHI for each voxel value more accurately detect changes, elucidating the sources of heterogeneity and enhancing the overall 3D-SHI. Scenario 2 differs from Scenario 1 solely by the insertion of a connected component for voxel values 1 and 8. The significant decrease in the 3D-SHI values for these voxel values accounts for the difference from Scenario 1. Furthermore, the global 3D-SHI values for Scenario 3 are significantly lower than those for the other scenarios, which corresponds to the 3D-SHI values for voxel values 4 and 5 being at a higher level, while the 3D-SHI for other voxel values shows no significant difference compared to the other two scenarios.

Thirdly, within the realm of big data, 3D-SHI can achieve near-real-time speeds in quantifying changes in spatial heterogeneity. Results from the three traditional methods exhibit minimal variation across the three scenarios, with spatial entropy calculations revealing almost no differences. This underscores the limitations of traditional methods in representing small-scale spatial aggregation phenomena within large datasets, while 3D-SHI effectively complements and enhances existing spatial characterization techniques. Moreover, 3D-SHI demonstrates a significant advantage in computation time. In a 3D dataset comprising 216000 samples, the 3D-SHI method exhibits superior performance compared to conventional techniques. Specifically, it outperforms Moran's I, which has an average computation time of approximately 80 s, and the semivariogram approach, which averages around 230 s. Although the implementation of Moran's I and spatial entropy on 60 2D slices significantly reduces runtime in comparison to full 3D processing, these slice-based methods inherently overlook heterogeneity along the axis perpendicular to the slices. In contrast, the 3D-SHI method effectively captures heterogeneity in all spatial dimensions while completing calculations over 800 times faster, requiring merely 0.19 s. This result conclusively demonstrates the superiority of our algorithm for large-scale geospatial data analysis.

5.3. Spatial and temporal dimensional heterogeneity of the North Atlantic Sea temperature areas

We first calculated 3D-SHI for six areas across different months, with the results presented in Figure 14. We find that heterogeneity shows significant differences both horizontally and vertically. Horizontally, the 3D-SHI for all layers in Area 1 is markedly greater than that in Area 2, indicating that Area 1 is more characterized by small-scale fragmented sea temperature connected components, leading to more frequent heat exchange phenomena. Vertically, the two areas exhibit similar spatial heterogeneity characteristics across different depth layers; the bathypelagic zone demonstrates significantly lower heterogeneity compared to the upper and middle layers. The heterogeneity in the upper layer is slightly less than that in the middle layer yet it remains high, exceeding 0.9, which reflects the spatial heterogeneity across different depth zones.

On the temporal dimension, heterogeneity in the various areas also displays distinct differences. Figure 14(b–d) illustrate the calculation results for the six sub-areas,

