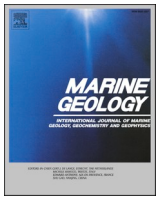
Marine Geology 486 (2025) 107566

Contents lists available at ScienceDirect

Marine Geology

journal homepage: www.elsevier.com/locate/margo



Testing spatial interpolation methods for deep-time organic carbon burial in epicontinental seas by taking Sunda Shelf as an example

Yida Yang [a,b], Pengfei Ma [c,*], Xiumian Hu [d], Yuan Gao [a,b], Chengshan Wang [a,b]

[a] State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Beijing 100083, China
[b] School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China
[c] State Key Laboratory of Marine Geology, Tongji University, Shanghai 200092, China
[d] State Key Laboratory of Mineral Deposits Research, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China

ARTICLE INFO

Editor Name: Dr. Natascha Riedinger

Keywords:
Epicontinental sea
Organic carbon burial
Spatial interpolation
Random forests
Deep time
Sunda Shelf

ABSTRACT

Quantifying the burial of organic carbon (OC) in epicontinental seas is crucial for understanding its role in regulating global long-term carbon cycle and climate. Utilizing spatial interpolation methods, prior works have quantified OC burial globally or regionally based on limited, unevenly distributed measurements. However, there remains a notable lack of comparative studies and assessments regarding their applicability and uncertainty in deep-time research. Taking the middle Miocene Sunda Shelf OC burial estimation as an example, four popular spatial interpolation methods are assessed quantitatively and qualitatively, including Thiessen polygons, Inverse Distance Weighting (IDW), Ordinary Kriging (OK) and Random Forests (RF). Based on quantitative and qualitative evaluation, the data-driven RF method demonstrates superior performance due to fewer assumptions, effectively capturing nonlinear relationships and complex spatial patterns in heterogeneous, non-Gaussian deep-time data, and demonstrating strong generalizability and robustness. High-resolution RF-based reassessment reveals significant spatial-temporal heterogeneity of OC burial on the Sunda Shelf between the Miocene Climatic Optimum (MCO) and Middle Miocene Climate Transition (MMCT). Although the overall OC burial and sediment accumulation rates (SAR) increase during the MMCT, notable spatial discrepancies emerge, with OC burial rates elevated near basin margins but decreased in distal regions. These variations primarily reflect the combined influences of eustatic sea-level fall and enhanced terrigenous input, highlighting the complex interplay of factors modulating OC burial efficiency. Machine learning methods such as RF prove highly effective in handling deep-time spatial data, but their application should be adapted to specific objectives, geological conditions, and data characteristics.

1. Introduction

Epicontinental seas, gently sloping shallow sea areas covering the continental shelves or interiors, have occurred widely throughout the geological history and have been an integral part of the global ocean system (Liu, 2001; Harries, 2009; Scotese, 2021). Owing to their distinct locations and the interactions between continents, oceans, and atmosphere, the composition and spatial distribution of sediment in epicontinental seas exhibit considerable complexity and diversity. These features not only provide direct evidence of past climatic, oceanic, and environmental changes but also aid in understanding the evolution of epicontinental seas, along with their profound impact on the earth system evolution (Alexander et al., 1991; Liu, 2001; Bastos et al., 2021; Kaya et al., 2022; Schwarz et al., 2022).

Organic carbon (OC), widely buried in the epicontinental seas, accounts for approximately 80 % of total OC burial in the global oceans, significantly impacting the global carbon cycle and climate system through various direct and indirect feedbacks (Bauer et al., 2013). A quantitatively accurate assessment of the OC burial spatiotemporal distribution in epicontinental seas is a key step for reconstructing the long-term carbon cycle model and interpreting the complex climate feedback mechanisms (Berner, 1999; Hu et al., 2016; Lin et al., 2023). Currently, there are two methodologies to quantify deep-time OC burial in marine sediments (Li et al., 2023): (i) "top-down" approach based on carbon isotopic mass balance model (e.g., Derry and France-Lanord, 1996) and (ii) "bottom-up" approach based on measured site or

* Corresponding author.
E-mail address: pma@tongji.edu.cn (P. Ma).

https://doi.org/10.1016/j.margeo.2025.107566
Received 21 January 2025; Received in revised form 5 April 2025; Accepted 19 April 2025
Available online 25 April 2025
0025-3227/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

section data (e.g., Kaya et al., 2022). In marine sediments, the significant carbon isotope fractionation associated with photosynthesis results in OC with lower $\delta^{13}C$ values, enabling the use of carbonate $\delta^{13}C$ signatures in sediments to calculate the global burial flux of organic carbon (Kump and Arthur, 1999). During OAE2, a worldwide oceanic anoxic event in the Cretaceous, previous studies estimated the total OC burial (~90 Eg, $1Eg = 10^{18}$ g) in marine sediments based on a carbon isotopic mass balance model (Owens et al., 2018). The "top-down" approach, however, raises concerns due to its reliance on simple assumptions regarding stable carbon input and constant carbon isotope composition (Li et al., 2023). Indeed, the sources of OC are diverse and the carbon isotope composition shows considerable variability, complicating calculations based on isotopic mass balance (Shields and Mills, 2017; Lin et al., 2023). Moreover, this "top-down" approach for calculating marine sediment OC burial seems applicable only to global, not regional studies. While the "bottom-up" approach, unaffected by variations in the source and isotopic composition of OC, enables straightforward calculation of OC burial using measurements of sediment thickness, dry bulk density, and total organic carbon (TOC) (e.g., Owens et al., 2018; Li et al., 2023). Measurements of a single site or section can provide long-term changes in OC content but cannot represent spatial changes and the total amount of OC buried in a region. Moreover, comprehensive and regular collection of sediment samples, which is time- and cost-consuming and impractical, makes "bottom-up" approach challenging (Bamisaiye, 2018; Lee et al., 2019). Therefore, it is necessary to derive data for unexplored regions from limited site or section data through rational spatial interpolation (Burrough and McDonnell, 2015).

Over the past half-century, with the development of Geographic Information Systems (GIS) and the accumulation of fundamental observation datasets, spatial interpolation techniques have been widely applied in various disciplines (e.g., Chen et al., 2010; Li et al., 2011; Heung et al., 2014; Antonakos, 2021). These methods can be categorized into traditional spatial interpolation and machine learning spatial interpolation, based on their model characteristics and complexity (Li et al., 2011; Hengl et al., 2018). Traditional spatial interpolation methods estimate values at unsampled locations by weighting the values of known sample points, based on spatial relationships among sample points and the assumption of spatial statistics (Li and Heap, 2008; Burrough and McDonnell, 2015). They fundamentally adhere to the first and second laws of geography describing spatial autocorrelation and heterogeneity (Zhu et al., 2018), including deterministic spatial interpolation methods represented by Thiessen polygons and Inverse Distance Weighting (IDW), as well as geostatistical methods such as Kriging (Li and Heap, 2008). Machine learning methods represent a distinct approach based on data-driven strategies, which are not limited by the assumption of traditional methods and exhibit significant advantages in learning and handling nonlinear, complex, large-scale spatial data problems (Li et al., 2011; Du et al., 2020). They can also integrate various data sources and environmental variables, offering more accurate, flexible, and reliable interpolation solutions. Random Forests (RF), a typical representative of machine learning methods, have been successfully applied in multiple fields (e.g., Li et al., 2011; Graw et al., 2021). For instance, Graw et al. (2021) used RF to map global seafloor sediment density, achieving higher accuracy than methods based on porosity grids. Though these spatial interpolation methods were predominantly applied in modern geographical research, some scholars have also made some limited attempts in deep-time geological studies. Ma et al. (2022) and Kaya et al. (2022) similarly employed the Thiessen polygon method to calculate the OC burial of the Sunda Shelf during the middle Miocene and the eastern epicontinental seas of Eurasia during the Paleocene–Eocene Thermal Maximum (PETM), respectively. However, these methods frequently neglect the spatial distribution differences and non-Gaussian nature of geological parameters such as sediment thickness and TOC content. Using the average value of one or a few drilling data to represent a vast area still involves considerable inaccuracies, particularly when the spatiotemporal distribution of available drilling data is uneven (Li and Heap, 2008). Moreover, based on current insights, it appears that these methods have not undergone comparative studies and reliability assessments, and there may be significant discrepancies between different spatial interpolation methods in estimating the total amount of OC burial during geological periods (Kaya et al., 2022).

In this study, to better explore the application and comparison of spatial interpolation methods in the study of epicontinental sea OC burial estimation during geological periods, we take the Sunda Shelf as a case study area, aiming to address the following problems and objectives: (1) systematically compare the performance of different spatial interpolation methods in predicting the spatial distribution of OC burial: Thiessen polygons used in the previous study (Ma et al., 2022), IDW, Kriging and Random Forest; (2) provide more reliable and accurate high-resolution spatial interpolation results of OC burial on the Sunda Shelf in the middle Miocene.

2. Regional settings

The Sunda Shelf, located in the tropical region of Southeast Asia (Fig. 1A), is the largest continental shelf outside the polar regions. It covers an area of approximately 1.8 million square kilometers, with an average depth of less than 100 m (Fig. 1B). This shelf was initially formed during the Early Mesozoic by the addition of continental fragments carried from Gondwana to Southeast Asia (Hall, 2012). Since then, the Sunda Shelf has remained within tropical low-latitude areas (Scotese, 2021), and potentially played a key role in global climate evolution (e.g., Hall, 2009; Park et al., 2020; Bayon et al., 2023). From the late Eocene to middle Oligocene, multiple rifting basins developed across the Sunda Shelf, accumulating thick successions of lacustrine shales and fluvio-deltaic sediments during the *syn-* and post- rift phases (Shoup et al., 2012; Morley and Morley, 2013; Pubellier and Morley, 2014; de Jonge-Anderson et al., 2024). Since the earliest Miocene, the Shelf has been relatively stable with gentle slopes and a wide epicontinental sea similar to the present day has appeared (Morley et al., 2016). This transgressive environment was primarily driven by regional subsidence and the South China Sea opening (Hall, 2012; Morley et al., 2016), and may have been further enhanced by the global Miocene sea level maximum (Miller et al., 2020). The occurrence of some planktonic assemblages indicates that much of the shelf was reaching shallow sea (Morley et al., 2021). Broad carbonate platforms also developed in the northern and eastern parts of the shelf (Fig. 1B; Leong, 2000; Yang et al., 2016). During the middle Miocene, extensive mangrove ecosystems flourished along the paleo-coastlines, contributing significant quantities of OC to the epicontinental sea (Morley and Morley, 2013; Collins et al., 2017). In addition, major paleo-rivers (e.g., Kelantan, proto-Mekong, Chao Phraya, Kapuas Rivers) delivered substantial terrigenous components from surrounding highlands (Morley et al., 2016; Breitfeld et al., 2022). Accelerated terrigenous input may potentially enhance marine productivity and OC burial (e.g., Galy et al., 2007; Xu et al., 2021b; Müller et al., 2024). As a critical region for OC burial during the middle Miocene, the Sunda Shelf may have played a crucial role in the modulating carbon cycle dynamics during the middle Miocene greenhouse (the Miocene Climatic Optimum, MCO) and the subsequent climate transition (the Middle Miocene Climate Transition, MMCT) (Ma et al., 2022).

