scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Spatiotemporal estimation of nutrient data from the northwest pacific and east asian seas

Gi Seop Lee[1], Jung Ho Lee[2] & Hong Yeon Cho[1]✉

Nutrient data obtained from field observations have the potential to enhance our understanding of oceanic biogeochemical cycling and productivity changes. In particular, long-term nutrient data can provide valuable information on the links between climate change and biogeochemical changes. However, unlike other observational variables such as sea surface temperature, nutrient data are limited in terms of their broad-scale observations and automated sensor-based measurements. In this study, we analyzed nitrate and phosphate data obtained from coastal regions in Northeast Asia and the northwest Pacific from 1980 to 2019 using the spatiotemporal kriging technique and provide results in a spatiotemporal grid format. The data are available at monthly intervals and may be attractive to researchers in the fields of oceanography, marine ecology, and marine biogeochemistry at the climate change scale. Furthermore, sharing the source code of the data production process can contribute to better long-term data reproduction in the future.

Background & Summary

The northwest Pacific and adjacent eastern Asian waters are known for their high primary productivity[1–3], which has been examined in various marine chemical and biological studies, including those on water quality changes and material cycling. Nutrient data play a crucial role in these studies, with nitrogen and phosphorus being especially significant, as they influence the growth and reproduction of phytoplankton and shape the area's phytoplankton species composition[4–9]. Human activities, such as artificial nitrogen input in densely populated regions such as eastern China and western Europe, affect the ocean's biogeochemical structure[10–13].

Coastal waters near East Asia, including the heavily impacted Yellow Sea and East China Sea, the unique East Sea with relatively lower human impact, and the northwest Pacific influenced by the Kuroshio current, are influenced by both natural and human factors, leading to complex changes in nutrient levels[14]. Nutrient supply from the deep sea has been reported to have decreased worldwide due to the strengthening of the stratification[1–3,15], while artificial supply through the atmosphere and rivers has increased in the East Asian waters[11,12,16–18]. Thus, understanding the long-term changes in nutrient levels in this region is crucial, and various perspectives are being studied to comprehend the phenomenon and predict future changes[6,19–21].

From the perspective of the spatial estimation of ocean information, studies providing gridded data such as OISST (Optimum Interpolation SST [Sea Surface Temperature]) are ongoing[22–24]. In addition, research is being conducted to remove the bias of numerical models using spatial estimation techniques such as Kriging[25]. However, these techniques mainly focus on temperature and salinity. Nutrient data, however, primarily rely on *in situ* observations, as satellite remote sensing and unmanned equipment cannot cover them. Efforts have been made to provide monthly gridded nutrient data for the North Pacific region[26–28], but 4D gridded nutrient data (x, y, z, t) that consider both spatial and temporal variations are limited to some reanalysis data using numerical models[29].

This study uses 40 years of nutrient concentration observations from 1980 to 2019 to spatially and temporally optimize the data for the northwest Pacific Ocean (N25–45°, E121–145°) into a gridded format. The estimated nutrient grid data were validated through a verification process and are presented (with validation errors shown) in the modeled results.

[1]Marine Bigdata AI Center, Korea Institute of Ocean Science and Technology, Busan, 49111, South Korea. [2]Data Tech Team, SK Telecom, Seoul, 02598, South Korea. ✉e-mail: hycho@kiost.ac.kr

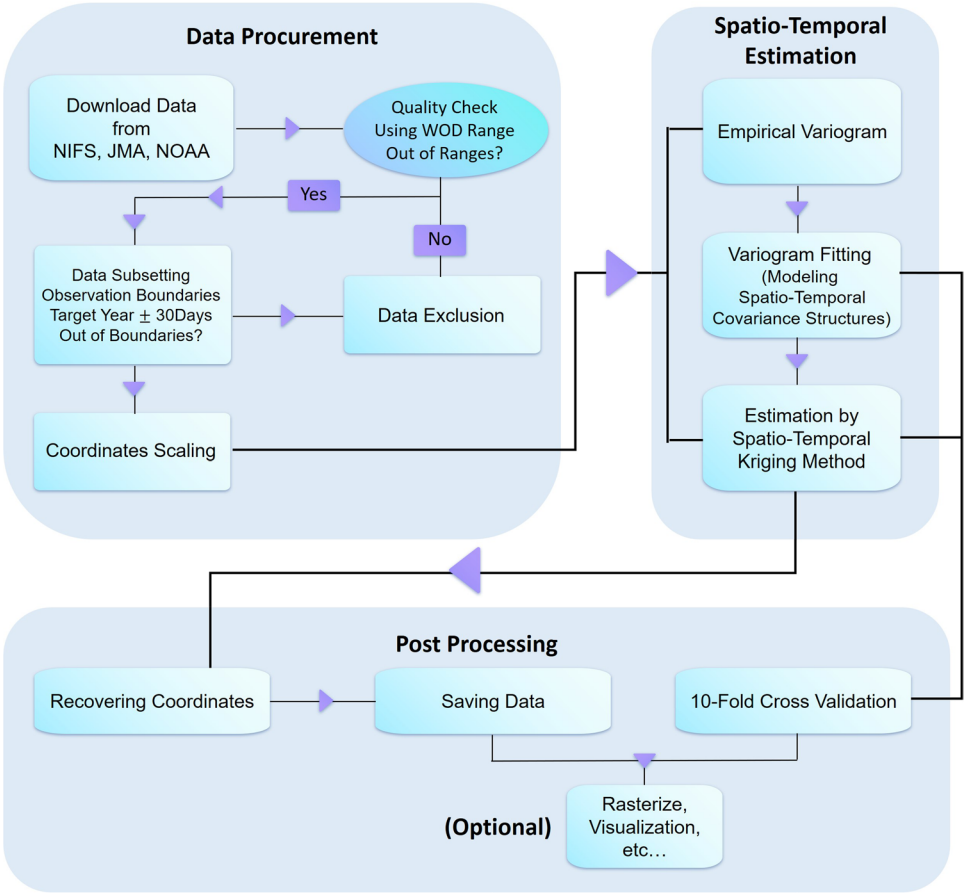

Fig. 1 Workflow of the study. Nutrient and depth data were downloaded from various sources. The nutrient data obtained from different sources were compiled into a common data format, and data within the valid range of time, space, and nutrient variables were selected. Then, the refined data were used for variogram modeling and spatiotemporal kriging. 10-fold cross-validation was performed to indicate estimation errors, and the data were saved.

Methods

The procedure used to create gridded data is summarized in Fig. 1. This section outlines the steps involved in transforming raw observational data into nutrient grid data and validating the results. All procedures were executed using the R programming language (R core team, 2023)[30]. The data production process consisted mainly of three steps: data collection and preprocessing, spatiotemporal estimation, and postprocessing and validation. These steps will be discussed in further detail below.