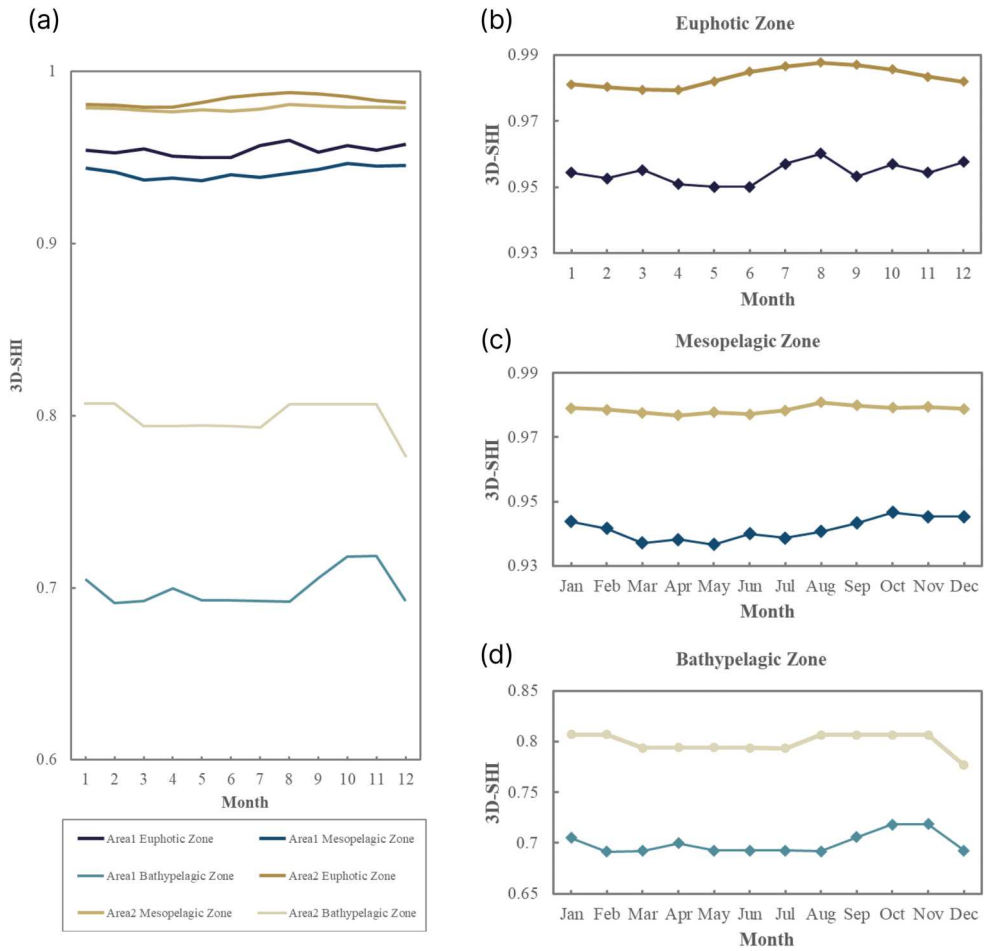


Figure 14. 3D-SHI calculation results for six sub-areas. (a) Summary of monthly 3D-SHI calculations for the six sub-areas. (b–d) Calculation results for each sub-area, where each column represents a different large area, and each row represents the same depth partition.

indicating that the spatial heterogeneity of all sub-areas fluctuates throughout the year, with a clear increase followed by a decline in heterogeneity between May and October. Additionally, in Area 1, the increase in spatial heterogeneity at different depth layers exhibits a clear time lag feature, with the timing of the decrease occurring progressively later as depth increases. This change in heterogeneity may be linked to the formation mechanism of NADW. In contrast, while Area 2 also shows an increase in heterogeneity across different depth layers, it lacks a pronounced time lag feature. The differences in the temporal fluctuations and lag characteristics of spatial heterogeneity between the two areas reflect the temporal changes in temperature structure.

5.4. Discussion

3D-SHI quantifies the scale of connected components within the Earth Cube Model to characterize three-dimensional spatial heterogeneity in big Earth data. This metric

processes several key features: firstly, it demonstrates high sensitivity to variations in the size of large connected components. Additionally, 3D-SHI can accurately detect changes for each individual value, thereby elucidating the sources of heterogeneity. Consequently, it serves as a robust indicator of spatial heterogeneity across different scenarios, including fluctuations in the magnitude of heterogeneity across various spatial distributions. Simulation experiments and case studies further illustrate that 3D-SHI effectively reveals spatiotemporal heterogeneity patterns by assessing spatial heterogeneity in three dimensions.

