Marine Pollution Bulletin 113 (2016) 216–223

Contents lists available at ScienceDirect

Marine Pollution Bulletin

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

Assessment of pollutant mean concentrations in the Yangtze estuary based on MSN theory

Jing Ren [a,b,1], Bing-Bo Gao [a,b,1], Hai-Mei Fan [c,*], Zhi-Hong Zhang [e], Yao Zhang [d], Jin-Feng Wang [b,f,*]

[a] Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China
[b] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[c] East China Sea Marine Environmental Monitoring Center, Shanghai, China
[d] National Marine Hazard Mitigation Service, State Oceanic Administration, Beijing, China
[e] Center for Environmental Risk and Damage Assessment, Chinese Academy for Environmental Planning, Beijing, China
[f] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China

A R T I C L E I N F O

Article history:
Received 11 March 2016
Received in revised form 7 September 2016
Accepted 11 September 2016
Available online 21 September 2016

Keywords:
MSN
Pollutant
Mean concentration
Variance of estimation error
Yangtze estuary

A B S T R A C T

Reliable assessment of water quality is a critical issue for estuaries. Nutrient concentrations show significant spatial distinctions between areas under the influence of fresh-sea water interaction and anthropogenic effects. For this situation, given the limitations of general mean estimation approaches, a new method for surfaces with non-homogeneity (MSN) was applied to obtain optimized linear unbiased estimations of the mean nutrient concentrations in the study area in the Yangtze estuary from 2011 to 2013. Other mean estimation methods, including block Kriging (BK), simple random sampling (SS) and stratified sampling (ST) inference, were applied simultaneously for comparison. Their performance was evaluated by estimation error. The results show that MSN had the highest accuracy, while SS had the highest estimation error. ST and BK were intermediate in terms of their performance. Thus, MSN is an appropriate method that can be adopted to reduce the uncertainty of mean pollutant estimation in estuaries.

1. Introduction

The water quality of the Yangtze estuary has been a substantial concern in recent years due to the aggravated pollution caused by increases in regular human activities, industrial discharge, and oil leakage events. According to the China Marine Environment Quality Bulletin, the water quality of the Yangtze estuary and Hangzhou bay has been listed as extremely unqualified among the nine major tidal estuaries in China. Because marine resources have been emphasized as a strategic resource for the national interest, the observation, assessment, and management of marine pollution has become critical (Floehr et al., 2015; Smith, 2003; Su et al., 2015).

Estuary water pollution involves eutrophication with various excessive nutrient components distributed by biochemical and physical processes (D. and D., 2004; Huang et al., 2006; Li et al., 2014; Zhang et al., 2007). Increasing numbers of studies have focused on the physical, biochemical, and coupled mechanisms involved in the distribution and transfer of pollutants (Shen et al., 2001; Sun et al., 2013; Wang et al., 2006; Chen et al., 2009). The spatial and temporal distributions of nutrients are related directly to the evolution of the contamination area and thus merit considerable attention from both academia and industry (Edmond et al., 1985; Gui lin et al., 2012; Pan and Shen, 2010; Zhang et al., 2011). However, statistical approaches adopted for the assessment of nutrient concentrations are rarely discussed, although accuracy in such estimates is highly desired. The most commonly adopted general-mean-value theory, which still remains useful, is partially violated under conditions of non-homogeneity. Furthermore, potential interdependence of the observed data of a variable in the same block area is rarely considered. Few efforts have been made to improve the statistical method itself for the estimation of the nutrient pollutant concentration, especially for the two major components, nitrate and phosphate.