Data procurement. The data analyzed in this study were acquired from the National Institute of Fisheries Science (NIFS) Serial Oceanographic Observation (SOO, https://www.nifs.go.kr/kodc/eng/eng_coo_list.kodc)[31], Japan Meteorological Agency (JMA) Oceanographic and Marine Meteorological Observations by Research Vessels (OMMORV, https://www.data.jma.go.jp/gmd/kaiyou/db/vessel_obs/data-report/html/ship/ship_e.php)[32], and topographic data from the National Oceanic and Atmospheric Administration (NOAA) Earth TOPOgraphy(https://www.ncei.noaa.gov/maps/grid-extract/)[33,34]. The data specifications, including file format, observational period, spatiotemporal resolution, and accessible URLs, are presented in Table 1.

The OMMORV data are currently accessible for download starting from 1997, with earlier data available in the JMA Data Report of Oceanographic Observations Special Issue[35]. To avoid duplications, any potential overlap in the data were excluded from the spatiotemporal estimation process.

Nutrient data. The **SOO data** can be obtained by specifying the region, line, station, observation date, and depth. These data comprise ten water quality parameters, including temperature, salinity, dissolved oxygen concentration, nitrate concentration, and phosphate concentration. The data have been collected since 1961 with a bimonthly (February, April, June, August, October, and December) measurement frequency. Although the station and line locations may have undergone slight changes, as of 2020, data from 207 stations along 25 lines have been accumulated.

Title	Provider	Format	Temporal Range	Resolution	Access URL
Serial Oceanographic Observation (SOO)	NIFS, Korea	.txt,.xls(excel),.csv(in korean)	1967–2022 1980–2022 (used)	Bimonthly (temporal)	https://www.nifs.go.kr/kodc/eng/eng_coo_list.kodc (in English) https://www.nifs.go.kr/kodc/soo_list.kodc (in Korean)
Oceanographic and Marine Meteorological Observations by Research Vessels, Hydrographic data (OMMORV)	JMA, Japan	.E (fixed width)	1964–2022 1980–2022 (used)	Seasonally (temporal)	https://www.data.jma.go.jp/gmd/kaiyou/db/vessel_obs/data-report/html/ship/ship_e.php (in English, >1997) https://warp.ndl.go.jp/info:ndljp/pid/11160873/www.data.jma.go.jp/gmd/kaiyou/db/vessel_obs/data-report/html/ship/efile_NoS2_e.html (in English, <2010)
ETOPO 2022	NOAA, US	.tif	—	15 arc sec (spatial)	https://www.ncei.noaa.gov/maps/grid-extract/

Table 1. Data specifications for SOO, OMMORV, and ETOPO 2022. The table includes information on data providers, file formats, observational periods, spatiotemporal resolutions, and accessible URLs.

Variables	Description	Unit	class	N of missing
Nat	Nation		Character	—
st	Station		Character	—
date_time	Date and Time		POSIXct	4,150
lat	Latitude	Degree(°)	Numeric	—
long	Longitude		Numeric	—
tm_x	Coordinate x, TM projected long	m	Numeric	—
tm_y	Coordinate y, TM projected lat		Numeric	—
depth	Depth		Numeric	766,492
wt	Water Temperature	°C	Numeric	632,880
sal	Salinity	psu	Numeric	633,813
pH	potential of Hydrogen	—	Numeric	1,611,517
DO	Dissolved Oxygen	ml/L (NIFS) μmol/kg (JMA)	Numeric	803,932
PO4	Phosphate	μmol/L (NIFS, JMA <2010 winter) μmol/kg (JMA ≥ 2010 spring)	Numeric	1,160,474
NO2	Nitrite		Numeric	1,291,329
NO3	Nitrate		Numeric	1,296,471
SiO4	Silicate		Numeric	1,619,079

Table 2. The basic description of the dataset merged from SOO and OMMORV. The preprocessed R object name and its description, unit, data type, and number of missing data are specified.

The **OMMORV data** can be obtained through the provided link associated with each research vessel. The hydrographic data, saved with an '.E' extension, were used in this study. The format of the data underwent a change in 2010 and is now classified into two versions, 'E2.x' and 'E3.x'. The data are a 126-byte ASCII record in a fixed width format, with observation information and items arranged at regular intervals. Detailed information on the format of the data is available separately[36].

To ensure compatibility, the nutrient data from both SOO and OMMORV underwent unit conversion. The original units of μmol/kg were changed to μmol/L, and the density was calculated based on the recorded water temperature and salinity data. This calculation was carried out assuming standard atmospheric pressure (10.1325 dbar) with the *gsw* library in the R programming language[37,38]. The SOO and OMMORV data were merged into a table of 2,023,251 observations and 16 variables, and the missing information for each variable is shown in Table 2.

The bathymetry used in this study utilized NOAA's ETOPO 2022 data, which is the second release of these data following ETOPO1[33,39]. The water depth data can be obtained either by downloading the desired area data from the following URL or by using the R Package *marmap*[40]. In this study, the water depth data within the range of N25–45° and E121–145° were processed at 10-minute intervals using the *getNOAA.bathy* function from the *marmap* library. The computation was performed on grid points that were densely set from the surface to the 9,000 m depth range using the standard depth from WOD18[41]. However, to reduce the computational load, the number of depth grids was reduced from 137 to 43. Depth intervals were created at 20 m intervals for the 0–200 m range, 100 m intervals for the 200–2000 m range, and 500 m intervals for the 2000–9000 m range to produce the grids.

Spatiotemporal estimation. The spatiotemporal kriging (STK) approach was used to transform the irregular nutrient data collected from the SOO and OMMORV sources into spatiotemporal grid data. This method

has been widely used in various fields[42–45] and distinguishes itself from those in previous studies by considering the vertical dimension in the 4-dimensional spatiotemporal estimation. The R libraries *gstat* and *spacetime* were utilized for maximum 3-dimensional spatiotemporal estimation[46–48]; however, a custom function had to be developed, as 4-dimensional coordinates are not supported by these libraries.

Kriging with External Drift (KED) was applied for nutrient estimation in unmeasured spatiotemporal areas using spatiotemporal coordinates as the auxiliary variables. KED is also referred to as universal kriging when the drift is limited to spatial coordinates[49,50]. The unmeasured point's estimation at a specific time point is represented as a weighted combination of the spatial trend and the residual from the regression model at the measured point as in Eq. (1).

$$\hat{z}(x_0) = \sum_{k=1}^{m} \omega_k f_k(x_0) + \omega_0 + \sum_{i=1}^{n} \lambda_i e(z(x_i) - \bar{z}) \tag{1}$$

where x_i and x_0 represent the coordinates of the observation and the target location, respectively, and the 4-dimensional coordinate structure including horizontal, vertical, and temporal dimensions is represented as $[x, y, z, t] = [x_1, \cdots, x_m]$. The subscript i denotes the observed location, and 0 denotes the location of interest for prediction. $f_k(x_0)$ is a function representing the average spatiotemporal variation, and a linear function is used. ω_k is the coefficient of the regression function $f_k(x_0)$, and ω_0 is the Lagrange parameter to remove bias. e is the residual of $f_k(x_0)$, and λ_i represents the weight coefficient of e.