Moreover, 3D-SHI exhibits considerable potential for high-dimensional datasets, such as those incorporating temporal data or other Earth system attribute variables. Based on the Earth Cube Model we proposed, each voxel in the high-dimensional dataset is extended from a single attribute variable to a high-dimensional spatiotemporal attribute vector. Unlike single-variable datasets, where thresholds can be directly established for discretization to identify connected components, high-dimensional vectors necessitate the use of vector similarity measures (such as cosine similarity (Han, Kamber, and Pei 2012) or Kullback–Leibler divergence (Kullback and Leibler 1951)) to define new connectivity standards and generate novel connected components. This methodology produces an Earth Cube Model that resembles the discretized version of single-attribute variables but incorporates standardized values. By inputting this model into our algorithm, we are able to quantify the spatial heterogeneity of high-dimensional datasets. However, it is essential to note that this metric has certain limitations, which can be highlighted in the following aspects: First, the results of the 3D-SHI calculation are influenced by dataset resolution. Datasets with varying resolutions for the same spatial area yield Earth Cube Models with differing total voxel counts, resulting in discrepancies in connected component scales. This variation can lead to differences in calculations between datasets of different resolutions. Second, when calculating 3D-SHI with continuous attribute values, it is necessary to establish a *Width* parameter for discretization. Variations in *Width* will also affect the 3D-SHI results. Furthermore, it is important to recognize that 3D-SHI is a non-linear metric. In Section 5, we computed 3D-SHI values for scenarios with different levels of spatial heterogeneity, observing a sequential decrease from 1 to 0. The first four scenarios, characterized by relatively dispersed or randomly distributed heterogeneity, show 3D-SHI values decreasing from 1 to 0.655. In contrast, the latter two scenarios which exhibit more aggregated spatial distributions, see a rapid decline in 3D-SHI to 0.5 and subsequently to 0. This swift decrease is attributable to the larger contribution of extensive connected components to 3D-SHI. Consequently, while 3D-SHI is more suitable for comparing the relative levels of spatial heterogeneity across different datasets, it may be less effective in providing a definitive understanding of heterogeneity magnitude based on absolute values.

6. Conclusions

This paper proposes an index called 3D-SHI to characterize the 3D spatial heterogeneity of big Earth data. By designing the Earth Cube model, we provide the foundation for this method, and introduce new concepts such as connected components and homogeneous connected components to capture 3D spatial heterogeneity. The 3D-SHI formulation,

along with the CCM and optimized Union-Find, enables near real-time computation of spatial heterogeneity.

We demonstrated the effectiveness of our method by comparing it with three typical statistical approaches for detecting heterogeneity. The 3D-SHI method accurately captures changes in spatial heterogeneity, particularly when large-scale connected components undergo spatial shifts. It offers a novel means of characterizing spatial heterogeneity across three-dimensional scales, thereby complementing existing methodologies. Subsequently, we applied our method to a 3D simulation dataset and two regional sea temperature datasets from the North Atlantic, evaluating the applicability of 3D-SHI to large Earth data. This method exhibits strong indicative capabilities in scenarios characterized by large-scale connected component distributions. Furthermore, calculating the 3D-SHI for each voxel value allows for the identification of the contributions of different voxel values to global spatial heterogeneity, thereby elucidating the causes of changes in heterogeneity. The case study of sea temperatures in the North Atlantic illustrates that 3D-SHI reveals significant spatial heterogeneity across both spatial and temporal dimensions. Horizontally, the Greenland-Iceland area displays lower heterogeneity compared to the North Atlantic Current area. Vertically, depth bands in both regions exhibit similar patterns of heterogeneity. Temporally, heterogeneity fluctuates, with oscillatory patterns observed in the upper and middle layers of the Greenland-Iceland area, while the North Atlantic Current area experiences distinct variations. The bathypelagic zones in both regions demonstrate greater stability. Additionally, in the Greenland-Iceland area, heterogeneity increases with depth from June to November, revealing a lag that is not present in the North Atlantic Current area.

This study shows that 3D-SHI can effectively quantify 3D spatial heterogeneity and is particularly suitable for measuring the spatial heterogeneity of 3D big Earth data. Future directions for this work encompass several short-term extensions, including the expansion of use cases for our method and the development of 3D-SHI indicators to assess local area spatial heterogeneity. Additionally, we aim to incorporate spatial information into connected components to identify regions exhibiting the greatest spatial heterogeneity. Furthermore, we plan to extend this method to accommodate spatio-temporal big data and large datasets with multiple attributes. A long-term objective of this research is to study spatial heterogeneity within 3D large geospatial data. Compared to traditional 2D geospatial data, this endeavor involves larger scales, higher spatial and temporal resolutions, and more complex spatial distribution characteristics. These factors present challenges for simplifying the CCM data structure in the future, with the goals of conserving storage space and enhancing algorithm efficiency.

Acknowledgments

The authors thank the editor and the anonymous reviewers for their helpful comments on an earlier draft of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Key Research and Development Program of China (grant number 2022YFB3904102) and Innovation Project of LREIS (grant number 08R8A092YA).

Data availability statement

The 4th version of the Institute of Atmospheric Physics (IAPv4) ocean temperature dataset that we used in this article is freely accessible at <https://doi.org/10.12157/IOCAS.20240117.002>.