Regional mean nutrient pollutant concentrations are important indicators in spatial and temporal variation analysis that decision-makers use most; these concentration estimations are usually generated by typical interpolation and statistical methods. However, estimation uncertainty remains an issue that cannot be avoided or neglected (Cambule et al., 2014; Liu et al., 2014; Murphy et al., 2010). In most cases, the field data of the surface seawater layer are collected at sampling gauge locations at certain depths. Thus, the estimation of average pollutant concentration is actually a process that uses limited or finite datasets to estimate the continuous area. Classical statistical methods and model-based inference are each able to handle this circumstance (Haining, 1988; J. and V., 2002; Matheron, 1963; Shimada and Taro, 2015; Wang et al., 2009). Classical statistical methods can achieve an

* Corresponding authors.
E-mail addresses: fhm@eastsea.gov.cn (H.-M. Fan), wangjf@lreis.ac.cn (J.-F. Wang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.marpolbul.2016.09.021

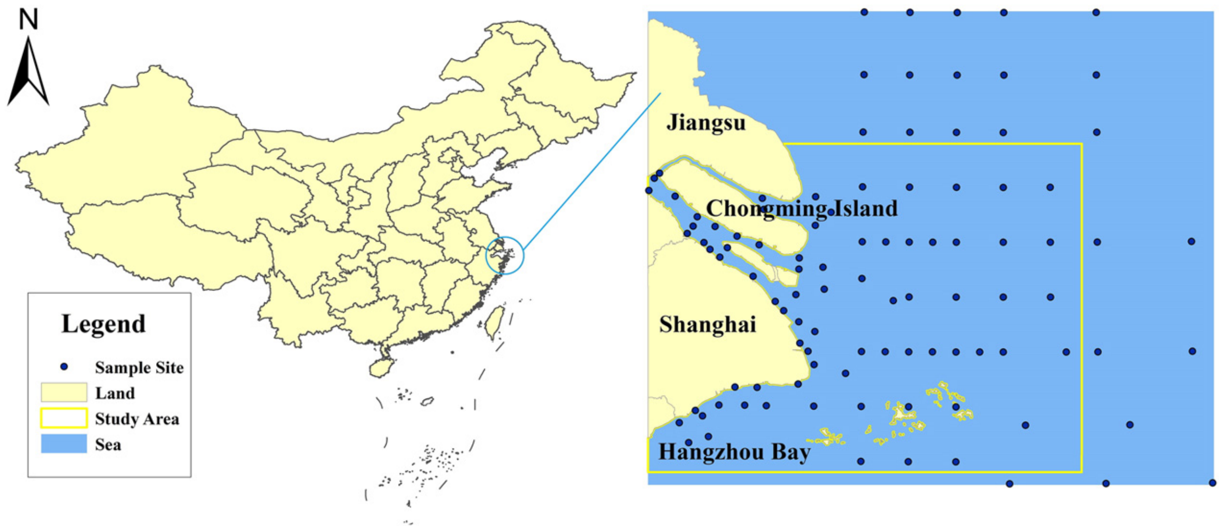


Fig. 1. Approximate position of the study area (circle of the left plot) and distribution map of monitoring sites.

unbiased estimation if the study area of interest approximately follows the independent identical distribution. Model-based inference methods provide more efficient estimation because they can guarantee an unbiased and optimal estimation, as both the spatial autocorrelation of target variables and minimum estimation variance are taken into account (Mishra et al., 2010; Thompson and Kolka, 2005). However, the homogeneous assumption for the survey region using this type of method strongly violates the stratified distribution in the real world and therefore hinders accurate assessment.

Non-homogeneity exists in a variety of natural phenomena and geophysical environments (e.g., moisture of soil, forest community) and is also inherent in estuary pollution because of complex physical and biochemical processes. Coastal water is affected by the impact from the land and sea, and because of hydrological and hydrodynamic conditions, changes within a small spatial scale can be substantial. In other words, it is a heterogeneous region (Hu, 1995). Wang et al. proposed the mean of surface with heterogeneity (MSN) method, which could provide preferable solutions to the non-homogeneous area and yield an optimized linear unbiased mean estimation (Wang et al., 2010a). Partition would be a reasonable way to divide a nonhomogeneous region into sub-areas that could be treated approximately as homogenous pieces. For the partition methods, different empirical or mathematical statistical approaches are applied according to various study targets (e.g., evolution stage, biological environment, nutrient salt, sediment-based and cluster-analysis-based estuary stratification) (Liu et al., 2011; Zhu et al., 2008). In this paper, a new assessment system is introduced and applied to reduce uncertainty regarding the nutrient pollutant's mean concentration. A hybrid-distance-based SOFM cluster method was adopted to stratify the Yangtze estuary. Then, the MSN method was used to calculate the mean pollutant concentration for each element for the whole study area. The results from the commonly used methods, block Kriging (BK) mean estimation (Kern and Coyle,