Several internal basins, including Pattani, Malay, Penyu, West Natuna, Mekong, Wan'an, and Sarawak, distributed across the Sunda Shelf (Fig. 1B). These internal basins were semi-isolated and likely exhibited different sedimentary responses influenced by regional tectonic during the Oligocene – Miocene (Morley et al., 2016). Nevertheless, they were all affected by global climate-driven sea-level fluctuations, resulting in multiple, frequent transgressive–regressive sedimentary sequences (Morley et al., 2021). This prominent characteristic has facilitated stratigraphic correlation among the internal basins across the Sunda Shelf and enabled the establishment of a unified

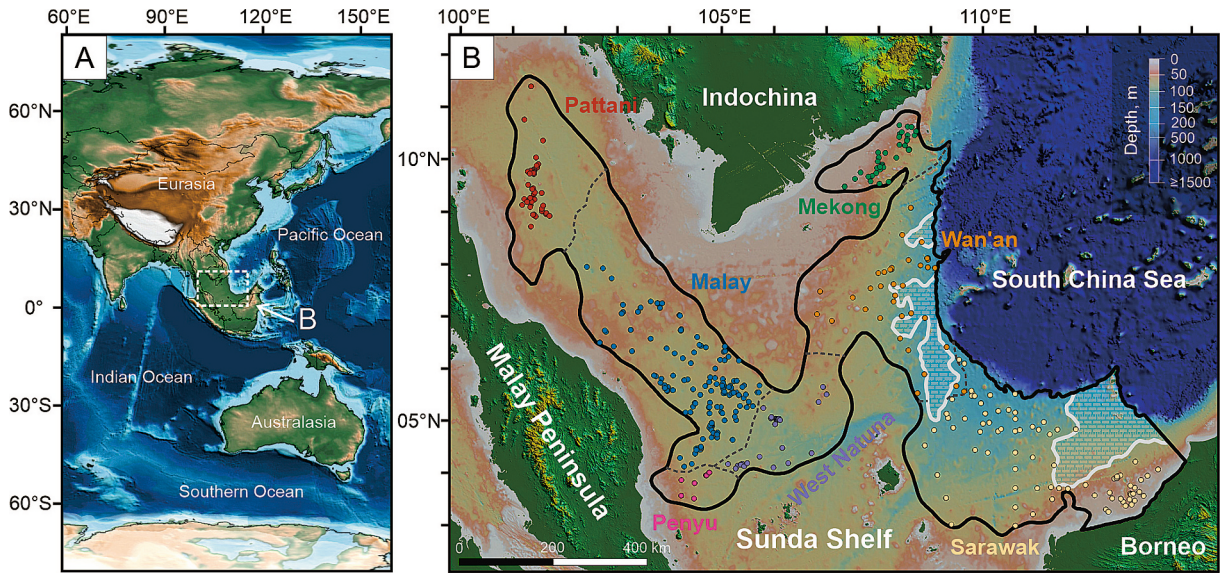


Fig. 1. Geographic map of Sunda Shelf and distribution of commercial drilling sites (modified from Ma et al., 2022). (A) The Sunda Shelf was located in the tropical region at the junction of the Pacific, Indian Oceans and Eurasia, Australasia continents in middle Miocene (~15 Ma; Scotese, 2021). (B) Distribution of drilling sites used in Spatial interpolation across the Sunda Shelf (367 sites in total). Different colors represent different basins: Pattani Basin (41 sites), Malay Basin (137 sites), Penyu Basin (7 sites), West Natuna Basin (28 sites), Mekong Basin (37 sites), Wan'an Basin (37 sites), Sarawak Basin (80 sites). The sedimentation boundaries of Sunda Shelf (solid black line) and internal basins (light black dotted line) were conservatively defined based on sediment thickness isopach maps (2000 m; Ramli, 1988; Straume et al., 2019). The light blue brick-shaped polygons, which represent Miocene carbonate platforms (Leong, 2000; Kob and Ali, 2008; Yang et al., 2016), were excluded in this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

chronostratigraphic framework (Morley et al., 2021; Ma et al., 2022). Hydrocarbon-bearing systems are widely developed on the Sunda Shelf resulting in major oil and gas fields (Leong, 2000; Doust and Sumner, 2007; Zhu et al., 2024). Since the 1960s, extensive oil and gas exploration has been conducted across the shelf, yielding a wealth of drilling data that provide valuable insights into regional OC burial patterns and their climate impact (Doust and Sumner, 2007). Through systematic literature review and data compilation, Ma et al. (2022) first provided a preliminary estimate of OC burial on the Sunda Shelf during the middle Miocene. Thus, there are three key reasons to choose the Sunda Shelf as a case study for testing spatial interpolation methods for deep-time OC burial in epicontinental seas. First, the Sunda Shelf featured an extensive epicontinental sea and accumulated substantial organic-rich sediments during the middle Miocene. Second, OC burial in this region exhibits pronounced spatial and temporal heterogeneity, providing an ideal setting to evaluate the performance and applicability of spatial interpolation methods in deep-time research. Third and equally important, this region offers abundant drilling data and the existing OC burial estimates can serve as a valuable baseline for methodological validation (Ma et al., 2022).

3. Materials and methods

3.1. Data collection and processing

In this study, we conducted a comparative study based on drilling data from the Sunda Shelf basins compiled by Ma et al. (2022). The compiled dataset comprises 367 published drilling sites across multiple shelf basins recording middle Miocene sediments (Fig. 1B). Among them, Malay and Sarawak basins contain the most drillings (137 and 80, respectively), accounting for more than half of the total, whereas the Penyu basin has the smallest area and the fewest drillings. Based on the South East Asia cycles chronostratigraphic framework (the SEA cycles; Morley et al., 2021), the drilling data are roughly assigned to two sedimentary intervals, corresponding to the MCO (SEA 52–53; 16.9–14.6 Ma) and the MMCT(SEA 54,56,58; 14.6–13.8 Ma), respectively (Holbourn et al., 2014, 2015). The compiled dataset includes site-specific attributes (site name, latitude, longitude and basin information) and stratigraphic parameters for each climatic period (sediment thickness, shale or mudstone percentages and TOC contents). Due to limitations in available published literature, not all drillings contain complete datasets of the aforementioned parameters (e.g., shale or mudstone percentages, TOC contents), which is common in this study. To ensure data usability, a conservative data completion method was adopted (e.g., Cartapanis et al., 2016; Ma et al., 2022; Li et al., 2023): for drillings lacking stratigraphic parameters within some basins, the corresponding average values of these basins were applied. While this method may introduce potential uncertainties and obscure local heterogeneity, these limitations are inherent in deep-time spatial interpolation studies and datasets and collecting complete datasets may be the only solution in the future. Drilling site coordinates are unified in the World Geodetic System 1984 (WGS84) (in degrees) to ensure the consistency and accuracy of site location.

Based on these data, we estimated the OC burial of the Sunda Shelf during the MCO and MMCT as follows (Kaya et al., 2022; Ma et al., 2022):

$$M_{OC} = \sum_{i=1}^{n} S_i * (H_{sh})_i * \rho_{sh} * TOC_i \quad (1)$$

Assuming the Sunda Shelf is divided into n polygons (or grid cells), which is primarily determined by the chosen spatial interpolation method and spatial resolution. Here, S_i is the area of a specific polygon (or grid cell), $(H_{sh})_i$ is the thickness of organic-rich shale/mudstone for MCO or MMCT within a specific polygon (or grid cell), and TOC_i is the weighted average TOC content for the specific climate period of that polygon (or grid cell). The average density of organic-rich shale/mudstone, ρ_{sh}, is approximately 2540 kg/m^3. $(H_{sh})_i$ is calculated by multiplying the actual observed strata thickness H with the proportion of shale/mudstone in the strata P_{sh}. TOC_i is obtained by the weighted average of actual TOC measured values and its controlled strata thickness within a specific climate period. For drillings within a specific basin that lack P_{sh} and TOC data, the average P_{sh} and TOC of the basin or block were applied, following Ma et al. (2022). In order to adequately capture

the spatial heterogeneity of OC burial and reduce spatial interpolation complexity, we use areal OC burial mass for specific climate period (MCO or MMCT) (abbreviate as $(OC_{burial})_i$, in Mt./km^2) to replace $(H_{sh})_i$ and TOC_i in this study. This approach simplifies the formula (1) to:

$$(OC_{burial})_i = (H_{sh})_i * \rho_{sh} * TOC_i \quad (2)$$

$$M_{OC} = \sum_{i=1}^{n} S_i * (OC_{burial})_i \quad (3)$$

This variable, $(OC_{burial})_i$, represents the cumulative organic carbon mass buried per unit area within a specific polygon (or grid cell) during the MCO or MMCT period. To facilitate spatial interpolation and regional OC burial estimation, this composite variable was calculated at each drilling site and included in the final dataset used for final interpolation. The original dataset used in this study was sourced from Ma et al. (2022). After further processing, the final dataset can be found in Supplementary data.

3.2. Methods for spatial interpolation

Here, four popular spatial interpolation methods were applied to evaluate the performance of different methods in predicting the spatial distribution of OC_{burial} on the Sunda Shelf: Thiessen polygon, IDW, Kriging and Random Forest machine learning method. To improve the overall accuracy, reliability, and comparability of the spatial interpolation models, we conducted an exploratory analysis of the sample data before interpolating. Moreover, the spatial resolution was set to 0.035° × 0.035° in IDW, Kriging, and RF method for sufficient spatial interpolation detail.

The Thiessen polygon method was used in a previous study (Ma et al., 2022), which divides the studied area into multiple polygons by drawing perpendicular bisectors between sample points (Li and Heap, 2008). Each polygon contains only one sample point, ensuring any point within a subregion is closer to its sample point than to any other (Webster and Oliver, 2007). In this method, the entire study area was divided into 367 polygons based on collected drilling sites, and we calculated the area (S_i) of each polygon controlled by a drilling site. For each polygon, its OC_{burial} came from the corresponding drilling site and was applied in formula (3) to calculate the OC burial.

The Inverse Distance Weighting (IDW) method has been widely used in various fields because of its simplicity and operability: rainfall distribution (Chen and Liu, 2012), soil pollution (Qiao et al., 2018) and so on. It assumes all known sample points could influence any unknown point, with this influence decreasing as distance increases (Li and Heap, 2008). In this study, IDW estimated the OC_{burial} of target points using a linear combination of nearby known values at drilling sites weighted by an inverse function of the distance between them. The performance of IDW largely depends on the type of inverse function, in which the optimal choice of distance decay parameter (P) is the key (Lu and Wong, 2008).

Kriging is a generic name for a family of generalized least squares regression algorithms based on spatial autocorrelation, using linear combinations of data with varying weights for regional prediction (Webster and Oliver, 2007). Spatial autocorrelations are quantified through semivariogram or covariance functions, revealing the spatial dependency among observations. Effective variogram modeling is crucial for the application of Kriging spatial interpolation method (Oliver and Webster, 2015). Moreover, the application of Kriging is most efficient when done on certain characteristics of data distribution: normality and spatial stationarity (Webster and Oliver, 2007). However, the OC burial in marine sediments typically exhibits a strongly skewed, positive distribution due to the spatiotemporal heterogeneity controlled by multiple geological factors (e.g., Burdige, 2007; Bianchi et al., 2018). In order to satisfy the normality assumption of Kriging interpolation functions, the original OC_{burial} data in this study were transformed to a standard normal distribution (mean = 0, variance = 1, Fig. 2) using the quantile transformer function. Then, the QT-transformed OC_{burial} of any target points were interpolated by Ordinary Kriging (OK), the most used variant of Kriging assuming a constant but unknown mean across the study area (Li and Heap, 2008). Finally, the back transformation was applied to get the distribution results of OC_{burial} on the Sunda Shelf and calculated the OC burial.

The Random Forest (RF) method is an ensemble learning technique based on decision trees and bagging, widely used in classification, regression, and other machine learning tasks (Breiman, 2001). It constructs numerous decision trees with high predictive accuracy and low correlation, integrating their results to enhance performance and generalization capability. In spatial interpolation, random forests can fully capture complex nonlinear relationships between target and environment variables, crucial for spatial prediction (Li et al., 2011; Hengl et al., 2015). Some environment information, such as bathymetry, temperature, marine productivity and mud content, have been proved as important covariates for OC burial spatial interpolation in marine sediments (e.g. Lee et al., 2019; Diesing et al., 2021, 2024). However, these variables were excluded from this study due to the challenges in acquiring high-resolution and reliable continuous spatiotemporal data at the local scale for the Sunda Shelf during the Miocene. Nevertheless, assuming the sedimentation boundaries as the paleo-coastline of Sunda shelf during Miocene, the Euclidean distance to the sedimentation boundaries was calculated as the only one environment variable in this study. Previous studies have highlighted the significance of Euclidean distance to the sedimentation boundaries in the spatial interpolation of OC burial in marine sediments (Lee et al., 2019; Diesing et al., 2021). Clay content typically increases with distance from the coastline due to hydrodynamic sorting, and these fine-grained sediments exhibit a notable capacity to adsorb organic matter (Mayer, 1994; Kennedy and Wagner, 2011). A weak positive correlation relationship was also observed in this study, as indicated by the Spearman correlation analysis between OC_{burial} and the Euclidean distance to sedimentation boundaries (Fig. S1 in Supplementary Material). Tobler's first law of geography states that "everything is related to everything else but near things are more related to each other" (Tobler, 1970), a principle that is also applicable in deep-time earth. In addition to the Euclidean distance to sedimentation boundaries, we also considered the coordinates of drilling sites (latitude and longitude), the target variables of nearest sites and their distances to the prediction location as covariates for random forest training and prediction. The optimal number of nearest sites, determined by the degree of spatial correlation and sample size (Sekulić et al., 2020), was set to 25 here. These variables can fully exploit the available spatial information (Hengl et al., 2018; Sekulić et al., 2020).