ORCID

Peixiao Wang  <http://orcid.org/0000-0002-1209-6340>

References

- Borrelli, J. J., and R. A. Relyea. 2022. "A Review of Spatial Structure of Freshwater Food Webs: Issues and Opportunities Modeling within-Lake Meta-ecosystems." *Limnology and Oceanography* 67 (8): 1746–1759. <https://doi.org/10.1002/lno.12163>.
- Cai, W. J., G. J. Wang, B. Dewitte, L. X. Wu, A. Santoso, K. Takahashi, Y. Yang, A. Carréric, and M. J. McPhaden. 2018. "Increased Variability of Eastern Pacific El Nino under Greenhouse Warming." *Nature* 564 (7735): 201–206. <https://doi.org/10.1038/s41586-018-0776-9>.
- Chen, Y. G., and L. S. Huang. 2018. "Spatial Measures of Urban Systems: From Entropy to Fractal Dimension." *Entropy* 20 (12): 991. <https://doi.org/10.3390/e20120991>.
- Cheng, L., Y. Pan, Z. Tan, H. Zheng, Y. Zhu, W. Wei, J. Du, et al. 2024. "IAPv4 Ocean Temperature and Ocean Heat Content Gridded Dataset." *Earth System Science Data* 16 (8): 3517–3546. <https://doi.org/10.5194/essd-16-3517-2024>.
- Cheng, Lijing, Karina von Schuckmann, John P. Abraham, Kevin E. Trenberth, Michael E. Mann, Laure Zanna, Matthew H. England, et al. 2022. "Past and Future Ocean Warming." *Nature Reviews Earth & Environment* 3 (11): 776–794. <https://doi.org/10.1038/s43017-022-00345-1>.
- Fan, Zhang, Xuan Xin, and Deng Xiangzheng. 2021. "Research Progress and Prospect on the Non-Uniform Distribution of Atmospheric CO₂ Concentration and Its Influence on Surface Warming." *Journal of Geo-Information Science* 23 (8): 1362–1371. <https://doi.org/10.12082/dqxkx.2021.210041>.
- Fortin, M. J., P. M. A. James, A. MacKenzie, S. J. Melles, and B. Rayfield. 2012. "Spatial Statistics, Spatial Regression, and Graph Theory in Ecology." *Spatial Statistics* 1:100–109. <https://doi.org/10.1016/j.spasta.2012.02.004>.
- Fotheringham, A. S., and M. Sachdeva. 2022. "On the Importance of Thinking Locally for Statistics and Society." *Spatial Statistics* 50:100601. <https://doi.org/10.1016/j.spasta.2022.100601>.
- Fröhle, J., P. V. K. Handmann, and A. Biastoch. 2022. "Major Sources of North Atlantic Deep Water in the Subpolar North Atlantic from Lagrangian Analyses in an Eddy-Rich Ocean Model." *Ocean Science* 18 (5): 1431–1450. <https://doi.org/10.5194/os-18-1431-2022>.
- Galler, Bernard A., and Michael J. Fischer. 1964. "An Improved Equivalence Algorithm." *Communications of the ACM* 7 (5): 301–303. <https://doi.org/10.1145/364099.364331>.
- Gao, Yunqiu, Xuhui Lee, Shoudong Liu, Ning Hu, Xiao Wei, Cheng Hu, Cheng Liu, Zhen Zhang, and Yichen Yang. 2018. "Spatiotemporal Variability of the Near-Surface CO₂ Concentration across an Industrial-Urban-Rural Transect, Nanjing, China." *Science of the Total Environment* 631–632:1192–1200. <https://doi.org/10.1016/j.scitotenv.2018.03.126>.
- Gao, H., J. Y. Tham, P. Xue, and W. Lin. 2008. "Complexity Analysis of Morphological Area Openings and Closings with Set Union." *IET Image Processing* 2 (4): 231–238. <https://doi.org/10.1049/iet-ipr:20070174>.