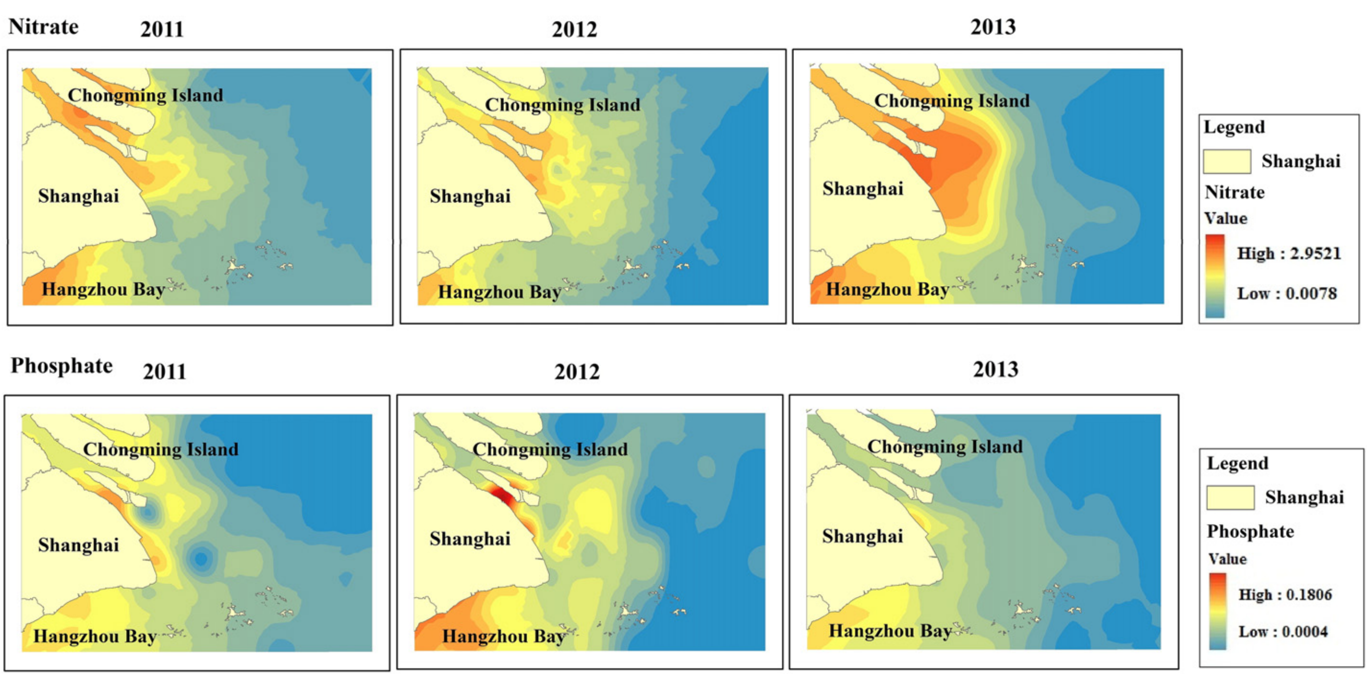


Fig. 2. Concentration (mg/L) map of nitrate (above) and phosphate (below) from 2011 to 2013. Red color represents high concentration; light blue color represents low concentration. The maximum and minimum concentrations are given in the legend.

Table 1
Sample data from 2011 to 2013.