3.3. Evaluation criteria

In this study, we first applied spatial stratification approach to divide the whole dataset (367 sites) into two parts (Fig. S2 in Supplementary Material): a calibration dataset (329 sites) and a test dataset (38 sites). The test dataset is evenly distributed across the Sunda Shelf, preserving the relative proportions of each internal basin in the whole dataset and ensuring at least one test site per internal basin (Fig. S2 in Supplementary Material). The test dataset was not involved in the training and hyperparameter tuning of the spatial interpolation model, serving an unseen dataset for the final evaluation of each spatial interpolation method. In the calibration dataset, we employed the 10-fold Cross-Validation (CV) method for model training and validating to get the optimal model of each spatial interpolation method. At each modeling phase, one fold from the calibration dataset was selected as the validation set, and the remaining folds were combined and used as the training set for model training. This process was repeated, changing the validation set until all folds have been used as the validation set. The selection and final evaluation of the optimal model in each spatial interpolation method were assessed based on the mean absolute error (MAE) and the

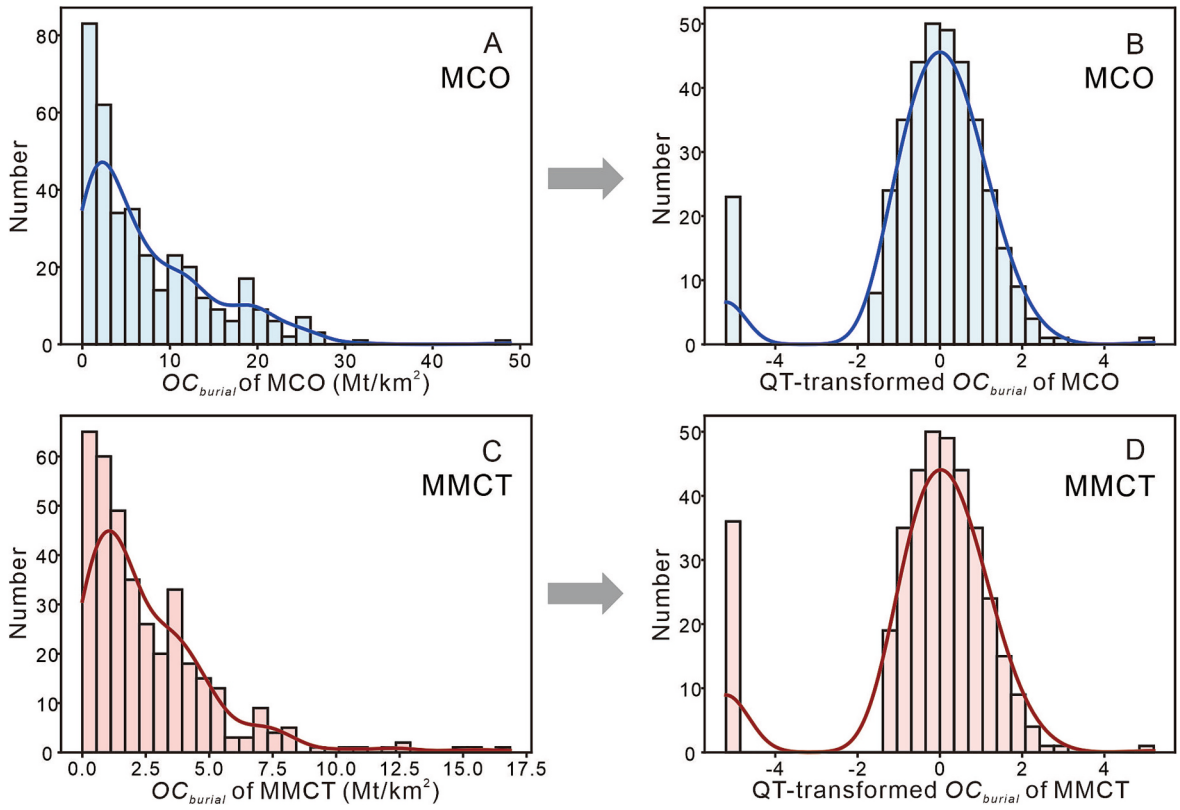


Fig. 2. Histograms and Kernal density estimation distribution of original and QT-transformed OC_{burial}. The original OC_{burial} of MCO (A) and MMCT (C) exhibit significant positively skewed distribution. After applying quantile transformation, QT-transformed OC_{burial} of MCO (B) and MMCT (D) approximates a normal distribution.

root mean square error (RMSE). MAE and RMSE summarize the average discrepancies between observed and predicted values, making them among the common overall measures of model performance (Hodson, 2022). RMSE emphasizes larger errors, making it sensitive to outliers, while MAE treats all errors equally, reflecting overall prediction stability. Lower MAE and RMSE values indicate higher spatial interpolation accuracy and robustness. Additionally, we also evaluated the coefficient of determination (R^2), which measures the proportion of variance in the observed values that is explained by the predicted values. The higher R^2, the stronger the explanatory ability of spatial interpolation model. The formulas for MAE, RMSE, and R^2 are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P(x_i) - O(x_i)| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[P(x_i) - O(x_i)]^2} \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[O(x_i) - P(x_i)]^2}{\sum_{i=1}^{n}[O(x_i) - \overline{O}]^2} \quad (6)$$

Where $P(x_i)$ is the predicted value at point x_i, $O(x_i)$ is the observed value at point x_i, $\overline{O}$ represents the average of observed values for n sample points, and n is the number of samples.

The preliminary data processing, analysis, spatial modeling, and cross validation were conducted interactively through Python (Version 3.10) and ArcGIS Pro (Version 3.0.2).

4. Results

4.1. Statistical characteristics of OC_{burial}

The OC_{burial} serves as the primary spatial interpolation target variable in this study, with its descriptive statistical results of the whole Sunda shelf and individual basins presented in Table 1. Descriptive statistics show that the MCO period yielded higher OC_{burial} levels across the Sunda Shelf compared to the subsequent MMCT period. During the MCO period, the Sunda Shelf OC_{burial} ranged from 0 to 48.8482 Mt./km^2, with a mean of 7.4908 Mt./km^2 and a standard deviation of 7.2633 Mt./km^2. During the MMCT period, the Sunda Shelf OC_{burial} ranged from 0 to 16.8531 Mt./km^2, with a mean of 2.5832 Mt./km^2 and a standard deviation of 2.5649 Mt./km^2. Coefficient of variation (CV) is defined as the ratio of standard deviation to the mean, which is a normalized measure of the data variability (Brown, 1998). During the MCO and MMCT period, the CV of OC_{burial} across these internal basins varied from 22.63 % to 100.57 %, and for the whole shelf it was 96.96 % and 99.29 %, respectively. The distribution of OC_{burial} across the Sunda Shelf also exhibited significant dispersion and positive skewness distribution characteristics where the skewness was greater than 1 and kurtosis was greater than 2 (MCO: 1.383 and 2.570; MMCT: 2.072 and 6.430) (Fig. 2A and C). These statistical characteristics are common in deep-time geoscience data due to the complexity and spatiotemporal heterogeneity in earth system (e.g., Wellmer, 1998; Reimann and Filzmoser, 2000). After applying quantile transformation, the QT-transformed OC_{burial} approximates a normal distribution but with a pronounced peak on the left tail (Fig. 2B and D). This abnormal deviation is likely attributable to the presence of numerous extreme low values in the original dataset, which results in an aggregation of data points in the lower range after quantile transformation (Hastie et al., 2009).

Table 1
Descriptive statistics of original OC_{burial} (Mt/km^2).

Stage	Basin	N	Min.	Max.	Mean	SD	CV(%)	Skewness	Kurtosis
MCO (16.9–14.6 Ma)	All Basins	367	0	48.8482	7.4908	7.2633	96.96	1.383	2.570
	Pattani	41	0	3.1820	0.9071	0.9123	100.57	1.099	0.340
	Malay	137	0	31.5811	12.5153	6.9750	55.73	0.223	−0.505
	Penyu	7	4.0287	7.7399	5.5906	1.2651	22.63	0.669	−0.017
	West Natuna	28	0	6.0141	2.4137	1.5030	62.27	0.276	0.363
	Wan'an	37	0	15.5191	4.7019	3.6915	78.51	1.328	1.530
	Mekong	37	0.5048	2.3133	1.5840	0.4835	30.52	−0.446	−0.180
	Sarawak	80	0	48.8482	8.2255	7.5341	91.59	2.525	9.928
MMCT (14.6–13.8 Ma)	All Basins	367	0	16.8531	2.5832	2.5649	99.29	2.072	6.430
	Pattani	41	0	3.4070	0.7996	0.7479	93.53	1.382	2.379
	Malay	137	0	16.8531	3.9562	3.1187	78.83	1.282	3.082
	Penyu	7	1.8311	3.6866	2.5645	0.6242	24.34	0.912	0.672
	West Natuna	28	0	2.7822	1.3786	0.7154	51.89	−0.613	0.295
	Wan'an	37	0.1355	4.8124	2.0100	1.1399	56.71	0.445	−0.082
	Mekong	37	0.2883	1.3483	0.9242	0.2821	30.52	−0.484	−0.073
	Sarawak	80	0	14.7499	2.6019	2.3962	92.09	2.236	7.606

Notes: N = number of samples; Min. = minimum; Max. = maximum; SD = standard deviation; CV = coefficient of variation.

The spatial trend of OC_{burial} remained essentially consistent between the MCO and MMCT periods (Fig. 3), with generally higher in the Malay and Sarawak basins, and relatively lower in the Pattani, Mekong and West Natuna basins (Table 1). This consistent spatial pattern was also reflected in sediment thickness, with the greatest mean thickness found in the Sarawak and Wan'an basins, while the West Natuna and Mekong basins consistently exhibited the lowest means (Text S1 in Supplementary Material). During the MCO period, the mean OC_{burial} was about 12.5153 Mt./km^2 in the Malay Basin and 8.2255 Mt./km^2 in the Sarawak Basin. They all decreased by ~70 % to 3.9562 Mt./km^2 in the Malay Basin and to 2.6019 Mt./km^2 in the Sarawak Basin during the MMCT period, but still exhibited high variability (Table 1). In other basins, the mean OC_{burial} was all also decreased from the MCO to MMCT period, but spatial variability within these basins remained substantial with moderate to high CV (Table 1). This indicates the significant inherent spatial heterogeneity within internal basins or across the shelf as a whole. Moreover, within each internal basin on the Sunda Shelf, the spatial distribution of OC_{burial} displayed a conservative trend of higher value at the center and lower at the edge (Fig. 3). This spatial trend was further supported by Spearman correlation analyses between OC_{burial} and the Euclidean distance to sedimentation boundaries (Fig. S1 in Supplementary Material), although positive correlations were weak during both the MCO ($\rho = \sim 0.202$, $p < 0.01$) and MMCT ($\rho = \sim 0.121$, $p < 0.05$) periods.

4.2. Hyperparameters optimization for OC_{burial} spatial interpolation

Selecting and optimizing model hyperparameters is crucial for achieving the best interpolation performance among different spatial interpolation methods mentioned in Section 3.2 (Li and Heap, 2008; Table 2). Each hyperparameter set was iteratively evaluated using the 10-fold CV method to identify the set with the smallest RMSE, which represents the optimal interpolation performance for each method (e.g., Kim et al., 2022). For the Thiessen polygon spatial interpolation, we

Table 2
Hyperparameter sets used for each spatial interpolation method in this study to obtain the optimal interpolation results.

Methods	Hyperparameter	Range
Thiessen	–	–
IDW	P	1,2,3,4
OK	Model	Exponential, Spherical, Gaussian, Linear and Circular
RF	n_trees、max_features	n_trees (50–500, 50 step size), max_features (5–50, 5 step size)

Notes: P = Distance decay parameter; Model = Model of semivariogram function; n_trees = number of trees; max_features = the maximum number of features selected in each split.

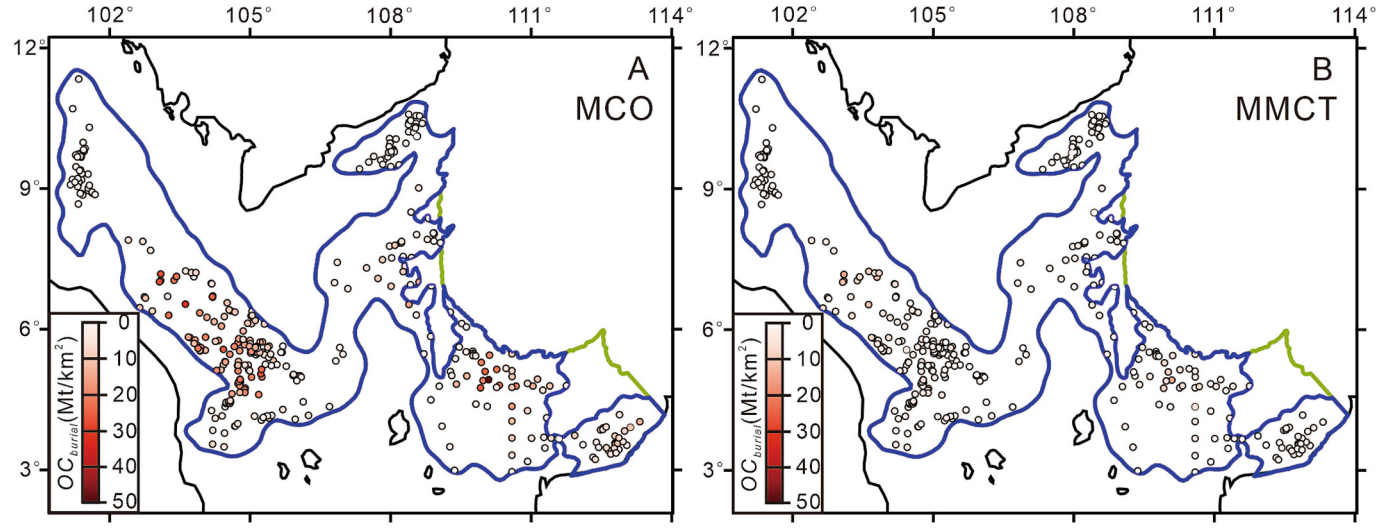


Fig. 3. Spatial distribution of OC_{burial} on the Sunda Shelf during MCO (A) and MMCT (B) periods. The OC_{burial} values range from 0 to 48.85 Mt./km^2 in the MCO (A) and from 0 to 16.85 Mt./km^2 in the MMCT (B).

adopted a similar interpolation method and assumptions as Ma et al. (2022) to set up a control experiment for other methods comparison (Fig. 4A and B). The distance decay parameter, as one of the most crucial parameters in the IDW model, directly affects how neighboring points influence unknown points. A higher distance decay parameter increases the weight of nearby points, whereas a lower one gives more weight to distant points (Lu and Wong, 2008). In the IDW spatial interpolation, distance decay parameters of 1, 2, 3, and 4 were validated (Table 2), with the optimal parameters determined to be 2 for the MCO and 1 for the MMCT period (Table S2 in Supplementary Material). This parameter changes also coincided with the changes in spatial autocorrelation of OC_{burial} from the MCO to MMCT period (Fig. 3), as reflected in Moran's I values of 0.63 (z-score = 16.60) for MCO and 0.48 (z-score = 17.66) for MMCT.