- Gaudart, J., R. Giorg, B. Poudiougou, O. Toure, S. Ranque, O. Doumbo, and J. Demongeot. 2007. "Spatial Cluster Detection without Point Source Specification: The Use of Five Methods and Comparison of Their Results." *Revue D'Épidémiologie et de Santé Publique* 55 (4): 297–306. <https://doi.org/10.1016/j.respe.2007.04.003>.
- Geary, Robert C. 1954. "The Contiguity Ratio and Statistical Mapping." *The Incorporated Statistician* 5 (3): 115–146. <https://doi.org/10.2307/2986645>
- Goodchild, M. F. 2004. "The Validity and Usefulness of Laws in Geographic Information Science and Geography." *Annals of the Association of American Geographers* 94 (2): 300–303. <https://doi.org/10.1111/j.1467-8306.2004.09402008.x>.
- Guo, H. D., Z. Liu, H. Jiang, C. L. Wang, J. Liu, and D. Liang. 2017. "Big Earth Data: A new Challenge and Opportunity for Digital Earth's Development." *International Journal of Digital Earth* 10 (1): 1–12. <https://doi.org/10.1080/17538947.2016.1264490>.
- Guo, H. D., L. Z. Wang, F. Chen, and D. Liang. 2014. "Scientific Big Data and Digital Earth." *Chinese Science Bulletin* 59 (35): 5066–5073. <https://doi.org/10.1007/s11434-014-0645-3>.
- Guo, H. D., L. Z. Wang, and D. Liang. 2016. "Big Earth Data from Space: A New Engine for Earth Science." *Science Bulletin* 61 (7): 505–513. <https://doi.org/10.1007/s11434-016-1041-y>.
- Han, Jiawei, Micheline Kamber, and Jian Pei. 2012. "2 - Getting to Know Your Data." In *Data Mining*, edited by Jiawei Han, Micheline Kamber, and Jian Pei, 39–82. 3rd ed. Boston: Morgan Kaufmann.
- He, C. P., M. R. Ji, M. L. Grieneisen, and Y. Zhan. 2022. "A Review of Datasets and Methods for Deriving Spatiotemporal Distributions of Atmospheric CO₂." *Journal of Environmental Management* 322:116101. <https://doi.org/10.1016/j.jenvman.2022.116101>.
- He, H. L., Z. Ling, S. C. Wu, X. Lyu, Z. Zeng, R. Z. Tian, Y. Wang, and J. Sun. 2024. "In Situ Observation of Ocean Response to Tropical Cyclone in the Western North Pacific during 2022." *Frontiers in Marine Science* 11:1445071. <https://doi.org/10.3389/fmars.2024.1445071>.
- Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, et al. 2020. "The ERA5 Global Reanalysis." *Quarterly Journal of the Royal Meteorological Society* 146 (730): 1999–2049. <https://doi.org/10.1002/qj.3803>.
- Hu, Shijian, Dunxin Hu, Cong Guan, Nan Xing, Jianping Li, and Junqiao Feng. 2017. "Variability of the Western Pacific Warm Pool Structure Associated with El Niño." *Climate Dynamics* 49 (7): 2431–2449. <https://doi.org/10.1007/s00382-016-3459-y>.
- Huang, F., S. D. Jiang, W. F. Zhan, B. Bechtel, Z. H. Liu, M. Demuzere, Y. Huang, et al. 2023. "Mapping Local Climate Zones for Cities: A Large Review." *Remote Sensing of Environment* 292:113573. <https://doi.org/10.1016/j.rse.2023.113573>.
- Kullback, S., and R. A. Leibler. 1951. "On Information and Sufficiency." *The Annals of Mathematical Statistics* 22 (1): 79–86. <https://doi.org/10.1214/aoms/1177729694>.
- Larkin, Christina S., Mohamed M. Ezat, Natalie L. Roberts, Henning A. Bauch, Robert F. Spielhagen, Riko Noormets, Leonid Polyak, et al. 2022. "Active Nordic Seas Deep-Water Formation during the Last Glacial Maximum." *Nature Geoscience* 15 (11): 925–931. <https://doi.org/10.1038/s41561-022-01050-w>.
- Li, Q., G. Shi, W. Shangguan, V. Nourani, J. Li, L. Li, F. Huang, et al. 2022. "A 1 km Daily Soil Moisture Dataset over China Using *in situ* Measurement and Machine Learning." *Earth System Science Data* 14:5267–5286. <https://doi.org/10.5194/essd-14-5267-2022>.
- Li, Habin, Zhengquan Wang, and Qingcheng Wang. 1998. "Theory and Methodology of Spatial Heterogeneity Quantification." *Yingyong Shengtai Xuebao* 9 (6): 651–657.
- Liang, Zhongtong, Wenhan Tang, Ning Zeng, Qixiang Cai, Pengfei Han, Yu Zhang, Weijun Quan, Bo Yao, Pucai Wang, and Zhiqiang Liu. 2022. "Simulation of Surface Atmospheric CO₂ Concentration in Beijing-Tianjin-Hebei Region Based on WRF Model." *Transactions of Atmospheric Sciences* 45 (3): 387–396. <https://doi.org/10.13878/j.cnki.dqkxxb.20220228018>.
- Liao, Xiaoyu, Zhaoli Liu, Yiyong Wang, and Jiming Jin. 2013. "Spatiotemporal Variation in the Microclimatic Edge Effect between Wetland and Farmland." *Journal of Geophysical Research: Atmospheres* 118 (14): 7640–7650. <https://doi.org/10.1002/jgrd.50573>