	Nitrate (mg/L)				Phosphate (mg/L)			
Year	Max	Min	Mean	CV (%)	Max	Min	Mean	CV (%)
2011	2.9521	0.0288	0.9418	78.67	0.1163	0.0011	0.0407	82.44
2012	2.4067	0.0060	0.8600	85.57	0.1806	0.0008	0.0395	84.69
2013	2.6004	0.0078	1.0818	81.29	0.0898	0.0004	0.0324	76.98

2000), stratified sampling (ST) inference, and simple random sampling (SS) inference, were compared to those of MSN to determine if MSN is a preferable way to estimate mean concentrations of marine pollutants of estuary.

2. Data and methodology

The Yangtze estuary, located between Shanghai and Jiangsu in the most important economic area of China's east coast, provides the regional industrial and agricultural water sources, ecological character, and navigation, as well as pollution discharge. The estuary is 167 km long and is trumpet-shaped. The width of the narrow section is 5.6 km, while the wide mouth spans up to 90 km. Data were provided by East China Sea Marine Environmental Monitoring Center and were collected each season. Monitoring indicators cover a broad range of parameters, such as pH, dissolved oxygen (DO), chemical oxygen demand (COD), phosphate (PO_4-P), nitrate (here we refer to as nitrate nitrogen NO_3-N +, nitrite nitrogen NO_2-N +, and ammonia nitrogen NH_3-N), temperature, salinity, petroleum, mercury (Hg), copper (Cu), lead (Pb), cadmium (Cd), arsenic (As), and chlorophyll. We used the sample data for August from the years 2011 to 2013 and considered only two main contaminants, nitrate and phosphate. According to China's national criterion, GB17378.4-2007, specified methods to evaluate the concentration of nitrate nitrogen, nitrite nitrogen, and ammonia nitrogen were the Zn-Cd reduction method, naphthyl ethylenediamine dihydrochloride spectrophotometric method, and hypobromite oxidation method. The study area and the distribution of sample sites are shown in Fig. 1. The total numbers of samples for August for years 2011 to 2013 were 101, 100 and 102, respectively.

2.1. Self-organizing feature map (SOFM)

The SOFM was presented by Kohonen (Kohonen, 1998) and is an unsupervised clustering method that organizes the input data through a self-learning procedure achieved by competitive learning. The difference between SOFM and a general neural network is that SOFM also adjusts the weight of adjacent neurons next to winner neurons. The underlying theory is based on the inherent topological structure exhibited by the human brain. The result is that the input vectors next to each other will acquire a similar topological structure in terms of the output layer.

It is advantageous to divide the study area into several sub-areas so that the dispersion variance within a stratum is small but the difference between sub-areas is large. In this study, a hybrid distance based on the SOFM clustering method is used to stratify the study area (Jiao et al., 2011). Hybrid distance refers to spatial distance and attribute distance. The equation is defined as follows:

$$D_{ij} = w_s D_{ij}^{(S)} + w_a D_{ij}^{(A)} \tag{1}$$

where D_{ij} is the hybrid distance, $D_{ij}^{(s)}$ represents spatial distance and $D_{ij}^{(A)}$ is the attribute distance, and w_s and w_a refer to the weight for space and attribute, respectively.

The advantage of using the hybrid distance is to provide a balance between spatial distance and attribute distance. This approach attempts to enhance the cohesion in the attribute domain and ensure continuity