The optimal coefficients (ω_k, ω_0, λ_i) that satisfy the condition of minimizing the error variance in Eq. (2) are derived in the form of Eq. (3), and it is solved as shown in Eq. (4).

$$\min [\hat{z}(x_0) - z(x_0)]^2 \tag{2}$$

The process of finding the solution in the form of a matrix equation is shown in Eqs. (3, 4), and the block matrices that make up the overall matrix equation are constituted as in Eq. (5–11). The bold symbols indicate the block matrix.

$$\boldsymbol{C}^{\boldsymbol{KED}} \cdot \boldsymbol{\lambda}^{\boldsymbol{KED}} = \boldsymbol{C}_{\boldsymbol{0}}^{\boldsymbol{KED}} \tag{3}$$

$$\lambda^{\boldsymbol{KED}} = \boldsymbol{C}^{\boldsymbol{KED}^{-1}} \cdot \boldsymbol{C}_{\boldsymbol{0}}^{\boldsymbol{KED}} \tag{4}$$

$$\boldsymbol{\lambda}^{\boldsymbol{KED}} = \begin{bmatrix} \boldsymbol{\lambda_i} \\ \omega_0 \\ \boldsymbol{\omega_k} \end{bmatrix}, \; i = 1, \ldots, n; \; k = 1, \ldots, m \tag{5}$$

$$\boldsymbol{C}^{\boldsymbol{KED}} = \begin{bmatrix} \boldsymbol{\sigma_{ij}^2} & \boldsymbol{I_n} & \boldsymbol{X} \\ \boldsymbol{I_n^T} & \boldsymbol{0} & \boldsymbol{0_m} \\ \boldsymbol{X^T} & \boldsymbol{0_m^T} & \boldsymbol{0} \end{bmatrix} \tag{6}$$

$$\boldsymbol{C}_{\boldsymbol{0}}^{\boldsymbol{KED}} = \begin{bmatrix} \boldsymbol{\sigma_{0j}^2} \\ 1 \\ \boldsymbol{X_0} \end{bmatrix} \tag{7}$$

where $\boldsymbol{X}$ is the observed coordinates, $\boldsymbol{X_0}$ is the target coordinates, and ~ represents the min-max scaled coordinates. $\boldsymbol{I_n}$ is a unit vector of $n \times 1$, $\boldsymbol{0_m}$ is a zero vector of $1 \times m$, and $\boldsymbol{0}$ is a zero matrix of $m \times m$. The superscript T on a matrix represents the transpose.

$$\boldsymbol{X} = \begin{bmatrix} (\tilde{x}_1)_1 & \cdots & (\tilde{x}_m)_1 \\ \vdots & \ddots & \vdots \\ (\tilde{x}_1)_n & \cdots & (\tilde{x}_m)_n \end{bmatrix} = \begin{bmatrix} \tilde{x}_1 & \tilde{y}_1 & \tilde{z}_1 & \tilde{t}_1 \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{x}_n & \tilde{y}_n & \tilde{z}_n & \tilde{t}_n \end{bmatrix}, = \begin{bmatrix} \tilde{x}_0 \\ \tilde{y}_0 \\ \tilde{z}_0 \\ \tilde{t}_0 \end{bmatrix} \tag{8, 9}$$

$$\boldsymbol{\sigma_{ij}^2} = \boldsymbol{\gamma_{st}^{SM}}(\boldsymbol{h_{ij}}, \boldsymbol{u_{ij}}), \; \boldsymbol{\sigma_{0j}^2} = \boldsymbol{\gamma_{st}^{SM}}(\boldsymbol{h_{0i}}, \boldsymbol{u_{0i}}) \tag{10, 11}$$

where the matrices $\boldsymbol{\sigma_{ij}^2}$ and $\boldsymbol{\sigma_{0j}^2}$ are calculated based on the spatiotemporal variogram $\gamma_{st}^{SM}(h, u)$ estimated from the observational data. The variogram represents the change in covariance with respect to distance and time, reflecting the strength of the correlation between data points as a function of spatial and temporal distance. $\gamma_{st}^{SM}(h, u)$ is expressed as a function of the spatial distance h and the temporal distance u (12, 13).

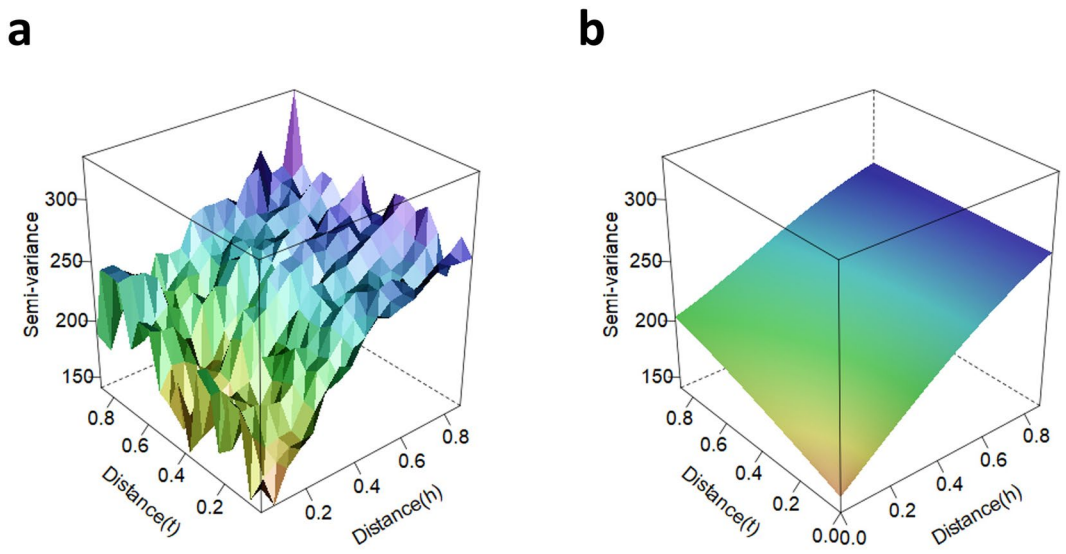

Fig. 2 Example of empirical (**a**) and fitted (**b**) variograms. The nitrate data from 2013 were used and fitted using the Sum-Metric model. The spatial and temporal distances were min-max scaled. The semivariance of the z-axis represents the inverse correlation with respect to spatiotemporal distance.