- Luo, P., Y. Z. Song, D. Zhu, J. Y. Cheng, and L. Q. Meng. 2023. "A Generalized Heterogeneity Model for Spatial Interpolation." *International Journal of Geographical Information Science* 37 (3): 634–659. <https://doi.org/10.1080/13658816.2022.2147530>.
- Mahadevan, A. 2003. "Spatial Heterogeneity and Its Relation to Processes in the Upper Ocean." Paper Presented at the 10th Cary Conference, Inst Ecosyst Studies, Millbrook, NY, Apr 29–May 01.
- Matheron, Georges. 1963. "Principles of Geostatistics." *Economic Geology* 58 (8): 1246–1266. <https://doi.org/10.2113/gsecongeo.58.8.1246>.
- Meyer, Rena, Peter Engesgaard, Anne-Sophie Høyer, Flemming Jørgensen, Giulio Vignoli, and Torben O Sonnenborg. 2018. "Regional Flow in a Complex Coastal Aquifer System: Combining Voxel Geological Modelling with Regularized Calibration." *Journal of Hydrology* 562:544–563. <https://doi.org/10.1016/j.jhydrol.2018.05.020>
- Moran, P. A. 1950. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37 (1-2): 17–23. <https://doi.org/10.1093/biomet/37.1-2.17>.
- Orgiazzi, A., C. Ballabio, P. Panagos, A. Jones, and O. Fernández-Ugalde. 2018. "LUCAS Soil, the Largest Expandable Soil Dataset for Europe: A Review." *European Journal of Soil Science* 69 (1): 140–153. <https://doi.org/10.1111/ejss.12499>.
- Pelletier, J. D., P. D. Broxton, P. Hazenberg, X. Zeng, P. A. Troch, G. Niu, Z. C. Williams, M. A. Brunke, and D. Gochis. 2016. *Global 1-km Gridded Thickness of Soil, Regolith, and Sedimentary Deposit Layers*. Oak Ridge, Tennessee, USA: ORNL DAAC. <https://doi.org/10.3334/ORNLDAAC/1304>.
- Potapov, P., M. C. Hansen, I. Kommareddy, A. Kommareddy, S. Turubanova, A. Pickens, B. Adusei, A. Tyukavina, and Q. Ying. 2020. "Landsat Analysis Ready Data for Global Land Cover and Land Cover Change Mapping." *Remote Sensing* 12 (3):426. <https://doi.org/10.3390/rs12030426>.
- Qin, S. S., Q. L. Zhang, and B. S. Yin. 2015. "Seasonal Variability in the Thermohaline Structure of the Western Pacific Warm Pool." *Acta Oceanologica Sinica* 34 (7): 44–53. <https://doi.org/10.1007/s13131-015-0696-6>.
- Resplandy, L., A. Hogikyan, J. D. Müller, R. G. Najjar, H. W. Bange, D. Bianchi, T. Weber, et al. 2024. "A Synthesis of Global Coastal Ocean Greenhouse Gas Fluxes." *Global Biogeochemical Cycles* 38 (1): e2023GB007803. <https://doi.org/10.1029/2023gb007803>.
- Rew, R., and G. Davis. 1990. "NetCDF – an Interface for Scientific-Data Access." *IEEE Computer Graphics and Applications* 10 (4): 76–82. <https://doi.org/10.1109/38.56302>.
- Roemmich, D., M. H. Alford, H. Claustre, K. Johnson, B. King, J. Moum, P. Oke, et al. 2019. "On the Future of Argo: A Global, Full-Depth, Multi-disciplinary Array." *Frontiers in Marine Science* 6: 439. <https://doi.org/10.3389/fmars.2019.00439>.
- Shen, Dayong, Kaoru Takara, and Yuling Liu. 2016. 3D GIS Modeling of Soft Geo-Objects: Taking Rainfall, Overland Flow, and Soil Erosion as an Example." In *Geospatial Technologies, Environment, and Society Applications*. InTech. <https://doi.org/10.5772/61680>.
- Shen, D. Y., Kaoru Takara, Yasuto Tachikawa, and Yueh-Lin Liu. 2006. "3D Simulation of Soft Geo-Objects." *International Journal of Geographical Information Science* 20 (3): 261–271. <https://doi.org/10.1080/13658810500287149>
- Shen, Fengbei, Chengdong Xu, Jinfeng Wang, Maogui Hu, Guanlin Guo, Tingting Fang, Xingbao Zhu, Hongying Cao, Huan Tao, and Yixuan Hou. 2024. "A New Method for Spatial Three-Dimensional Prediction of Soil Heavy Metals Contamination." *CATENA* 235:107658. <https://doi.org/10.1016/j.catena.2023.107658>
- Sun, J., S. S. Jin, X. Ju, H. L. He, R. B. Ding, X. M. Hu, and X. J. Xiong. 2024. "Influence of Spatial Heterogeneity in Sea Surface Temperature on Tropical Cyclone Intensity over the Western North Pacific." *Journal of Geophysical Research-Atmospheres* 129 (4): e2023JD038971. <https://doi.org/10.1029/2023jd038971>.
- Swenson, M. S., and Ams Ams. 2000. "The Global Drifter Program." Paper Presented at the 16th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and hydrology, Long Beach, Ca, Jan 09–14.