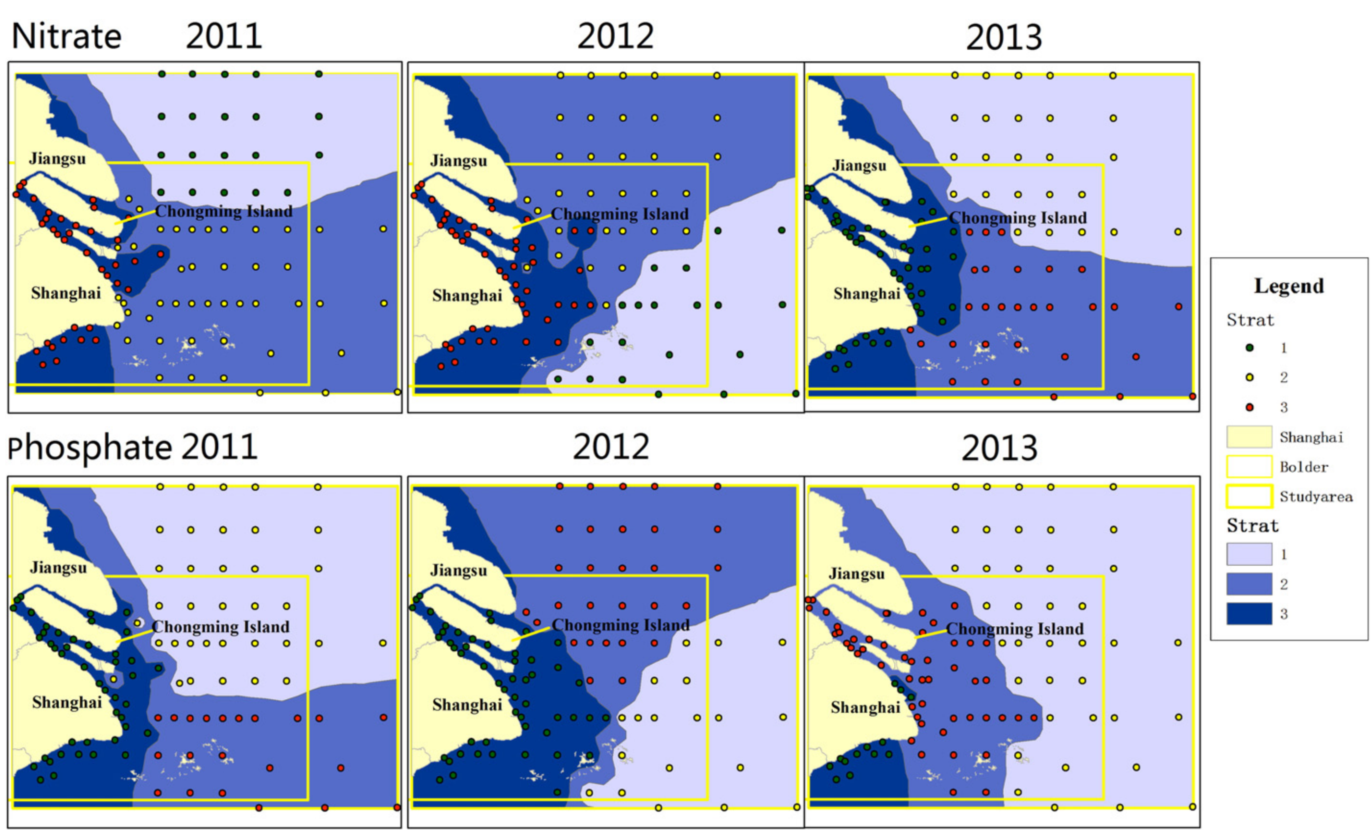


Fig. 3. Stratification result of nitrate and phosphate from 2011 to 2013. The stratum near the river mouth has high concentrations, the stratum at the open sea has low concentrations, and the remaining part is a stratum with medium concentration.

in the spatial domain. The implementation procedure for the hybrid-distance-based SOFM involves the following steps:

1) Data standardization. Because hybrid distance involves two aspects, spatial and attribute domain, the space coordinates are treated as a special variable. The measurement units for the variables are different, so standardization is necessary.