$$h_{ij} = \sqrt{(\tilde{x}_i - \tilde{x}_j)^2 + (\tilde{y}_i - \tilde{y}_j)^2 + (\tilde{z}_i - \tilde{z}_j)^2} \tag{12.1}$$

$$h_{0i} = \sqrt{(\tilde{x}_0 - \tilde{x}_i)^2 + (\tilde{y}_0 - \tilde{y}_i)^2 + (\tilde{z}_0 - \tilde{z}_i)^2} \tag{12.2}$$

$$u_{ij} = \sqrt{(\tilde{t}_i - \tilde{t}_j)^2} \tag{13.1}$$

$$u_{0i} = \sqrt{(\tilde{t}_0 - \tilde{t}_i)^2} \tag{13.2}$$

The $\gamma_{st}^{SM}(h, u)$ was fitted using the Sum-Metric (14) model[51,52]. The Spherical model (15) was applied equally to the γ_s, γ_t, and γ_{joint} models.

$$\gamma_{st}^{SM}(h, u) = \gamma_s(h_{ij}) + \gamma_t(u_{ij}) + \gamma_{joint}\left(\sqrt{h_{ij}^2 + (\kappa u_{ij})^2}\right) \tag{14}$$

$$\gamma(h) = Var(z) - C_0\left(1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right) + b \tag{15}$$

In the variogram model, h is the separation distance (generally, spatial distance), a is the range, and b is the nugget. The minimum of the observational data is $C_0 + b$, and κ is an anisotropic parameter for time. Each parameter was optimally estimated using the L-BFGS-B algorithm[46,53]. An example of spatiotemporal variogram modeling using observed nitrate data from 2013 is provided in Fig. 2.

When computing the actual $\hat{z}(x_0)$, only the estimated λ_i is used, and the regression coefficient ω_k, which determines the spatial average variation, is calculated using Eq. (16) as a constraint.

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i z_i \tag{16}$$

The prediction results of Eq. (16) are accompanied by the indicator of estimation uncertainty, the error variance, as provided in Eq. (17).

$$\sigma_{KED}^2 = \sigma^2 - \sum_{i=1}^{n} \lambda_i \sigma_{0i}^2 + \sum_{k=0}^{m} \omega_k f_k(x_0) \tag{17}$$

Data Records

The reproduced data are provided in comma-separated values (.csv) and R Data (.RData) file formats, with processing codes written in the R language[54]. The code for reading, analyzing, and visualizing the data can also be used to update the data. The data provided in CSV format consist of spatiotemporal coordinates (x, y, z, t), estimated values of nitrate or phosphate, and error variances of kriging. The error variances provide quantitative

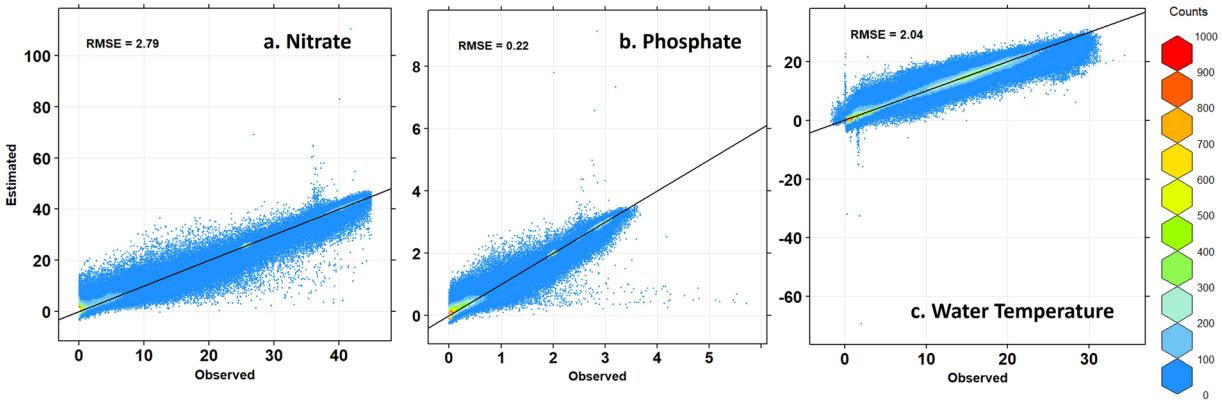

Fig. 3 10-fold cross-validation results using observational data from 1980 to 2019, spanning a total of 40 years(**a**: nitrate, **b**: phosphate, **c**: water temperature). Some values fall outside of the valid range, but many data points are densely distributed around the 1:1 line (black solid line). Due to the nature of KED, concentration values calculated as negative values can be converted to 0 or detection limit values.

Evaluation Metrics	Water Temperature (°C)	Nitrate (μmol/L)	Phosphate (μmol/L)
RMSE	2.04	2.79	0.22
MAE	1.41	1.84	0.14
R^2_{adj}	0.93	0.97	0.96

Table 3. The error evaluation metrics for the 10-fold cross validation, including the root mean square error (RMSE), mean absolute error (MAE), and adjusted R squared.

information on the magnitude of estimation errors and can be utilized in future conditional simulations. R Data (.RData) is a binary data format that can be directly loaded into memory using the *load* function in the R programming language for immediate use.

The dataset is projected in the Lambert azimuthal equal-area projection method with the following coordinate reference system (CRS):

"$+proj=laea\ +lat\_0=34.53333\ +lon\_0=137.0698\ +x\_0=0\ +y\_0=0\ +datum=WGS84\ +units=m\ +no\_defs\ +ellps=WGS84\ +towgs84=0,0,0$"

The data can be converted back to the longitude and latitude coordinate system using the following CRS:

"$+proj=longlat\ +datum=WGS84\ +no\_defs\ +ellps=WGS84\ +towgs84=0,0,0$"

Coordinate transformation using R can be performed using spatial data libraries such as *sp* and *sf*[55–57].

However, the conversion process may introduce slight errors, resulting in longitude and latitude coordinates with nonuniform degree intervals. Therefore, interpolation methods such as aggregation, nearest neighbor, or bilinear interpolation may be necessary for the stretched grid.

Technical Validation

The performance of the estimation model was evaluated using 10-fold cross-validation for spatial estimation results obtained through STK (Fig. 3, Table 3). Note that Simple and Ordinary Kriging always predict values that are less than or equal to the maximum observed value, while KED can predict values that are greater or smaller than the neighboring observed values. Therefore, if an estimated value falls outside the range of the WOD18 standard, it may need to be adjusted to a value within the range before interpretation. In this case, negative concentration values were replaced with 0. The root mean square error (RMSE), mean absolute error (MAE), and adjusted coefficient of determination (R^2_{adj}) were used as performance evaluation metrics (18)[58]

$$R^2_{adj} = \frac{n-1}{(n-m-1)(1-R^2)} \tag{18}$$

Where n is the number of data points and $m(=4)$ is the number of predictor variables used in the estimation.