In the OK spatial interpolation, we assessed the impact of different semivariogram model selections on the spatial interpolation results: exponential, spherical, Gaussian, linear, and circular models (Table 2). Variogram model parameters were auto-tuned via ArcGIS Pro Kriging function (Johnston et al., 2001). During the MCO period, the exponential semivariogram model achieved the lowest MAE (~3.8621) and RMSE (~6.1117), making it the optimal choice, although the differences between models were small (Table S3 in Supplementary Material). During the MMCT period, the evaluation results for all five semivariogram models were nearly identical (MAE = ~1.356, RMSE = ~ 2.336; Table S3 in Supplementary Material). Here, the spherical model was selected as the optimal semivariogram model for the MMCT period, in consideration of its stable performance and widespread use in shelf sediment interpolation (e.g., Webster and Oliver, 2007; Li and Heap, 2008). Moreover, the R^2 values of OK during both the MCO and MMCT periods were relatively low (<0.3).

In the RF spatial interpolation, two critical parameters, the n_trees and max_features significantly influence model performance and generalization (Houborg and McCabe, 2018). Herein, we employed the Grid Search method to find the optimal combination of n_trees and max_features (Table 2; Fig. S4 in Supplementary Material). Other hyperparameters, such as max_depth, min_samples_split and others, were set to their default values as implemented in the scikit-learn package in Python. During the MCO period, the optimal settings were 100 n_trees and 20 max_features (MAE = ~2.8415, RMSE = ~4.5115, R^2 = ~0.6084); for MMCT, the optimal settings were 400 n_trees and 10 max_features (MAE = ~1.2128, RMSE = ~1.9390, R^2 = ~0.4383). All three spatial interpolation methods (IDW, OK and RF) exhibited consistently lower prediction accuracy (MAE and RMSE) but higher spatial explanatory power (R^2) during the MCO compared to the MMCT period (Table S2–4 in Supplementary Material).

The optimized spatial interpolation results of OC_{burial} for each method (Thiessen, IDW, OK and RF) are presented in Fig. 4. All methods interpolated similar OC_{burial} spatial patterns for both the MCO and MMCT periods, particularly highlighting the Malay and Sarawak basins as principal areas of OC burial. The value ranges and distributions of interpolated OC_{burial} were consistent with the original data characteristics (Table 1; Fig. 3).

4.3. Comparative performance of traditional and machine learning spatial interpolation methods

Following the optimal spatial interpolation results presented in Section 4.2 (Fig. 4), the performance of traditional (Thiessen, IDW and OK) and machine learning (RF) spatial interpolation methods was quantitatively evaluated using an independent unseen test dataset (Fig. 5). The evaluation metrics included the mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2) (see Section 3.3). Among them, the Thiessen method, as a control experiment, consistently exhibited the poorest performance, with the highest MAE and RMSE values and the lowest R^2 value during both the MCO and MMCT periods. Notably, it even had a negative R^2 value during the MMCT period. In contrast, the IDW, OK and RF methods demonstrated significantly better performance than the Thiessen method while demonstrating similar predictive accuracies on the unseen test dataset (Fig. 5). Specifically, during the MCO period, the OK method slightly outperformed the others, presenting the lowest errors (MAE = ~2.9576, RMSE = ~4.8430) and highest R^2 (~0.5967). For the MMCT period, the RF method achieved marginally better performance on the unseen test dataset, with the lowest errors (MAE = ~1.1114, RMSE = ~1.6037) and the highest R^2 (~0.5246).

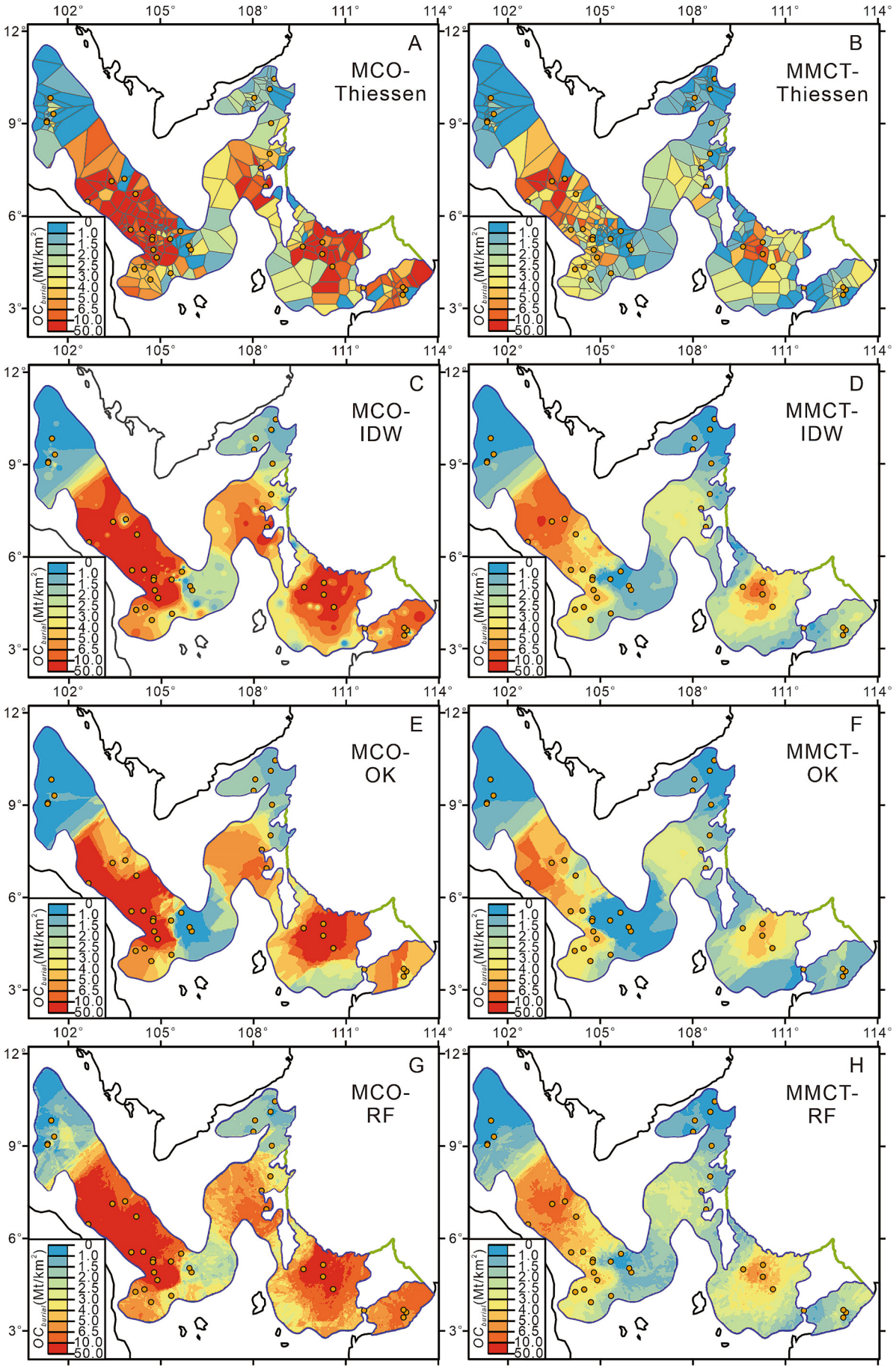
Additionally, a visual examination of interpolation results was further conducted to identify anomalous interpolation patterns and inspect spatial interpolation performance (e.g., Li et al., 2011; Chen et al., 2016; Maxwell et al., 2021). Although quantitative evaluation metrics exhibited similar interpolation accuracies among the IDW, OK and RF methods, visual examination of interpolation results still revealed notable spatial differences and abnormal results. Notably, the four interpolation results consistently revealed a notable and similar abnormal demarcation of OC_{burial} between the Pattani and Malay basins in both periods (Fig. 4), coinciding with regions of sparse drilling coverage (Fig. 3). The Thiessen method exhibited clear spatial discontinuities and abrupt transitions, particularly evident near internal basin boundaries and areas with sparse drilling coverage (Fig. 4A and B). In contrast, the IDW, OK and RF methods generated relatively more continuous and smoother spatial transitions. However, notable differences still existed among these three methods. Specifically, the IDW method displayed pronounced "bull's-eye" artifacts characterized by abrupt circular transitions around individual high-value or low-value data points, particularly noticeable during the MCO period (Fig. 4C and D). The OK method produced generally smooth maps but exhibited unnatural zoning patterns and distinct spatial demarcations, especially during the MMCT period (Fig. 4E and F). Comparatively, the RF method presented the most naturally continuous spatial interpolation results (Fig. 4G and H), effectively mitigating the abnormal zone observed in the OK maps and the localized bull's-eye artifacts present in IDW. Moreover, in the visual examination, RF captured more subtle spatial variations and transitions within internal basins, with gradual gradations of OC_{burial} from basin margins towards basin centers (Fig. 4G and H).

5. Discussion

5.1. Advantages of machine learning in deep-time spatial interpolation?

Quantitative evaluations on the unseen test dataset (Section 4.3) indicated broadly similar predictive accuracy among the IDW, OK and RF methods, with only minor performance differences observed (Fig. 5). Nevertheless, this similarity in performance should be interpreted cautiously, considering potential bias and uncertainty arising from the evaluation on a single pre-divided independent test dataset (Table S2 in Supplementary Material). If the optimal hyperparameters are predetermined and a cross-validation approach is applied to evaluate model average performance on multiple unseen datasets (e.g., Li et al., 2011), the performance of OK observed in the single-dataset evaluation may diminish, whereas IDW and RF could consistently outperform OK during both climate periods (Tables S2–S4 in Supplementary Material). Moreover, visual examination clearly demonstrated that the RF method generated notably smoother and more natural spatial distributions compared to traditional spatial interpolation methods (Thiessen, IDW, and OK; Fig. 4). Such quantitative and qualitative assessments tentatively suggest potential advantages of RF for interpolating deep-time organic carbon burial in epicontinental seas.

Traditional methods (Thiessen, IDW, and OK) simplify and assume data distribution pattern based on the spatial relationships of sample points by mathematical or statistical models (Webster and Oliver, 2007; Li and Heap, 2008). In this study, the Thiessen method consistently showed the poorest performance on the unseen test dataset during both



(caption on next page)

Fig. 4. Spatial interpolation maps of OC_{burial} by Thiessen method (A, B), IDW (C, D), OK (E, F), and RF (G, H). Yellow points indicate the unseen test dataset used for the generalization evaluation of each spatial interpolation method. The left panels (A, C, E, G) represent the MCO climate period, while the right panels (B, D, F, H) correspond the MMCT climate period. The spatial interpolation maps of different methods exhibit similar spatial trends within the same climatic period. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the MCO and MMCT periods. Specifically, for the MMCT period, the Thiessen method yielded even worse performance with a negative R^2 (~−0.1524), suggesting its inability to capture the spatial variability of OC_{burial} effectively. The effective application of Thiessen method is closely tied to the spatial distribution of data points, with significant uncertainty and risk in representing an area with a single site in sparsely sampled regions (Li and Heap, 2008), like the northern Malay Basin and the West Nutuna Basin (Fig. 1). Particularly in complex and variable deep-time geological settings, the Thiessen method may fail to capture spatial characteristics based on limited deep-time geological data. Moreover, IDW interpolation results also generated unusual bull's-eye patterns that were especially in MCO period (Fig. 4C and D). This phenomenon arises from the high sensitivity of IDW to extreme data values, a well-documented limitation in numerous case studies (Li et al., 2011; Chen et al., 2016). Likewise, OK interpolation was sensitive to the statistical characteristics and spatial distribution patterns of data points (Li and Heap, 2008). Although OK demonstrated slightly better performance than RF on the unseen test dataset during the MCO period, RF showed a clear advantage during the MMCT period, outperforming OK in both evaluation metrics and the ability to capture finer spatial details and features in the visual examination (Figs. 4 and 5). The observed differences in performance between OK and RF across the two periods can be attributed to the variability of spatial autocorrelation. During the MCO period, OC_{burial} exhibited stronger spatial autocorrelation (Moran's I value = ~0.63, z-score = 16.60), enabling OK to perform well by effectively capturing spatial continuity through the variogram. However, the spatial autocorrelation weakened during the MMCT period (Moran's I value = ~0.48, z-score = 17.66), potentially limiting the effectiveness of OK. Furthermore, a pronounced Pure Nugget Effect was also observed during the MMCT period (partial sill = ~0.01, nugget = ~3.6573), indicating that spatial variability was predominantly driven by random noise or unobserved micro-scale processes that Kriging failed to capture (Webster and Oliver, 2007). These observations underscore the limitations of OK, which relies on specific statistical assumptions and spatial structures that are often difficult to meet in deep-time geological datasets, even after relevant transformations (Fig. 2).