$$y_i = \frac{y_i' - y_{min}}{y_{max} - y_{min}} \tag{2}$$

2) Initiating the weight vector. The output node is assigned a minor but random number, no >0.5, as the initial weight $w_{ji} \le 0.5$, $i = 1, \cdots p$. The initial neighborhood is defined as $N_c(0)$, the learning rate as $\eta(0)$, and the maximum iterations as T; additionally, $t = 0$.
3) Calculating the best matching node. After initialization, an input vector $Y(t) = \{y_1(t), y_2(t) \cdots y_p(t)\}$ is chosen; then, the node that has the minimum hybrid distance between the input vector and output weight vector is identified. Subsequently, the weights of the winner node and its adjacent nodes are adjusted according to equation (4), and the learning rate $\eta(t)$ and neighborhood $N_c(t)$ are updated according to equations (5) and (6).

$$\|Y(t) - W_C(t)\| = min\{\|Y(t) - W_j(t)\|\}^{hybrid} \tag{3}$$

$$\Delta w_j(t) = \begin{cases} \eta(t)[y(t) - w_j(t)], & j \in N_c(t) \\ 0, & j \notin N_c(t) \end{cases} \tag{4}$$

$$\eta(t) = \eta(0)\left(1 - \frac{t}{T}\right) \tag{5}$$

$$N_c(t) = INT\left[N_c(0)\left(1 - \frac{t}{T}\right)\right] \tag{6}$$

4) The learning process is repeated until the iterations $t \ge T$. As the learning process approaches convergence, the learning rate will decrease to a minimum of zero, and the neighborhood will narrow to a very small area.

2.2. MSN

Through stratification, a non-homogeneous area is divided into smaller homogeneous sub-strata. A sub-stratum is a homogeneous area that meets the stationary hypothesis. For a homogeneous stratum, the mean value can be calculated by a weighted sample mean, and the assigned weight must meet certain conditions (optimal and unbiased). For a non-homogeneous area, the mean value can be calculated by a weighted homogeneous strata mean. To estimate the global mean with the MSN method, two steps are required: (1) the non-homogeneous surface dataset R is decomposed into a set of spatially homogeneous sub-strata R_h, and each subset surface is defined by $\boldsymbol{y}(\boldsymbol{s})$:

$$\mathrm{E}[y(s)|s \in R_h] = C \tag{7}$$

Table 2
PD values and P values of nitrate and phosphate from 2011 to 2013.

	Nitrate			Phosphate		
PD value	0.87	0.77	0.89	0.87	0.66	0.91
P value	0.00	0.00	0.00	0.00	0.00	0.00

Table 3
Parameters of semi-variograms of nitrate concentration from 2011 to 2013.

Time	Stratum	Nugget (C0) 10^{-3}	Sill (C0+C) 10^{-3}	C0/(C0+C) %	Range	Variance 10^{-2}
2011	Global	0.2140	1.4650	14.6	170,882	0.60
	Stratum1	0.0000	0.0040	0.0	36,963	0.13
	Stratum2	0.0480	0.1371	35.0	154,600	0.08
	Stratum3	0.0000	0.1203	0.0	13,600	0.003
2012	Global	0.2178	0.9115	23.9	139,172	0.54
	Stratum1	0.0000	0.0244	0.0	21,053	0.04
	Stratum2	0.0280	0.0723	38.7	123,100	0.22
	Stratum3	0.1157	0.2807	41.2	38,319	0.03
2013	Global	0.0930	1.4790	6.3	133,100	0.77
	Stratum1	0.0000	0.0156	0.0	35,982	0.01
	Stratum2	0.0137	0.3533	3.9	118,879	0.14
	Stratum3	0.0042	0.1693	2.5	34,513	0.09

The surface average over R is defined as:

$$\overline{Y_R} = R^{-1} \int_R R_h \overline{Y}_h d_s \tag{8}$$

$$a_h = R_h R^{-1} \tag{9}$$

where a_h is the area ratio of strata R_h, $\overline{Y}_h$ is the average of each sub-stratum and can be estimated by a weighted sample mean within the homogeneous strata R_h, $\overline{y}_h$ is the unbiased estimation of $\overline{Y}_h$, and equation (11) must be satisfied.