The performance of the model in predicting water temperature and nitrate and phosphate concentrations was evaluated using error metrics (RMSE, MAE, adjusted R-squared). Note that since various reanalysis datasets are available, sea water temperature estimates are not provided here, and only the performance evaluation results

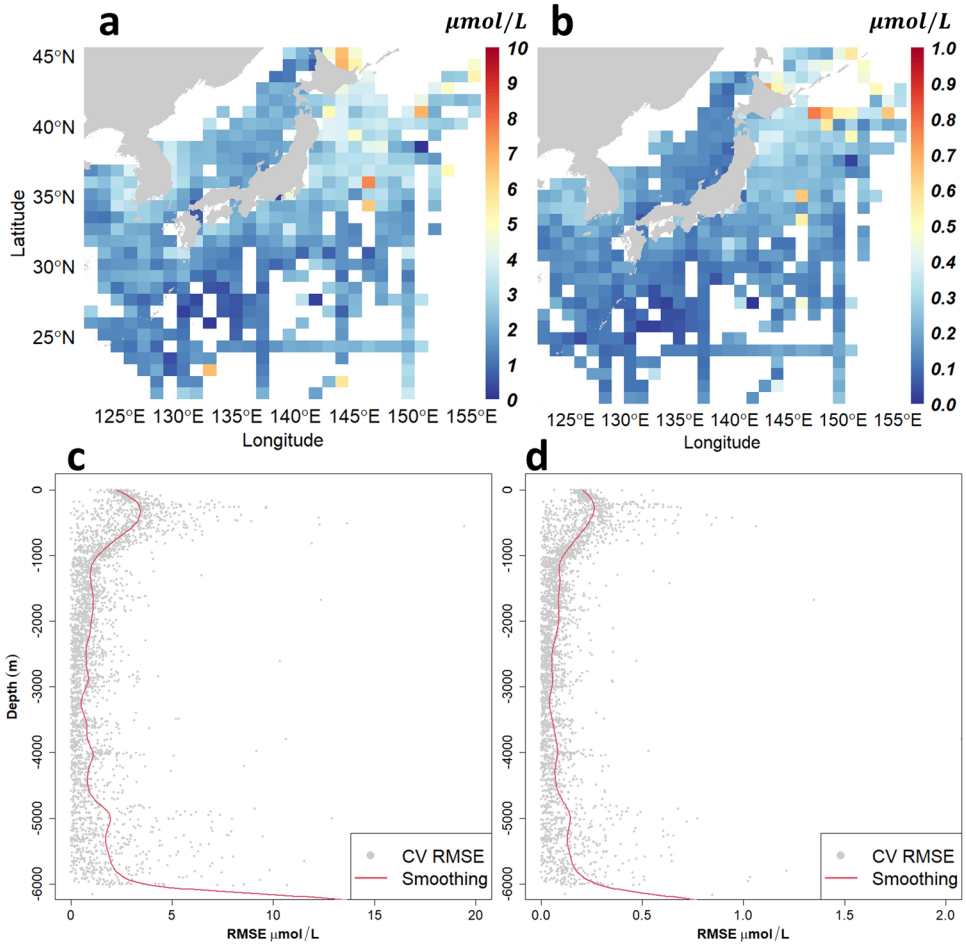

Fig. 4 Error distribution for horizontal (**a,b**) and vertical (**c,d**) directions. a and c represent the error distribution of nitrate concentration, while b and d represent the error distribution of phosphate concentration. In areas with high data density, errors are relatively low. The vertical distribution of errors shows an increasing trend in the thermocline (approximately 300 m).

are presented as supporting information. The estimation errors for water temperature were 2.05 °C (RMSE), 1.42 °C (MAE), and 0.93 (R^2_{adj}). The error metrics for nitrate concentration were 2.79 (RMSE), 1.84 (MAE), and 0.97 (R^2_{adj}), and those for phosphate concentration were 0.22 μmol/L (RMSE), 0.14 μmol/L (MAE), and 0.96 (R^2_{adj}). The spatial distribution of the errors showed that the area near Hokkaido in Japan had higher nutrient concentrations than other areas, with approximately 7 μmol/L for nitrate and approximately 0.8 μmol/L for phosphate(Fig. 4a,b).

The RMSE variation was analyzed according to depth(Fig. 4c,d). The RMSE was observed to be relatively higher within the 0–1000 m depth range, where the thermocline is located, but remained stable below a depth of 1000 m. However, an increase in RMSE was observed in the deep sea below 5000 m. The increased error in the thermocline and deep sea was attributed to the abrupt changes in nutrient concentration and lack of data, respectively.

Additionally, Compatibility with global-scale projects was assessed. The raw data utilized in this study was contrasted with the biogeochemical data product of GLODAPv2.2022(https://www.ncei.noaa.gov/data/oceans/ncei/ocads/data/0257247/)[59]. Since the data prior to 2010 was verified in previous research using CLIVAR and SIO datasets[11], the focus was on data from 2010 onwards. A total of 671 data points with precisely matching longitude, latitude, depth and time were compared(Fig. 5).

Subsequently, the data estimated by STK was also compared with the GLODAPv2.2022 data(Fig. 6). Among the gridded data from the period 2010–2019, grids that were spatiotemporally closest to certain GLODAPv2.2022 data were compared. The grid data closest to the GLODAPv2.2022 data were identified, and those within the 5% quantile distance were selected. The criteria for selection were a horizontal distance of approximately 15 km, a vertical distance of about 16 m, and a time difference within roughly 9 days. For NO3, 2,652 data points were contrasted, yielding an R^2_{adj} of 0.984 and a Residual Standard Error of around 2.03. For PO4, 2,676 data points were examined, with an R^2_{adj} of 0.981 and a Residual Standard Error of approximately 0.158.

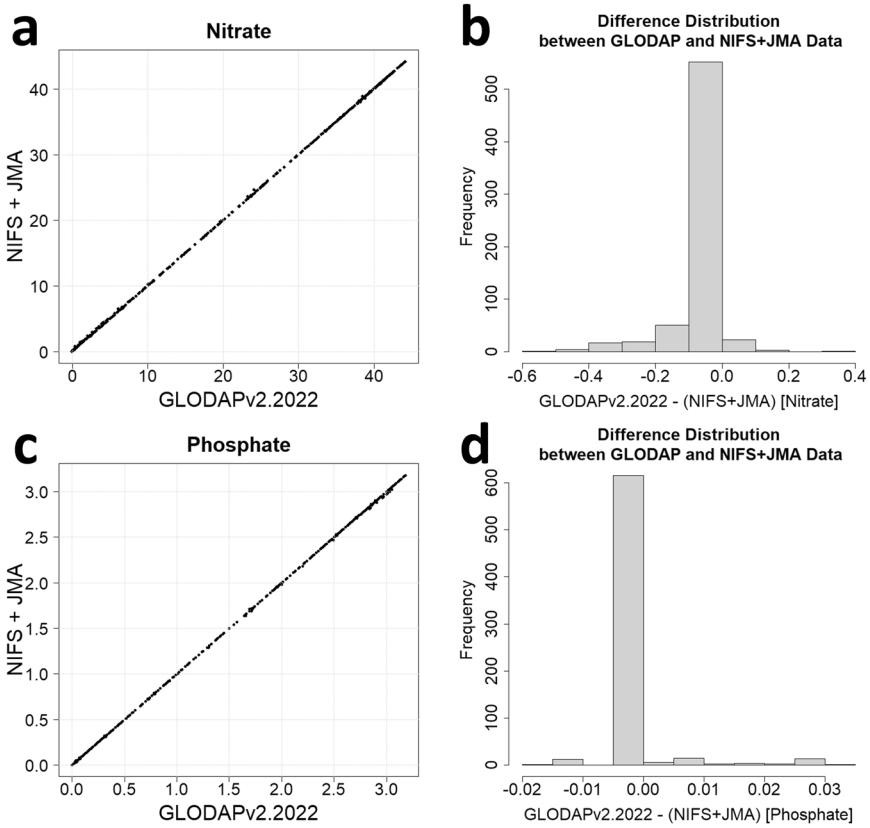

Fig. 5 A comparison of nutrient data used for STK estimation (from NIFS and JMA) with spatiotemporally matched cruise data from GLODAPv2.2022. The results of the comparison for nitrate (**a,c**) and phosphate (**b,d**) are presented, with the distribution of differences converging towards a central value of 0.