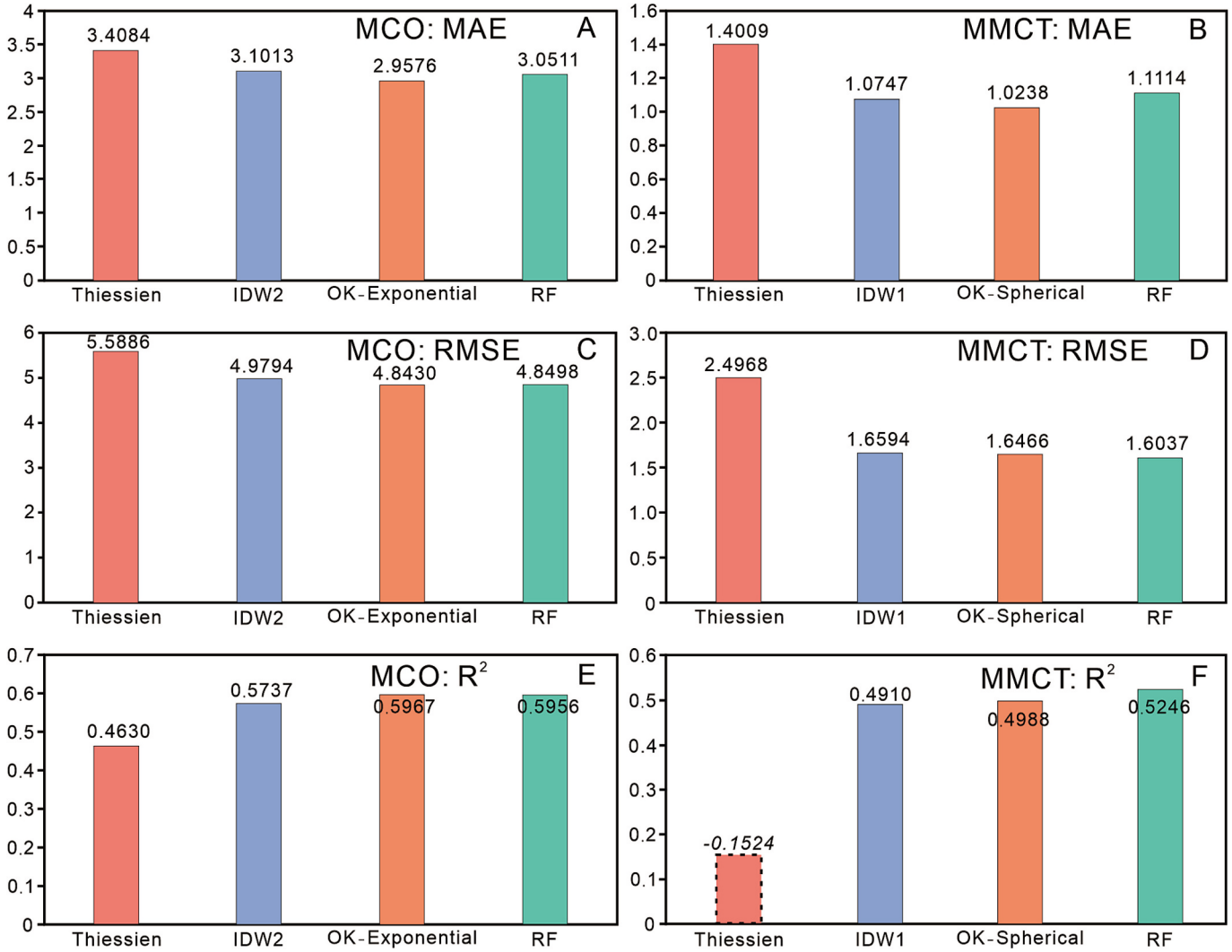


Fig. 5. Comparative performance of four spatial interpolation methods based on the unseen test dataset (Fig. S2 in Supplementary Material). IDW2 indicates that the distance decay parameter is 2, while IDW1 indicates it is 1. OK-Exponential refers to the model using an exponential semivariogram, and OK-Spherical refers to the model using a spherical semivariogram.

Emerging machine learning methods, notably Random Forest, are sufficiently flexible and require fewer assumptions (Breiman, 2001). These methods can reveal complex non-linear patterns in natural processes and have been widely applied in environmental science, marine ecology and other related fields (Hengl et al., 2018; Sekulić et al., 2020). Unlike OK, as a data-driven approach, RF is less constrained by assumptions of spatial autocorrelation or data stationarity, allowing it to flexibly capture complex spatial patterns even under weak autocorrelation or high heterogeneity conditions (Li et al., 2011; Maxwell et al., 2021). In this study, our RF model incorporated not only spatial coordinates but also auxiliary features such as distances to sedimentation boundaries and neighboring point values, thereby capturing subtle spatial heterogeneity across both climatic intervals (Fig. 4G and H). Despite obtaining deep-time continuous spatiotemporal OC burial related environmental information (such as paleodepth, paleotemperature, and paleoproductivity) on the Sunda Shelf remains challenging, the RF model can be readily updated and refined as new data become available. Based on the mean decrease in impurity of these auxiliary variables used in RF, the values of nearby points were identified as key features in predicting OC_{burial} during both MCO and MMCT periods (Fig. S5 in Supplementary Material). This feature is also reflected in IDW results with similar evaluation performance to RF (Fig. 5), further emphasizing the importance of nearby point values (Lu and Wong, 2008; Sekulić et al., 2020). However, IDW relies on a globally fixed distance-decay parameter, which may overlook localized spatial patterns and requires prior parameter selection (Lu and Wong, 2008). By dynamically assessing the importance of neighboring points (Fig. S5 in Supplementary Material), RF overcomes these constraints and more effectively captures fine-scale spatial variability without rigid distance-decay assumptions (e.g., Hu et al., 2022). This flexibility helps explain why RF produced smoother transitions with fewer interpolation artifacts (Fig. 4G and H), especially during the MMCT period when spatial autocorrelation was weaker. Although the values of nearby points were the most important predictors, their distances to the target points were less influential within the RF model (Fig. S5 in Supplementary Material). This may be related to the obvious cluster patterns in spatial distribution during both two climate periods, which still needs further investigation.

Notably, deep-time geological data often exhibit strong spatiotemporal heterogeneity, post-depositional alteration and non-Gaussian distributions, which pose substantial challenges for traditional spatial interpolation methods. RF can address these challenges and optimize predictions by leveraging ensemble decision trees and maximizing the use of all available information, even with spatially sparse and uneven data (Breiman, 2001; Li et al., 2011). Additionally, RF model allows for improved scientific interpretability and expert evaluation through variable importance assessments (Breiman, 2001; Cutler et al., 2007). In terms of uncertainty estimation, the unique model structure of RF allows for low bias and variance through averaging results from multiple weakly correlated trees and quantifying the prediction uncertainty range based on the results of multiple decision trees (Breiman, 2001; Coulston et al., 2016; Graw et al., 2021). While OK could provide uncertainty estimates for spatial interpolation, it still depends on assumptions of stationarity and well-defined variogram structures, which may not hold in deep-time geological studies. Conversely, IDW and Thiessen, as deterministic spatial interpolation methods, cannot directly quantify the uncertainty range of spatial interpolation (Li and Heap, 2008). They often obtain relative uncertainty ranges through quantile calculations in previous study (e.g. Kaya et al., 2022; Ma et al., 2022), but it mainly reflects the variability of the observed data.

Overall, these findings highlight the practical advantages of RF over traditional spatial interpolation methods for interpolating deep-time organic carbon burial in epicontinental seas by taking the Sunda Shelf as a case study. However, further work is required to validate the broader applicability and robustness of RF in deep-time spatial interpolation. For example, it remains unclear whether RF and similar machine learning methods can maintain their superiority on the deep-time global-scale interpolations, despite there are many applications in modern ocean (e.g., Lee et al., 2019; Diesing, 2020; Graw et al., 2021). Moreover, integrating geological models (such as depositional environments, tectonic settings, and sediment dynamics) into data-driven machine learning methods remains a key challenge for improving the objectivity and geoscientific interpretability of spatial predictions (Reichstein et al., 2019; Xu et al., 2021a).

5.2. Spatial-temporal patterns of OC Burial on the Sunda Shelf and controls

As discussed above, the RF method showed superior performance in this study. To fully utilize the collected data and improve the accuracy of the RF method, we re-optimized the hyperparameters on the whole dataset by 10-fold CV method (Fig. S6 in Supplementary Material). Based on the RF method and Eq. (3), the spatial-temporal distribution of OC burial mass on the Sunda Shelf during the MCO and MMCT periods was refined (Fig. 6).

During the MCO period, massive OC were deposited across most internal basins of the Sunda Shelf, especially in the Malay, Penyu, Wan'an, and Sarawak basins (Fig. 6A). This distribution pattern was broadly consistent with known areas of deep rifting systems and marine productivity under warm, high-stand sea-level conditions (Morley and Morley, 2013). During the MMCT, the hotspot areas of OC burial ($\geq$90 Mt) contracted to central regions of the Malay and Sarawak basins, while low-value areas expanded (Fig. 6C). Notably, the Pattani and Mekong basins exhibited low OC burial values during both the MCO and MMCT periods (Fig. 6A and C), suggesting persistently poor preservation conditions for OC in these regions. Previous studies have identified these areas as shallow rift systems, which were prone to seasonal flooding with fluvial sediments and lacked sustained anoxic conditions (Shoup et al., 2012; Morley and Morley, 2013). Despite the shifts in the extent of high-value areas for OC burial across the two climate periods, most basins still exhibited a mass gradient from the basin centers to the margins (Fig. 6A and C). Similar spatial patterns were also exhibited in the spatial distribution of sediment accumulation during both climate periods (Text S2 and Fig. S8 in Supplementary Material). The recalculated total OC burial masses for the two climate periods on the Sunda Shelf are larger than the median estimates by Ma et al. (2022) yet remain within their interquartile ranges (Table S6 in Supplementary Material). Specifically, the total OC burial on the Sunda Shelf during the MCO was $2.73 \pm 1.94 \times 10^3$ GtC, accounting for about 9.62 % of global net marine OC burial; during the MMCT period, it was $1.14 \pm 0.78 \times 10^3$ GtC, approximately 10.01 % of global net marine OC burial (Derry and France-Lanord, 1996). The Sunda Shelf played a crucial role in global marine OC burial during the middle Miocene. However, the burial efficiency during the MMCT exceeded that of the MCO relative to global net OC burial, as also reflected in the rates of total OC burial on the Sunda Shelf during these two climate periods (Fig. 7). Notably, a similar temporal pattern of OC burial efficiency is observed in the Monterey Formation of the Santa Barbara Basin, suggesting analogous controlling factors (e.g., see level change and sediment supply; Anttila et al., 2023). During the MMCT period, the OC burial rate increased by approximately 0.24×10^3 GtC/Myr compared to the MCO, corresponding to an extra of ~115.48 ppm/Myr of pCO_2 sequestration (Fig. 7), which is about twice the previous estimate (Ma et al., 2022). In terms of sediment accumulation rate on the Sunda Shelf, the MMCT period also came to the dominate, with a burial rate roughly twice that of the MCO (Fig. 7B and D).