- Taghizadeh-Mehrjardi, R., B. Minasny, F. Sarmadian, and B. P. Malone. 2014. "Digital Mapping of Soil Salinity in Ardakan Region, Central Iran." *Geoderma* 213:15–28. <https://doi.org/10.1016/j.geoderma.2013.07.020>.
- Tian, Li, Xiaopan Zhang, and Yanbin Yuan. 2014. "Improved Algorithm on Global Morans I of Regular Grid System." *Science of Surveying and Mapping* 39 (2): 110–113.
- Tobler, W. 2004. "On the First law of Geography: A Reply." *Annals of the Association of American Geographers* 94 (2): 304–310. <https://doi.org/10.1111/j.1467-8306.2004.09402009.x>.
- Van Lancker, Vera, Frederic Francken, Lars Kint, Nathan Terseleer, Dries Van den Eynde, Lies De Mol, Guy De Tré, Robin De Mol, Tine Missiaen, and Vasileios Hademenos. 2017. "Building a 4D Voxel-Based Decision Support System for a Sustainable Management of Marine Geological Resources." In *Oceanographic and Marine Cross-Domain Data Management for Sustainable Development*, edited by P. Diviacco, A. Leadbetter, and H. Glaves, 224–252. Hershey: IGI Global.
- Wang, Haodi, Shiyao Chen, Senliang Bao, and Kaijun Ren. 2022. "Research Progress of Global Gridded Ocean Environment Datasets." *Advance in Earth Sciences* 37 (8): 822–840.
- Wang, Chen, Changyuan Chen, Huimin Li, and Yijun He. 2024. "Feature-Oriented Reconstruction of Vertical Temperature Profile: A Feasibility Study in the Northwest Pacific Ocean." *Deep Sea Research Part I: Oceanographic Research Papers* 203:104201. <https://doi.org/10.1016/j.dsr.2023.104201>
- Wang, Xianghong, Jiping Liu, Shenghua Xu, and Yong Wang. 2013. "Visualization of Marine Environment Data Based on NetCDF Data Model." *Science of Surveying and Mapping* 38 (2): 59–61.
- Wang, Jinfeng, and Chengdong Xu. 2017. "Geodetector: Principle and Prospective." *Acta Geographica Sinica* 72 (1): 116–134.
- Willcock, S., J. Martinez-Lopez, N. Dandy, and J. M. Bullock. 2021. "High Spatial-Temporal Resolution Data across Large Scales Are Needed to Transform Our Understanding of Ecosystem Services." *Land* 10 (7): 759. <https://doi.org/10.3390/land10070759>.
- Wu, Q., F. X. Guo, H. Q. Li, and J. Y. Kang. 2017. "Measuring Landscape Pattern in Three Dimensional Space." *Landscape and Urban Planning* 167:49–59. <https://doi.org/10.1016/j.landurbplan.2017.05.022>.