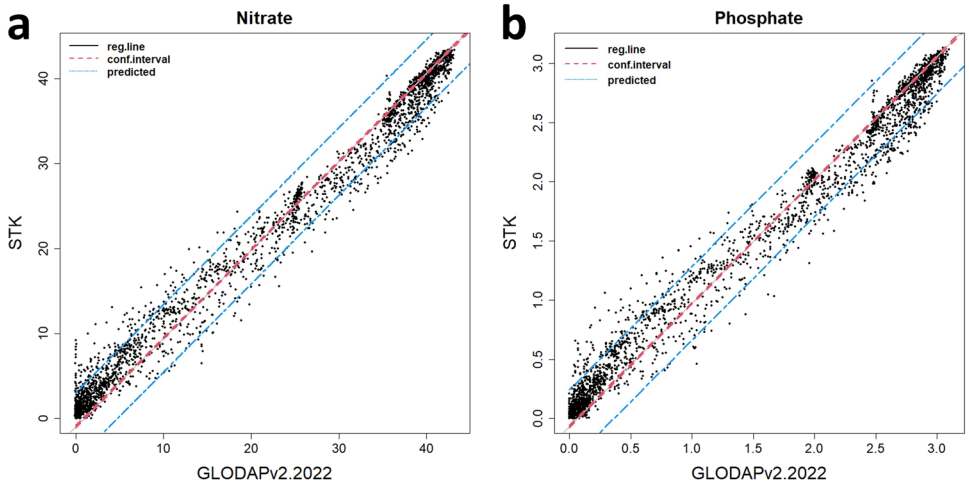

Fig. 6 Results from a comparison of nitrate (**a**) and phosphate (**b**) gridded data estimated by STK from 2010 to 2019 with GLODAPv2.2022 data. Information from the linear model is represented by the linear regression line (black solid line), 95% confidence interval (red dotted line), and 95% prediction interval (blue dotted line).

Usage Notes

This dataset was used to assess the nutrient dynamics in select areas of the northwest Pacific, both locally and regionally (Fig. 7). Gridded data can be examined through basic statistical analysis and spatial statistical methods such as EOF. These data can also be utilized for comparison with biogeochemical modeling outcomes. The

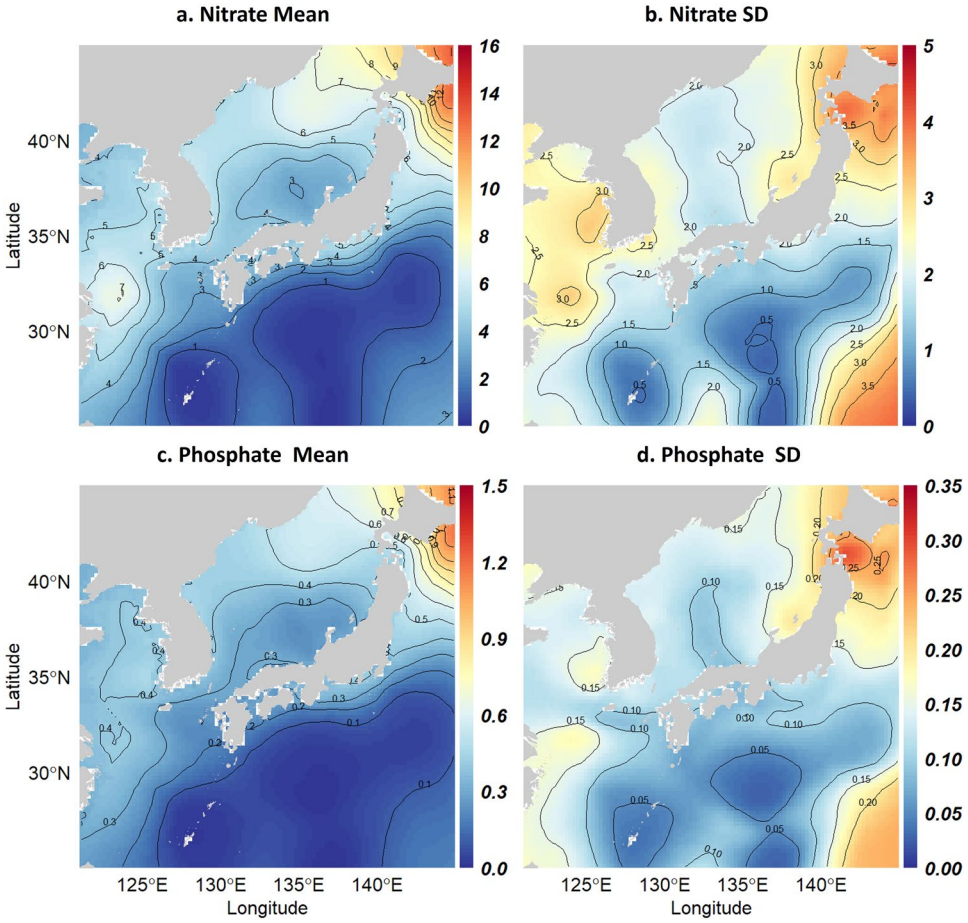

Fig. 7 Spatial distribution of the mean and standard deviation of nitrate (**a,b**) and phosphate (**c,d**) concentrations during the target period of estimation (1980–2019). The mean concentration and variability are highest in the Yellow Sea and near Hokkaido, Japan. The high standard deviation observed in the northwest Pacific off the southeast coast of Japan is attributed to the lack of data.