Spatial analysis of the OC burial rate changes across the Sunda Shelf from the MCO to MMCT demonstrated that not all locations experienced an increase during the MMCT (Fig. 7A). Elevated OC burial rates during the MMCT period were primarily concentrated in drainage systems (deltas or river mouths) along the margins of the Sunda Shelf (e.g., northern Malay Basin, Penyu Basin, northwestern Wan'an Basin and southwestern Sarawak Basin; see fig. 8 in Morley et al., 2016). In

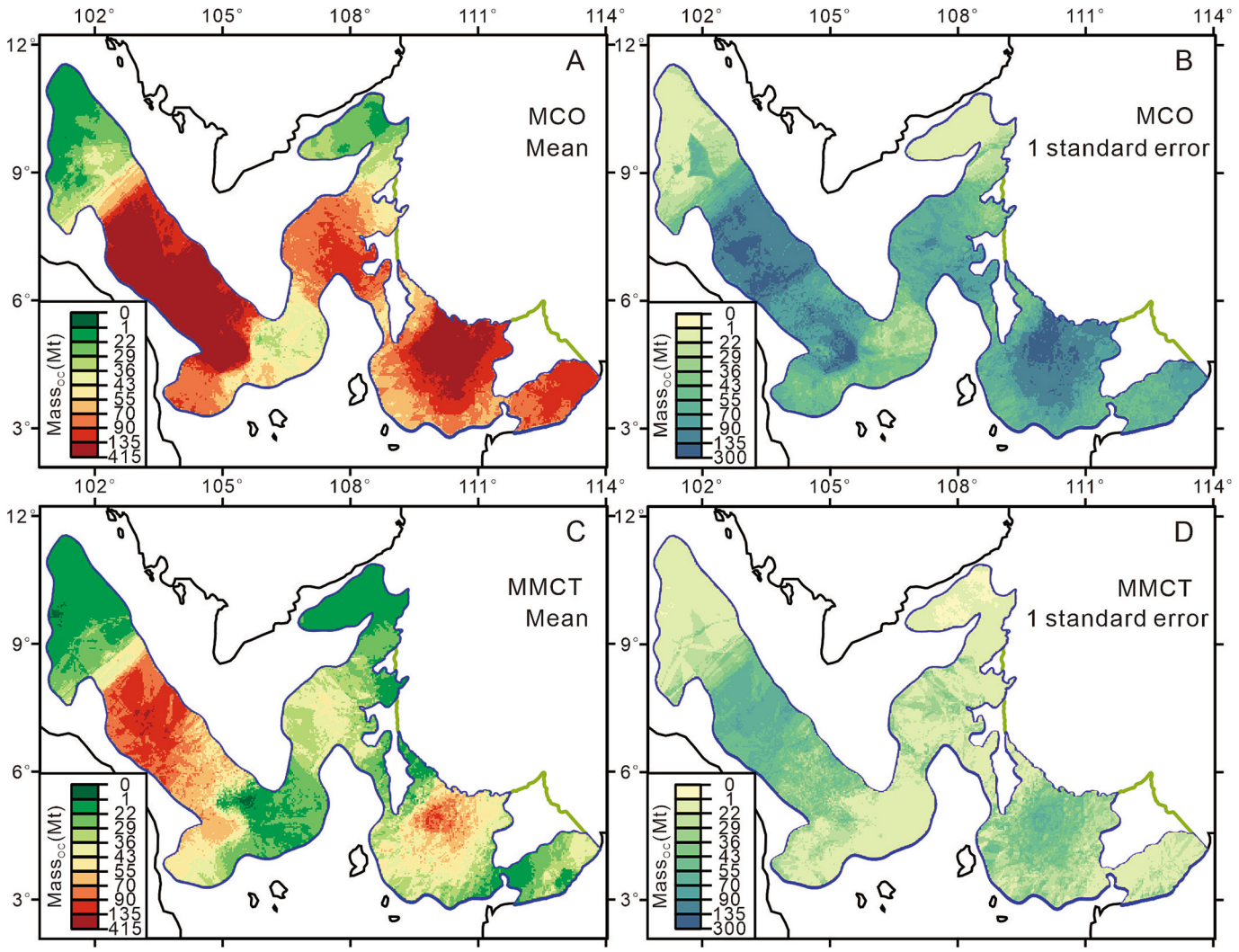


Fig. 6. Spatial distribution of OC burial mass and its one standard error during the MCO (A and B) and MMCT (C and D) periods based on RF method. The OC burial mass and its standard error at any location on the Sunda Shelf were derived from all the predictions made within the forest, following the method described in Graw et al. (2021).

contrast, in basin regions located distal to the drainage systems (e.g., southern Malay Basin, eastern Wan'an and Sarawak Basin), OC burial rate decreased during the MMCT period. However, this spatial heterogeneity contrasts with the general increase in sediment accumulation rates (SAR) across the Sunda Shelf during the MMCT period compared to the MCO (Fig. 7B). The broadly observed acceleration in sediment accumulation during the MMCT period may reflect an increase in terrigenous sediment supply, potentially associated with regional sea-level fall (Miller et al., 2020; Morley et al., 2021), active local tectonic activities (Morley et al., 2016), and increased monsoonal runoff (Nie et al., 2018; Ma et al., 2023). The gradual sea-level falling during the MMCT period also likely enhanced terrigenous organic matter input via expanded drainage systems and terrestrial vegetation (Morley et al., 2016), boosting overall OC burial rates despite the spatial heterogeneities existed (Fig. 7). Although we cannot obtain the high-resolution spatial pattern of organic matter type changes across the shelf, the changes of organic matter type in many individual basins and the widespread existence of terrigenous pollen support such characteristics (Morley et al., 2021; Ma et al., 2022). These findings lead confidence that there may be complex relationships between OC burial, sediment accumulation and terrigenous organic matter input across the Sunda Shelf, resulting in spatial-temporal differences in response to climate change during the middle Miocene.

The conceptual model of middle Miocene OC burial on the Sunda Shelf in Ma et al. (2022) focused primarily on the time-scale response of OC burial and sediment accumulation rates to climate and environmental change, while overlooking spatial-temporal discrepancies and patterns. Here, from a higher spatial resolution perspective, we propose a possible explanation for the spatial-temporal patterns in OC burial rate changes across two climate periods on the Sunda Shelf. At the basin margins within the Sunda Shelf, the gradual sea-level falling promoted expansion of drainage systems and terrestrial vegetation during the MMCT period (Morley et al., 2021), leading to the sufficient terrigenous organic matter input and high OC burial rate under higher SAR (yellow-red areas, Fig. 7A and B). As sediments dispersed basinward and hydrodynamic conditions gradually weakened, prolonged oxygen exposure reduced OC burial efficiency and rates significantly despite higher overall sediment supply (Fig. 7A and B). Therefore, the spatial-temporal patterns of OC burial rate on the Sunda Shelf from the MCO to MMCT period may reflect the local coupling relationship between sedimentation rates and OC burial efficiency in respond to different climate and sea-level conditions. But from the perspective of the Sunda Shelf as a whole, the Sunda Shelf during the MMCT was still characterized by an overall enhanced OC burial rate (Fig. 7D), contributing to further global cooling and *p*CO2 reduction through positive climate feedback (Ma et al., 2022).

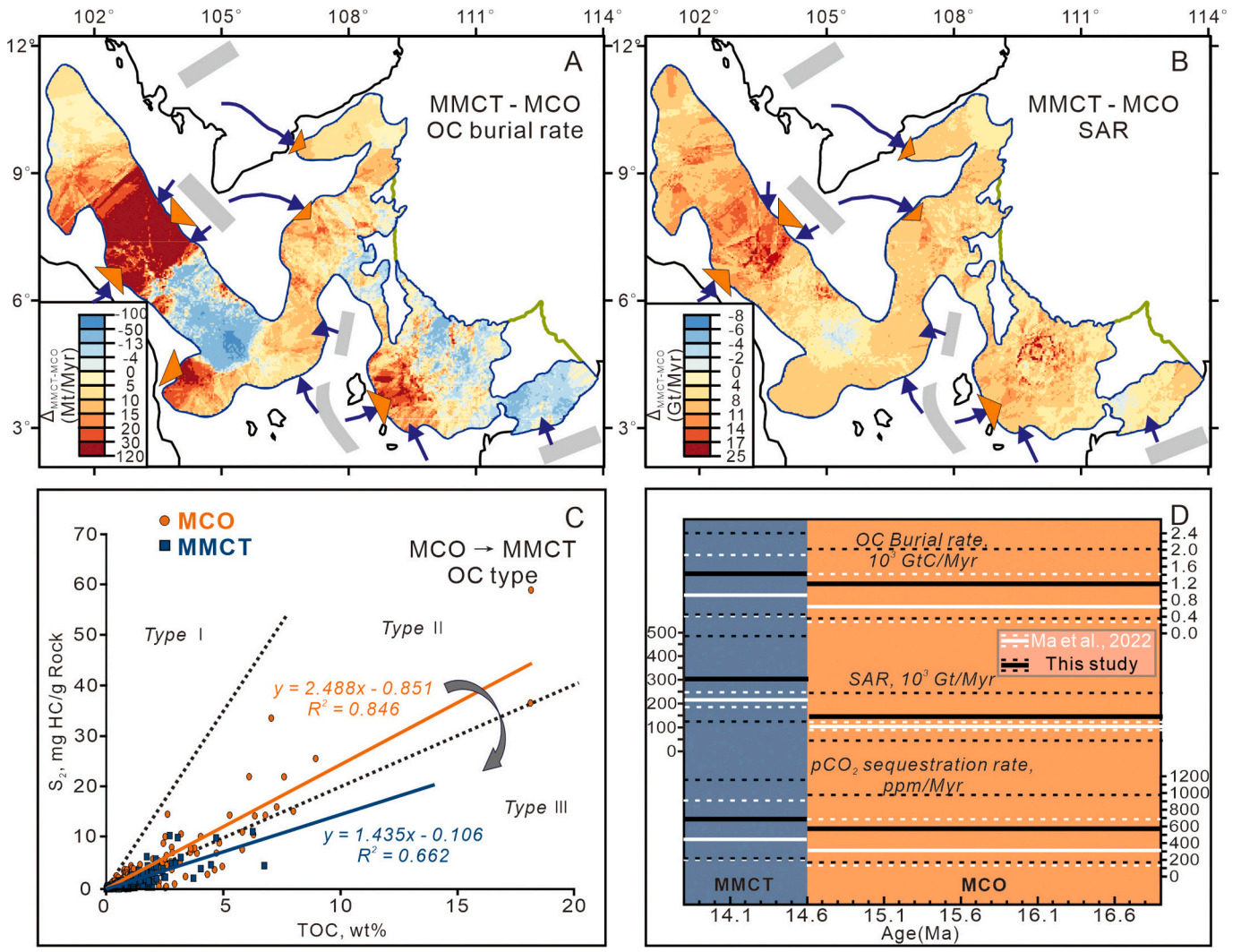


Fig. 7. Spatial-temporal patterns of OC burial and sediment accumulation on the Sunda Shelf during the middle Miocene. (A) and (B) show the spatial distribution changes of OC burial rates and sediment accumulation rates (SAR) across the shelf from the MCO to MMCT period. Gray rectangles indicate the possible high lands, while purple arrows and orange triangles represent the potential rivers and fan systems on the Sunda Shelf during the MCO and MMCT periods (see fig. 8 in Morley et al., 2016). (C) Organic type changes across the Sunda Shelf from the MCO to MMCT period based on TOC and S_2 (Ma et al., 2022). The Gray arrow indicates that the increased terrigenous OC input from the MCO to MMCT period. (D) Temporal changes in OC burial rate, sediment accumulation rate (SAR) and pCO_2 sequestration rate across the Sunda Shelf during MCO and MMCT periods. White lines represent the estimates from Ma et al. (2022), black lines are results calculated based on RF method in this study (see also table S4 in Supplementary Material). Solid lines indicate the mean rate, while dashed lines represent one standard error (this study) or the 25th/75th percentile values (Ma et al., 2022). The calculation method of CO_2 sequestration rate follows the method provided in Ma et al. (2022). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6. Conclusions and perspectives

Quantifying the total amount and spatiotemporal distribution patterns of OC burial in epicontinental seas is a crucial step for reconstructing the global carbon cycle and understanding climate evolution. Although several spatial interpolation methods have yielded promising results in estimating deep-time OC burial, comparative analyses evaluating their applicability and uncertainty remain scarce. In this study, utilizing 367 drilling data points from the middle Miocene Sunda Shelf collected by Ma et al. (2022), we quantitatively and qualitatively evaluated the performance of four prevalent spatial interpolation methods: Thiessen polygons, IDW, OK and RF. We assessed their performance using an unseen test dataset with three key evaluation metrics: MAE, RMSE, and R^2. The results demonstrated that the Random Forest machine learning method outperformed the others, showing high predictive accuracy and detailed spatial variations. This superiority is attributable to the data-driven advantage of RF. RF exhibited high generalizability and robustness, which need fewer assumptions and can effectively capture nonlinear relationships and complex spatial patterns in non-Gaussian deep-time geological data. Using the RF method, we also refined the spatial distribution pattern of OC burial on the Sunda Shelf during the middle Miocene. The estimated total OC burial for the MCO period was about $2.73 \pm 1.94 \times 10^3$ GtC, and about $1.14 \pm 0.78 \times 10^3$ GtC for the MMCT period. However, the MMCT period showed a higher OC burial rate than the MCO period, further supporting the conclusions of Ma et al. (2022). The analysis of spatial-temporal patterns in OC burial and sediment accumulation rates on the Sunda Shelf revealed a localized coupling between sedimentation rates and OC burial efficiency in respond to different climate and sea-level conditions. Overall, our case study shows that spatial-temporal analysis of OC can enhance the understanding of the burial history in epicontinental seas in relation to geological events.