- Xu, Yusheng, Xiaohua Tong, and Uwe Stilla. 2021. "Voxel-Based Representation of 3D Point Clouds: Methods, Applications, and Its Potential Use in the Construction Industry." *Automation in Construction* 126:103675. <https://doi.org/10.1016/j.autcon.2021.103675>.
- Xu, Yusheng, Sebastian Tuttas, Ludwig Hoegner, and Uwe Stilla. 2018. "Voxel-Based Segmentation of 3D Point Clouds from Construction Sites Using a Probabilistic Connectivity Model." *Pattern Recognition Letters* 102:67–74. <https://doi.org/10.1016/j.patrec.2017.12.016>
- Zhang Y., Y. Du, Feng M, and Alistair J. Hobday. 2023. "Vertical Structures of Marine Heatwaves." *Nature Communications* 14:6483. <https://doi.org/10.1038/s41467-023-42219-0>.
- Yang, Yiping, Rong Xiang, Lanlan Zhang, Fuchang Zhong, and Miming Zhang. 2020. "Is the Upward Release of Intermediate Ocean Heat Content a Possible Engine for Low-Latitude Processes?" *Geology* 48 (6): 579–583. <https://doi.org/10.1130/g47271.1>.
- Yang, J., and R. Yao. 2007. "Spatial Variability of Soil Water and Salt Characteristics in the Yellow River Delta." *Scientia Geographica Sinica* 27 (3): 348.
- Ying, Zhixia, Yongjie Liu, Chen Shi, Zhichun Lan, Huan Zhang, Yizhen Liu, and Gang Ge. 2021. "Simulating Root Distribution of Plant Individual with a Three-Dimensional Model." *Ecological Modelling* 455:109649. <https://doi.org/10.1016/j.ecolmodel.2021.109649>.
- Zhang, H. P., X. X. Zhou, Y. Yang, H. R. Wang, X. Y. Ye, and G. A. Tang. 2024. "Advancing Process-Oriented Geographical Regionalization Model." *Annals of the American Association of Geographers* 114 (10): 2388–2413. <https://doi.org/10.1080/24694452.2024.2380893>.
- Zhao, W. W., X. N. Fang, S. Daryanto, X. Zhang, and Y. P. Wang. 2019. "Factors Influencing Soil Moisture in the Loess Plateau, China: A Review." *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* 109 (3-4): 501–509. <https://doi.org/10.1017/s1755691018000671>.

- Zhao, R., L. P. Zhan, M. X. Yao, and L. C. Yang. 2020. "A Geographically Weighted Regression Model Augmented by Geodetector Analysis and Principal Component Analysis for the Spatial Distribution of PM_{2.5}." *Sustainable Cities and Society* 56:102106. <https://doi.org/10.1016/j.scs.2020.102106>.
- Zhu, Q. Q., Y. Lei, X. L. Sun, Q. F. Guan, Y. F. Zhong, L. P. Zhang, and D. R. Li. 2022. "Knowledge-Guided Land Pattern Depiction for Urban Land Use Mapping: A Case Study of Chinese Cities." *Remote Sensing of Environment* 272:112916. <https://doi.org/10.1016/j.rse.2022.112916>.