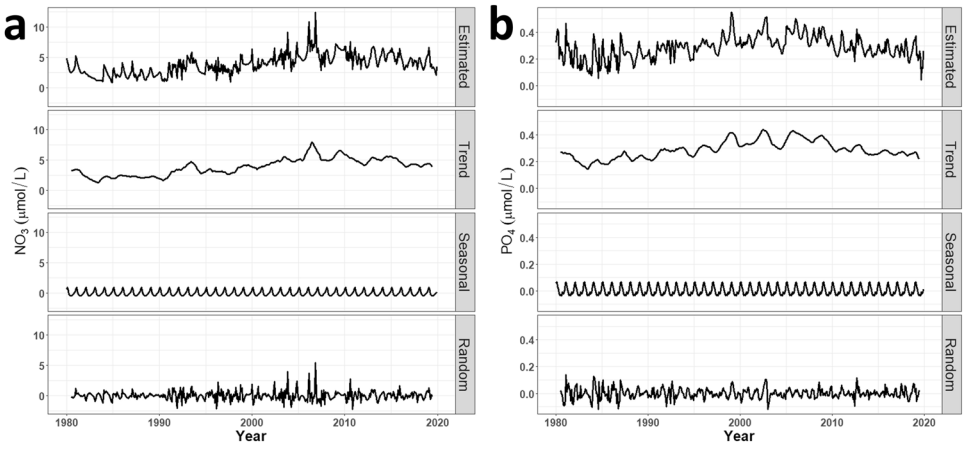

Fig. 8 The time series decomposition of the estimated nitrate (**a**) and phosphate (**b**) concentrations in the Yellow Sea region is displayed, where they were decomposed into trend, seasonal, and residual components. These outcomes can be utilized for research on the sources of variability and trend analysis.

surface (0–50 m averaged) nitrate concentration trend estimated in this study corroborates the decreasing nitrate concentration trend observed in previous studies[20,21] since approximately 2010 in the Yellow Sea (Fig. 8) .

	Originally Developed	Also Tested
OS	Windows 10 Enterprise	Linux Ubuntu 20.04
System	64 bit, x64 processor	
CPU	Intel i9-9900K 3.60 GHz 8 core 16 thread	Intel i9-10980XE 3.00 GHz 18 core 36 thread
RAM	64 GB	256 GB
R version	4.1.1	4.1.3

Table 4. The system environment used for development and testing is presented. The main development system for the *STK* library utilized a Windows operating system and was also tested on a Linux (Ubuntu) system.

Code availability

The R code scripts and dataset are available on 'Figshare' for reproducibility[54]. The author's GitHub online repository will be continuously updated to ensure sustainable usage of these codes(https://github.com/Gi-Seop/STK).

Tested system. All codes were tested in the following system environment (Table 4).

Received: 6 July 2023; Accepted: 27 September 2023;
Published online: 14 October 2023

References

1. Gregg, W. W., Conkright, M. E., Ginoux, P., O'Reilly, J. E. & Casey, N. W. Ocean primary production and climate: Global decadal changes. *Geophys Res Lett* **30**(15), 1809 (2003).
2. Joo, H. *et al*. Long-term pattern of primary productivity in the East/Japan Sea based on ocean color data derived from MODIS-aqua. *Remote Sens* **8**(1), 25 (2016).
3. Lee, S. H. *et al*. Seasonal carbon uptake rates of phytoplankton in the northern East/Japan Sea. *Deep Sea Res 2 Top Stud Oceanogr* **143**, 45–53 (2017).
4. Lin, C. L., Ning, X. R., Su, J. L., Lin, Y. & Xu, B. Environmental changes and the responses of the ecosystems of the Yellow Sea during 1976–2000. *J Mar Syst* **55**(3-4), 223–234 (2005).
5. Wang, Z., Qi, Y., Chen, J., Xu, N. & Yang, Y. Phytoplankton abundance, community structure and nutrients in cultural areas of Daya Bay, South China Sea. *J Mar Syst* **62**(1-2), 85–94 (2006).
6. Zhou, M. J., Shen, Z. L. & Yu, R. C. Responses of a coastal phytoplankton community to increased nutrient input from the Changjiang (Yangtze) River. *Cont Shelf Res* **28**(12), 1483–1489 (2008).
7. Fu, M., Wang, Z., Pu, X., Xu, Z. & Zhu, M. Changes of nutrient concentrations and N: P: Si ratios and their possible impacts on the Huanghai Sea ecosystem. *Acta Oceanol Sin* **31**(4), 101–112 (2012).
8. Fu, M. *et al*. Response of phytoplankton community to nutrient enrichment in the subsurface chlorophyll maximum in Yellow Sea Cold Water Mass. *Acta Ecol Sin* **36**(1), 39–44 (2016).
9. Yang, F., Wei, Q., Chen, H. & Yao, Q. Long-term variations and influence factors of nutrients in the western North Yellow Sea, China. *Mar Pollut Bull* **135**, 1026–1034 (2018).
10. Galloway, J. N. *et al*. Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science* **320**(5878), 889–892 (2008).
11. Kim, T. W., Lee, K., Najjar, R. G., Jeong, H. D. & Jeong, H. J. Increasing N abundance in the northwestern Pacific Ocean due to atmospheric nitrogen deposition. *Science* **334**(6055), 505–509 (2011).
12. Kim, I. N. *et al*. Increasing anthropogenic nitrogen in the North Pacific Ocean. *Science* **346**(6213), 1102–1106 (2014).
13. Liu, S. *et al*. Effects of anthropogenic nitrogen discharge on dissolved inorganic nitrogen transport in global rivers. *Glob Chang Biol* **25**(4), 1493–1513 (2019).
14. Shim, M. J. & Yoon, Y. Y. Long-term variation of nitrate in the East Sea, Korea. *Environ Monit Assess* **193**(11), 1–13 (2021).
15. Stramma, L. *et al*. Trends and decadal oscillations of oxygen and nutrients at 50 to 300 m depth in the equatorial and North Pacific. *Biogeosciences* **17**(3), 813–831 (2020).
16. Galloway, J. N. Nitrogen mobilization in Asia. *Nutr Cycl Agroecosyst* **57**(1), 1–12 (2000).
17. He, B. *et al*. Assessment of global nitrogen pollution in rivers using an integrated biogeochemical modeling framework. *Water Res* **45**(8), 2573–2586 (2011).
18. Kim, T. W. *et al*. Interannual nutrient dynamics in Korean coastal waters. *Harmful Algae* **30**, S15–S27 (2013).
19. Li, H. M., Tang, H. J., Shi, X. Y., Zhang, C. S. & Wang, X. L. Increased nutrient loads from the Changjiang (Yangtze) River have led to increased harmful algal blooms. *Harmful Algae* **39**, 92–101 (2014).
20. Wang, J., Yu, Z., Wei, Q. & Yao, Q. Long-term nutrient variations in the Bohai Sea over the past 40 years. *J Geophys Res Oceans* **124**(1), 703–722 (2019).
21. Wang, K. *et al*. Climate and human-driven variability of summer hypoxia on a large river-dominated shelf as revealed by a hypoxia index. *Front Mar Sci* **8**, 634184 (2021).
22. Reynolds, R. W. *et al*. Daily high-resolution-blended analyses for sea surface temperature. *J Clim* **20**(22), 5473–5496 (2007).
23. Banzon, V., Smith, T. M., Chin, T. M., Liu, C. & Hankins, W. A long-term record of blended satellite and *in situ* sea-surface temperature for climate monitoring, modeling and environmental studies. *Earth Syst Sci Data* **8**, 165–176 (2016).
24. Huang, B. *et al*. Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *J Clim* **34**, 2923–2939 (2020).
25. Chang, J. H., Hart, D. R., Munroe, D. M. & Curchitser, E. N. Bias Correction of Ocean Bottom Temperature and Salinity Simulations From a Regional Circulation Model Using Regression Kriging. *J Geophys Res Oceans* **126**(4), e2020JC017140 (2021).
26. Yasunaka, S. *et al*. Mapping of sea surface nutrients in the North Pacific: Basin-wide distribution and seasonal to interannual variability. *J Geophys Res Oceans* **119**(11), 7756–7771 (2014).
27. Yasunaka, S. *et al*. Long-term variability of surface nutrient concentrations in the North Pacific. *Geophys Res Lett* **43**(7), 3389–3397 (2016).
28. Yasunaka, S., Mitsudera, H., Whitney, F. & Nakaoka, S. I. Nutrient and dissolved inorganic carbon variability in the North Pacific. *J Oceanogr* **77**(1), 3–16 (2021).
29. Copernicus. *Global Ocean Biogeochemistry Hindcast* https://data.marine.copernicus.eu/products (2023).
30. R Core Team. R: A Language and Environment for Statistical Computing. (2023).
31. National Institute of Fisheries Science, Serial Oceanographic Observation. https://www.nifs.go.kr/kodc/eng/eng_coo_list.kodc (2023).
32. Japan Meteorological Agency. *Oceanographic and Marine Meteorological Observations by Research Vessels*. https://www.data.jma.go.jp/gmd/kaiyou/db/vessel_obs/data-report/html/ship/ship_e.php (2023).