It is critical to note that the choice of spatial interpolation method, alongside the research objectives and data characteristics involves a very wide range. This paper is limited to a comparative analysis of four prevalent methods applied to deep-time OC burial estimate in

epicontinental seas. There is no universally optimal spatial interpolation method, the best approach depends on the unique objectives and conditions specific to the dataset of each research area (Li and Heap, 2008). However, RF have demonstrated superior performance in spatial interpolation tasks, attributable to their inherent methodological strengths (Li et al., 2011; Hengl et al., 2018). Despite there are many challenges in data collection, uneven spatiotemporal distribution, and the absence of environmental auxiliary variables, the adaptability and robust data processing capabilities of machine learning methods still position them as promising tools for deep-time spatial interpolation studies. Furthermore, the growing availability of various proxies and model-derived gridded datasets in the big data era, provides a solid data foundation for the application of machine learning in deep-time spatiotemporal interpolation (Wang et al., 2021; Xu et al., 2021a; Vance et al., 2024). A more comprehensive reconstruction of the environmental evolution using machine learning (e.g., Dutkiewicz et al., 2017; Wang et al., 2023), could help reveal the mechanisms and environmental evolution patterns of the Earth system change in the past and future.

CRediT authorship contribution statement

Yida Yang: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Pengfei Ma:** Writing – review & editing, Validation, Supervision, Resources, Investigation, Funding acquisition, Data curation, Conceptualization. **Xiumian Hu:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Yuan Gao:** Writing – review & editing, Validation, Conceptualization. **Chengshan Wang:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

Code availability

The Python code (Jupyter notebook) used in this paper is available at GitHub: https://github.com/YangYidaaa/Yang-2025-SundaShelf. This Python code is developed to facilitate interaction with ArcGIS Pro for geospatial analysis and data processing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is a contribution to the Deep-time Digital Earth (DDE) Big Science Program. This work was supported by National Natural Science Foundation of China (42050102); Natural Science Foundation of Shanghai (21ZR1466800); and Fund of State Key Laboratory of Marine Geology (MGZ202304). We thank Prof. Chao Ma (Chengdu University of Technology) for his suggestions on data analysis and spatial interpolation. We are also grateful to Dr. Natascha Riedinger and two anonymous reviewers for their constructive comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margeo.2025.107566.

Data availability

The original data used in our manuscript was from the earlier study (Ma et al., 2022) published in AGU journal "Geophysical Research Letters" and had approval from the original authors. The original dataset can be found at Figshare: https://doi.org/10.6084/m9.figshare.21343662.v2. The processed dataset used in this paper is provided in Supplementary materials and also available at GitHub: https://github.com/YangYidaaa/Yang-2025-SundaShelf.

References

Alexander, C.R., DeMaster, D.J., Nittrouer, C.A., 1991. Sediment accumulation in a modern epicontinental-shelf setting: the Yellow Sea. Mar. Geol. 98 (1), 51–72. https://doi.org/10.1016/0025-3227(91)90035-3.

Antonakos, A., 2021. Spatial interpolation for the distribution of groundwater level in an area of complex geology using widely available GIS tools. Environ. Process. 8 (3), 993–1026. https://doi.org/10.1007/s40710-021-00529-9.

Anttila, E.S., Macdonald, F.A., Szymanowski, D., Schoene, B., Kylander-Clark, A., Danhof, C., Jones, D.S., 2023. Timing and tempo of organic carbon burial in the Monterey Formation of the Santa Barbara Basin and relationships with Miocene climate. Earth Planet. Sci. Lett. 620, 118343. https://doi.org/10.1016/j.epsl.2023.118343.

Bamisaiye, O.A., 2018. Subsurface mapping: selection of best interpolation method for borehole data analysis. Spat. Inf. Res. 26 (3), 261–269. https://doi.org/10.1007/s41324-018-0170-6.

Bastos, L.P.H., Rodrigues, R., Pereira, E., Bergamaschi, S., Alferes, C.L.F., Augland, L.E., et al., 2021. The birth and demise of the vast epicontinental Permian Irati-Whitehill sea: evidence from organic geochemistry, geochronology, and paleogeography. Palaeogeogr. Palaeoclimatol. Palaeoecol. 562, 110103. https://doi.org/10.1016/j.palaeo.2020.110103.

Bauer, J.E., Cai, W.-J., Raymond, P.A., Bianchi, T.S., Hopkinson, C.S., Regnier, P.A.G., 2013. The changing carbon cycle of the coastal ocean. Nature 504 (7478), 61–70. https://doi.org/10.1038/nature12857.

Bayon, G., Patriat, M., Godderis, Y., Trinquier, A., De Deckker, P., Kulhanek, D.K., et al., 2023. Accelerated mafic weathering in Southeast Asia linked to late Neogene cooling. Sci. Adv. 9 (13), eadf3141. https://doi.org/10.1126/sciadv.adf3141.

Berner, R.A., 1999. A new look at the long-term carbon cycle. GSA Today 9 (11), 1–6.

Bianchi, T.S., Cui, X., Blair, N.E., Burdige, D.J., Eglinton, T.I., Galy, V., 2018. Centers of organic carbon burial and oxidation at the land-ocean interface. Org. Geochem. 115, 138–155. https://doi.org/10.1016/j.orggeochem.2017.09.008.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Breitfeld, H.T., Hennig-Breitfeld, J., BouDagher-Fadel, M., Schmidt, W.J., Meyer, K., Reinprecht, J., et al., 2022. Provenance of Oligocene–Miocene sedimentary rocks in the Cuu Long and Nam Con Son basins, Vietnam and early history of the Mekong River. Int. J. Earth Sci. 111 (6), 1773–1804. https://doi.org/10.1007/s00531-022-02214-0.

Brown, C.E., 1998. Coefficient of variation. In: Applied Multivariate Statistics in Geohydrology and Related Sciences. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 155–157. https://doi.org/10.1007/978-3-642-80328-4_13.

Burdige, D.J., 2007. Preservation of organic matter in marine sediments: controls, mechanisms, and an imbalance in sediment organic carbon budgets? Chem. Rev. 107 (2), 467–485. https://doi.org/10.1021/cr050347q.

Burrough, P.A., McDonnell, R.A., 2015. Principles of Geographical Information Systems. Oxford University Press, USA.

Cartapanis, O., Bianchi, D., Jaccard, S.L., Galbraith, E.D., 2016. Global pulses of organic carbon burial in deep-sea sediments during glacial maxima. Nature communications 7 (1), 10796. https://doi.org/10.1038/ncomms10796.

Chen, F.W., Liu, C.-W., 2012. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Paddy Water Environ. 10 (3), 209–222. https://doi.org/10.1007/s10333-012-0319-1.

Chen, D., Ou, T., Gong, L., Xu, C.-Y., Li, W., Ho, C.-H., Qian, W., 2010. Spatial interpolation of daily precipitation in China: 1951-2005. Adv. Atmos. Sci. 27, 1221–1232. https://doi.org/10.1007/s00376-010-9151-y.

Chen, Y., Shan, X., Jin, X., Yang, T., Dai, F., Yang, D., 2016. A comparative study of spatial interpolation methods for determining fishery resources density in the Yellow Sea. Acta Oceanologica Sinica 35, 65–72. https://doi.org/10.1007/s13131-016-0966-y.

Collins, D.S., Avdis, A., Allison, P.A., Johnson, H.D., Hill, J., Piggott, M.D., et al., 2017. Tidal dynamics and mangrove carbon sequestration during the Oligo–Miocene in the South China Sea. Nat. Commun. 8 (1), 15698. https://doi.org/10.1038/ncomms15698.

Coulston, J.W., Blinn, C.E., Thomas, V.A., Wynne, R.H., 2016. Approximating prediction uncertainty for random forest regression models. Photogramm. Eng. Remote Sens. 82 (3), 189–197. https://doi.org/10.14358/PERS.82.3.189.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88 (11), 2783–2792. https://doi.org/10.1890/07-0539.1.

de Jonge-Anderson, I., Widyanita, A., Busch, A., Doster, F., Nicholson, U., 2024. New Insights into the structural and stratigraphic evolution of the Malay Basin using 3D seismic data: Implications for regional carbon capture and storage potential. Basin Research 36 (4), e12885. https://doi.org/10.1111/bre.12885.

Derry, L.A., France-Lanord, C., 1996. Neogene growth of the sedimentary organic carbon reservoir. Paleoceanography 11 (3), 267–275. https://doi.org/10.1029/95PA03839.

Diesing, M., 2020. Deep-sea sediments of the global ocean. Earth Syst. Sci. Data 12 (4), 3367–3381. https://doi.org/10.5194/essd-12-3367-2020.