33. NOAA National Centers for Environmental Information, ETOPO 2022 15 Arc-Second Global Relief Model. https://doi.org/10.25921/fd45-gt74 (2023).
34. NOAA National Oceanic and Atmospheric Administration, ETOPO (2022). https://www.ncei.noaa.gov/maps/grid-extract/ (2023).
35. Japan Meteorological Agency. *Data Report of Oceanographic Observations Special Issue*. https://warp.ndl.go.jp/info:ndljp/pid/11160873/www.data.jma.go.jp/gmd/kaiyou/db/vessel_obs/data-report/html/ship/efile_NoS2_e.html (2023).
36. Japan Meteorological Agency. *Description and format of data*. https://www.data.jma.go.jp/gmd/kaiyou/db/vessel_obs/data-report/html/ship/format_e.html (2023).
37. McDougall, T. J. & Barker, P. M. Getting started with TEOS-10 and the Gibbs Seawater (GSW) oceanographic toolbox. *Scor/Iapso WG* **127**, 1–28 (2011).
38. Kelley, D., Richards, C. & WG127 SCOR/IAPSO gsw: Gibbs Sea Water Functions. R package version 1.1-1. https://CRAN.R-project.org/package=gsw (2022).
39. Amante, C. & Eakins, B. W. ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis. *NOAA technical memorandum NESDIS NGDC-24, National Geophysical Data Center, NOAA* **10**(2009), V5C8276M (2009).
40. Pante, E. & Simon-Bouhet, B. marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS One* **8**(9), e73051 (2013).
41. Boyer, T. P., *et al* Technical Editor: Mishonov, A.V. World Ocean Database 2018. NOAA Atlas NESDIS **87** (2018).
42. Hengl, T., Heuvelink, G. B. M., Perčec Tadić, M. & Pebesma, E. J. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theor Appl Climatol* **107**(1), 265–277 (2012).
43. Gasch, C. K. *et al*. Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D+ T: The Cook Agronomy Farm data set. *Spat Stat* **14**, 70–90 (2015).
44. Gräler, B., Pebesma, E. J. & Heuvelink, G. B. Spatio-temporal interpolation using gstat. *R J* **8**(1), 204–218 (2016).
45. Park, N. W. Time-series mapping of PM10 concentration using multi-gaussian space-time kriging: a case study in the Seoul metropolitan area, Korea. *Adv Meteorol* **2016**, Article ID 9452080 (2016).
46. Pebesma, E. J. Multivariable geostatistics in S: the gstat package. *Comput Geosci* **30**(7), 683–691 (2004).
47. Pebesma, E. J. The role of external variables and GIS databases in geostatistical analysis. *Trans GIS* **10**(4), 615–632 (2006).
48. Pebesma, E. J. spacetime: Spatio-temporal data in R. *J Stat Softw* **51**(7), 1–30 (2012).
49. Matheron, G. Le krigeage universel (Universal kriging), Vol. 1. Cahiers du Centre de Morphologie Mathematique, Ecole des Mines de Paris, Fontainebleau, **83** (1969).
50. Hengl, T., Heuvelink, G. B. M., Stein, A. Comparison of kriging with external drift and regression kriging. Enschede, Netherlands: ITC (2003).
51. Bilonick, R. A. Monthly hydrogen ion deposition maps for the northeastern US from July 1982 to September 1984. *Atmos Environ* **22**(9), 1909–1924 (1988).
52. Snepvangers, J. J. J. C., Heuvelink, G. B. M. & Huisman, J. A. Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma* **112**(3-4), 253–271 (2003).
53. Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J Sci Comput* **16**(5), 1190–1208 (1995).
54. Lee, GS., Lee, JH. & Cho, HY. Spatio-Temporal Estimated Nutrient Data in Northwest Pacific and East Asian Seas, *Figshare*, https://doi.org/10.6084/m9.figshare.c.6634508.v1 (2023).
55. Pebesma, E. J. & Bivand, R. S. S classes and methods for spatial data: the sp package. *R news* **5**(2), 9–13 (2005).
56. Bivand, R. S, Pebesma, E. J., & Gómez-Rubio V. *Modelling areal data. In: Applied spatial data analysis with R*. (New York: Springer, 2013).
57. Pebesma, E. J. Simple features for R: standardized support for spatial vector data. R J., 10(1), 439 (2018).
58. Miles, J. *R-squared, adjusted R-squared. Encyclopedia of statistics in behavioral science*. (Wiley, 2005).
59. Lauvset, S. K. *et al*. GLODAPv2.2022: the latest version of the global interior ocean biogeochemical data product. Earth Syst. *Sci. Data* **14**(12), 5543–5572, https://doi.org/10.5194/essd-14-5543-2022 (2022).

Acknowledgements

The authors would like to acknowledge the officials of NIFS and JMA for providing nutrient data over several decades and officials of NOAA's ETOPO project for the bathymetry data. This research was financially supported by the Ministry of Oceans and Fisheries, Republic of Korea (PG53502).

Author contributions

Conception of the project: G.S. Lee. Contribution to R Script Writing: G.S. Lee, J.H. Lee. Manuscript Writing and Editing: G.S. Lee, H.Y. Cho. Final Manuscript Approval: All authors.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.Y.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023