Diesing, M., Thorsnes, T., Bjarnadóttir, L.R., 2021. Organic carbon densities and accumulation rates in surface sediments of the North Sea and Skagerrak. Biogeosciences 18 (6), 2139–2160. https://doi.org/10.5194/bg-18-2139-2021.
Diesing, M., Paradis, S., Jensen, H., Thorsnes, T., Bjarnadóttir, L.R., Knies, J., 2024. Glacial troughs as centres of organic carbon accumulation on the Norwegian continental margin. Commun. Earth Environ. 5 (1), 327. https://doi.org/10.1038/s43247-024-01502-8.
Doust, H., Sumner, H.S., 2007. Petroleum systems in rift basins – a collective approach inSoutheast Asian basins. Pet. Geosci. 13 (2), 127–144. https://doi.org/10.1144/1354-079307-746.
Du, P., Bai, X., Tan, K., Xue, Z., Samat, A., Xia, J., et al., 2020. Advances of four machine learning methods for spatial data handling: a review. J. Geovis. Spat. Anal. 4 (1), 13. https://doi.org/10.1007/s41651-020-00048-5.
Dutkiewicz, A., Müller, R.D., Wang, X., O'callaghan, S., Cannon, J., Wright, N.M., 2017. Predicting sediment thickness on vanished ocean crust since 200 Ma. Geochem. Geophys. Geosyst. 18 (12), 4586–4603. https://doi.org/10.1002/2017GC007258.
Galy, V., France-Lanord, C., Beyssac, O., Faure, P., Kudrass, H., Palhol, F., 2007. Efficient organic carbon burial in the Bengal fan sustained by the Himalayan erosional system. Nature 450 (7168), 407–410. https://doi.org/10.1038/nature06273.
Graw, J.H., Wood, W.T., Phrampus, B.J., 2021. Predicting global marine sediment density using the random forest regressor machine learning algorithm. J. Geophys. Res. Solid Earth 126 (1), e2020JB020135. https://doi.org/10.1029/2020JB020135.
Hall, R., 2009. Southeast Asia's changing palaeogeography. Blumea-Biodiv. Evol. Biogeogr. Plants 54 (1–3), 148–161. https://doi.org/10.3767/000651909X475941.
Hall, R., 2012. Late Jurassic–Cenozoic reconstructions of the Indonesian region and the Indian Ocean. Tectonophysics 570–571, 1–41. https://doi.org/10.1016/j.tecto.2012.04.021.
Harries, P.J., 2009. Epeiric seas: a continental extension of shelf biotas. In: Earth System: History and Natural Variability. EOLSS Publishers Company Limited, UK, pp. 138–155.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. Stanford University, Stanford, CA. https://doi.org/10.1007/978-0-387-84858-7.
Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., et al., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PloS One 10 (6), e0125814. https://doi.org/10.1371/journal.pone.0125814.
Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/peerj.5518.
Heung, B., Bulmer, C.E., Schmidt, M.G., 2014. Predictive soil parent material mapping at a regional-scale: a Random Forest approach. Geoderma 214–215, 141–154. https://doi.org/10.1016/j.geoderma.2013.09.016.
Hodson, T.O., 2022. Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not. Geosci. Model Dev. 15 (14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022.
Holbourn, A., Kuhnt, W., Lyle, M., Schneider, L., Romero, O., Andersen, N., 2014. Middle Miocene climate cooling linked to intensification of eastern equatorial Pacific upwelling. Geology 42 (1), 19–22. https://doi.org/10.1130/G34890.1.
Holbourn, Ann, Kuhnt, W., Kochhann, K.G.D., Andersen, N., Sebastian Meier, K.J., 2015. Global perturbation of the carbon cycle at the onset of the Miocene Climatic Optimum. Geology 43 (2), 123–126. https://doi.org/10.1130/G36317.1.
Houborg, R., McCabe, M.F., 2018. Daily retrieval of NDVI and LAI at 3 m resolution via the fusion of CubeSat, Landsat, and MODIS data. Remote Sens. (Basel) 10 (6), 890. https://doi.org/10.3390/rs10060890.
Hu, L., Shi, X., Bai, Y., Qiao, S., Li, L., Yu, Y., et al., 2016. Recent organic carbon sequestration in the shelf sediments of the Bohai Sea and Yellow Sea, China. J. Mar. Syst. 155, 50–58. https://doi.org/10.1016/j.jmarsys.2015.10.018.
Hu, L., Chun, Y., Griffith, D.A., 2022. Incorporating spatial autocorrelation into house sale price prediction using random forest model. Trans. GIS 26 (5), 2123–2144. https://doi.org/10.1111/tgis.12931.
Johnston, K., Hoef, J.M.V., Krivoruchko, K., Lucas, N., 2001. Using ArcGIS Geostatistical Analyst. ESRI, USA.
Kaya, M.Y., Dupont-Nivet, G., Frieling, J., Fioroni, C., Rohrmann, A., Altıner, S.Ö., et al., 2022. The Eurasian epicontinental sea was an important carbon sink during the Palaeocene-Eocene thermal maximum. Commun. Earth Environ. 3 (1), 124. https://doi.org/10.1038/s43247-022-00451-4.
Kennedy, M.J., Wagner, T., 2011. Clay mineral continental amplifier for marine carbon sequestration in a greenhouse ocean. Proc. Natl. Acad. Sci. 108 (24), 9776–9781. https://doi.org/10.1073/pnas.1018670108.
Kim, J., Lee, Y., Lee, M.-H., Hong, S.-Y., 2022. A comparative study of machine learning and spatial interpolation methods for predicting house prices. Sustainability 14 (15), 9056. https://doi.org/10.3390/su14159056.
Kob, M.R.C., Ali, M.Y., 2008. Regional controls on the development of carbonates in East Natuna basin and Luconia area. In: Paper presented at the Proceedings, Indonesian Petroleum Association 32nd Annual Convention & Exhibition. IPA08-G-078.
Kump, L.R., Arthur, M.A., 1999. Interpreting carbon-isotope excursions: carbonates and organic matter. Chem. Geol. 161 (1–3), 181–198. https://doi.org/10.1016/S0009-2541(99)00086-8.
Lee, T.R., Wood, W.T., Phrampus, B.J., 2019. A machine learning (kNN) approach to predicting global seafloor total organic carbon. Global Biogeochem. Cycles 33 (1), 37–46. https://doi.org/10.1029/2018GB005992.
Leong, K.M. (Ed.), 2000. The Petroleum Geology and Resources of Malaysia. Malaysian National Petroleum Corporation.
Li, J., Heap, A.D., 2008. A Review of Spatial Interpolation Methods for Environmental Scientists. Geoscience Australia, Australia.
Li, J., Heap, A.D., Potter, A., Daniell, J.J., 2011. Application of machine learning methods to spatial interpolation of environmental variables. Environ. Model. Software 26 (12), 1647–1659. https://doi.org/10.1016/j.envsoft.2011.07.004.
Li, Z., Zhang, Y.G., Torres, M., Mills, B.J.W., 2023. Neogene burial of organic carbon in the global ocean. Nature 613 (7942), 90–95. https://doi.org/10.1038/s41586-022-05413-6.
Lin, B., Liu, Z., Zhao, M., Sompongchaiyakul, P., Zhang, H., Blattmann, T.M., et al., 2023. Compositions and sources of sedimentary organic carbon on the tropical epicontinental sea. Geochim. Cosmochim. Acta 351, 32–44. https://doi.org/10.1016/j.gca.2023.04.030.
Liu, J., 2001. Post-Glacial Sedimentation in a River -Dominated Epicontinental Shelf: The Yellow Sea Example (Doctoral Dissertation). The College of Willian & Mary, Virginia. Retrieved from. http://scholarworks.wm.edu/etd/1539616747.
Lu, G.Y., Wong, D.W., 2008. An adaptive inverse-distance weighting spatial interpolation technique. Comput. Geosci. 34 (9), 1044–1055. https://doi.org/10.1016/j.cageo.2007.07.010.
Ma, P., Liu, Z., Jiang, M., Cheng, H., Zhang, L., Cai, D., 2022. Carbon sequestration of the Middle Miocene Sunda Shelf facilitated global climate change. Geophys. Res. Lett. https://doi.org/10.1029/2022gl100638.
Ma, P., Ma, C., Yang, S., Fernandez, A.R., 2023. East Asian summer monsoon evolution recorded by the middle Miocene pelagic reddish clay, South China Sea. Global Planet. Change 222, 104072. https://doi.org/10.1016/j.gloplacha.2023.104072.
Maxwell, K., Rajabi, M., Esterle, J., 2021. Spatial interpolation of coal properties using geographic quantile regression forest. Int. J. Coal Geol. 248, 103869. https://doi.org/10.1016/j.coal.2021.103869.
Mayer, L.M., 1994. Surface area control of organic carbon accumulation in continental shelf sediments. Geochim. Cosmochim. Acta 58 (4), 1271–1284. https://doi.org/10.1016/0016-7037(94)90381-6.
Miller, K.G., Browning, J.V., Schmelz, W.J., Kopp, R.E., Mountain, G.S., Wright, J.D., 2020. Cenozoic sea-level and cryospheric evolution from deep-sea geochemical and continental margin records. Sci. Adv. 6 (20), eaaz1346. https://doi.org/10.1126/sciadv.aaz1346.
Morley, R.J., Morley, H.P., 2013. Mid Cenozoic freshwater wetlands of the Sunda region. J. Limnol. 72, 2. https://doi.org/10.4081/jlimnol.2013.s2.e2.
Morley, R.J., Morley, H.P., Swiecicki, T., 2016. Mio-Pliocene Palaeogeography, Uplands and River Systems of the Sunda Region Based on Mapping Within a Framework of Vim Depositional Cycles. https://doi.org/10.29118/IPA.0.16.506.G.
Morley, R.J., Hasan, S.S., Morley, H.P., Jais, J.H.M., Mansor, A., Aripin, M.R., et al., 2021. Sequence biostratigraphic framework for the Oligocene to Pliocene of Malaysia_ High-frequency depositional cycles driven by polar glaciation. Palaeogeogr. Palaeoclimatol. Palaeoecol. 561, 110058. https://doi.org/10.1016/j.palaeo.2020.110058.
Müller, D., Liu, B., Geibert, W., Holtappels, M., Sander, L., Miramontes, E., et al., 2024. Depositional controls and budget of organic carbon burial in fine-grained sediments of the North Sea – the Helgoland Mud Area as a test field. EGUsphere [preprint] 2024. https://doi.org/10.5194/egusphere-2024-1632.
Nie, J., Ruetenik, G., Gallagher, K., Hoke, G., Garzione, C.N., Wang, W., et al., 2018. Rapid incision of the Mekong River in the middle Miocene linked to monsoonal precipitation. Nat. Geosci. 11 (12), 944–948. https://doi.org/10.1038/s41561-018-0244-z.
Oliver, M.A., Webster, R., 2015. Basic Steps in Geostatistics: The Variogram and Kriging. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-15865-5.
Owens, J.D., Lyons, T.W., Lowery, C.M., 2018. Quantifying the missing sink for global organic carbon burial during a Cretaceous oceanic anoxic event. Earth Planet. Sci. Lett. 499, 83–94. https://doi.org/10.1016/j.epsl.2018.07.021.
Park, Y., Maffre, P., Goddéris, Y., Macdonald, F.A., Anttila, E.S.C., Swanson-Hysell, N.L., 2020. Emergence of the Southeast Asian islands as a driver for Neogene cooling. Proc. Natl. Acad. Sci. 117 (41), 25319–25326. https://doi.org/10.1073/pnas.2011033117.
Pubellier, M., Morley, C.K., 2014. The basins of Sundaland (SE Asia): Evolution and boundary conditions. Mar. Pet. Geol. 58, 555–578. https://doi.org/10.1016/j.marpetgeo.2013.11.019.
Qiao, P., Lei, M., Yang, S., Yang, J., Guo, G., Zhou, X., 2018. Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing. Environ. Sci. Pollut. Res. 25 (16), 15597–15608. https://doi.org/10.1007/s11356-018-1552-y.
Ramli, M.N., 1988. Stratigraphy and palaeofacies development of Carigali's operating areas in the Malay Basin, South China Sea. Bulletin Geological Society of Malaysia 22, 153–187. https://doi.org/10.7186/bgsm22198808.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, F., 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566 (7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1.
Reimann, C., Filzmoser, P., 2000. Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. Environ. Geol. 39, 1001–1014. https://doi.org/10.1007/s002549900081.
Schwarz, E., Remírez, M., Lazo, D.G., Veiga, G.D., Isla, M.I., Echevarria, C., et al., 2022. A review on depositional systems, bioevents and paleogeography of the Valanginian-Hauterivian Neuquén Sea: refining sedimentary and biological signals linked to the dynamics of epeiric seas. Earth Sci. Rev. 234, 104224. https://doi.org/10.1016/j.earscirev.2022.104224.
Scotese, C.R., 2021. An atlas of Phanerozoic paleogeographic maps: the seas come in and the seas go out. Annu. Rev. Earth Planet. Sci. 49 (1), 679–728. https://doi.org/10.1146/annurev-earth-081320-064052.

Sekulić, A., Kilibarda, M., Heuvelink, G.B.M., Nikolić, M., Bajat, B., 2020. Random forest spatial interpolation. Remote Sensing 12 (10), 1687. https://doi.org/10.3390/rs12101687.

Shields, G.A., Mills, B.J.W., 2017. Tectonic controls on the long-term carbon isotope mass balance. Proc. Natl. Acad. Sci. 114 (17), 4318–4323. https://doi.org/10.1073/pnas.1614506114.

Shoup, R., Morley, R., Swiecicki, T., Clark, S., 2012. Tectono-stratigraphic framework and tertiary paleogeography of southeast Asia: Gulf of Thailand to South Vietnam Shelf. In: AAPG, Search and Discovery Article# 30246.

Straume, E.O., Gaina, C., Medvedev, S., Hochmuth, K., Gohl, K., Whittaker, J.M., et al., 2019. GlobSed: updated total sediment thickness in the world's oceans. Geochem. Geophys. Geosyst. 20 (4), 1756–1772. https://doi.org/10.1029/2018GC008115.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Economic geography 46 (sup1), 234–240. https://doi.org/10.2307/143141.

Vance, T.C., Huang, T., Butler, K.A., 2024. Big data in Earth science: emerging practice and promise. Science 383 (6688), eadh9607. https://doi.org/10.1126/science.adh9607.

Wang, C., Hazen, R.M., Cheng, Q., Stephenson, M.H., Zhou, C., Fox, P., et al., 2021. The Deep-Time Digital Earth program: data-driven discovery in geosciences. Natl. Sci. Rev. 8 (9), nwab027. https://doi.org/10.1093/nsr/nwab027.

Wang, X., Algeo, T.J., Li, C., Zhu, M., 2023. Spatial pattern of marine oxygenation set by tectonic and ecological drivers over the Phanerozoic. Nat. Geosci. 1–7. https://doi.org/10.1038/s41561-023-01296-y.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. John Wiley & Sons, Ltd., USA https://doi.org/10.1002/9780470517277. Retrieved from.

Wellmer, F.W., 1998. Skewed distributions. In: Statistical Evaluations in Exploration for Mineral Deposits. Springer, Berlin, Heidelberg, pp. 53–65. https://doi.org/10.1007/978-3-642-60262-7_8.

Xu, L., Chen, N., Chen, Z., Zhang, C., Yu, H., 2021a. Spatiotemporal forecasting in earth system science: methods, uncertainties, predictability and future directions. Earth-Sci. Rev. 222, 103828. https://doi.org/10.1016/j.earscirev.2021.103828.

Xu, Z., Wan, S., Colin, C., Clift, P.D., Chang, F., Li, T., et al., 2021b. Enhancements of Himalayan and Tibetan erosion and the produced organic carbon burial in distal tropical marginal seas during the Quaternary glacial periods: an integration of sedimentary records. J. Geophys. Res. Earth 126 (3), e2020JF005828. https://doi.org/10.1029/2020JF005828.

Yang, Z., Zhang, G.X., Zhang, L., 2016. The evolution and main controlling factors of reef and carbonate platform in Wan'an Basin. Earth Sci. 41 (8), 1349–1360. https://doi.org/10.3799/dqkx.2016.107.

Zhu, A., Lu, G., Liu, J., Qin, C., Zhou, C., 2018. Spatial prediction based on Third Law of Geography. Ann. GIS 24 (4), 225–240. https://doi.org/10.1080/19475683.2018.1534890.

Zhu, R., Wang, H., Wang, H., Wang, X., Wan, B., Zhang, W., et al., 2024. Multi-spherical interactions and mechanisms of hydrocarbon enrichment in the Southeast Asian archipelagic tectonic system. Sci. China Earth Sci. 67 (2), 566–583. https://doi.org/10.1007/s11430-023-1254-4.