$$\overline{y}_h = \sum_{i=1}^{n_h} w_{hi} y_{hi} \tag{10}$$

$$\sum_{i=1}^{n_h} w_{hi} = 1 \tag{11}$$

From the above, the surface average can be calculated as follows:

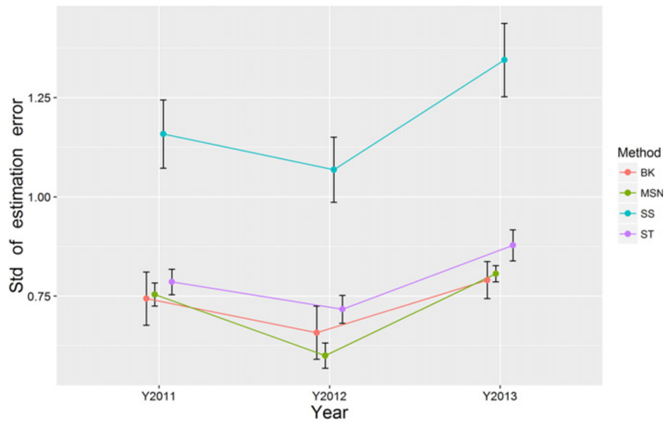
$$\overline{y}_h = R^{-1} \sum_{h=1}^{H} R_h \sum_{i=1}^{n_h} w_{hi} y_{hi} \tag{12}$$

The concentration is the calculation of weight, which insures the condition of equation (13) and minimizes the mean squared estimation error.

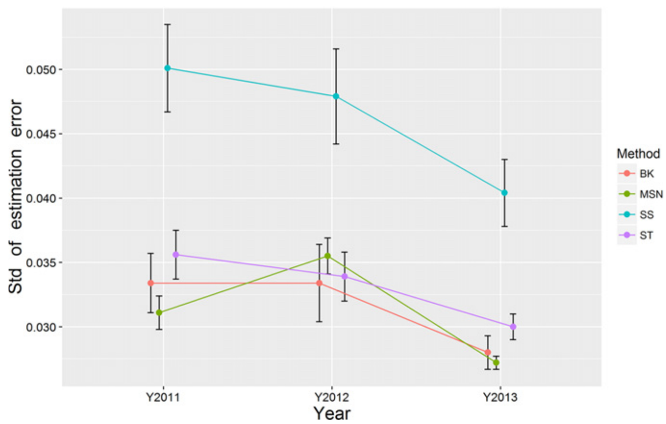
$$\sigma_R{}^2 = E[\overline{y}_R - \overline{Y}_R]^2 \tag{13}$$

Table 4
Parameters of semi-variograms of phosphate concentration from 2011 to 2013.

Time	Stratum	Nugget (C0) 10^{-3}	Sill (C0 + C) 10^{-3}	C0/(C0 + C) %	Range	Variance 10^{-2}
2011	Global	0.2330	2.2860	10.2	164,700	0.102
	Stratum1	0.0001	0.0643	0.2	16,500	0.015
	Stratum2	0.0780	0.8070	9.7	149,000	0.024
	Stratum3	0.0012	0.1814	0.7	14,200	0.007
2012	Global	0.4290	2.5880	16.6	188,100	0.112
	Stratum1	0.0001	0.0370	0.3	24,745	0.065
	Stratum2	0.0536	0.1637	32.7	16,384	0.003
	Stratum3	0.0000	0.7244	0.0	15,917	0.014
2013	Global	0.0783	2.0153	3.9	207,163	0.062
	Stratum1	0.0001	0.0339	0.3	32,000	0.005
	Stratum2	0.0241	0.1162	20.7	50,000	0.008
	Stratum3	0.0000	0.0481	0.0	8834	0.003



(a) Mean estimation result of nitrate



(b) Mean estimation result of phosphate

Fig. 4. Mean estimation result (lines of different colors represent different methods, the corresponding point on the line represents the estimated mean for the year, error bars for each point reflect the estimation error).

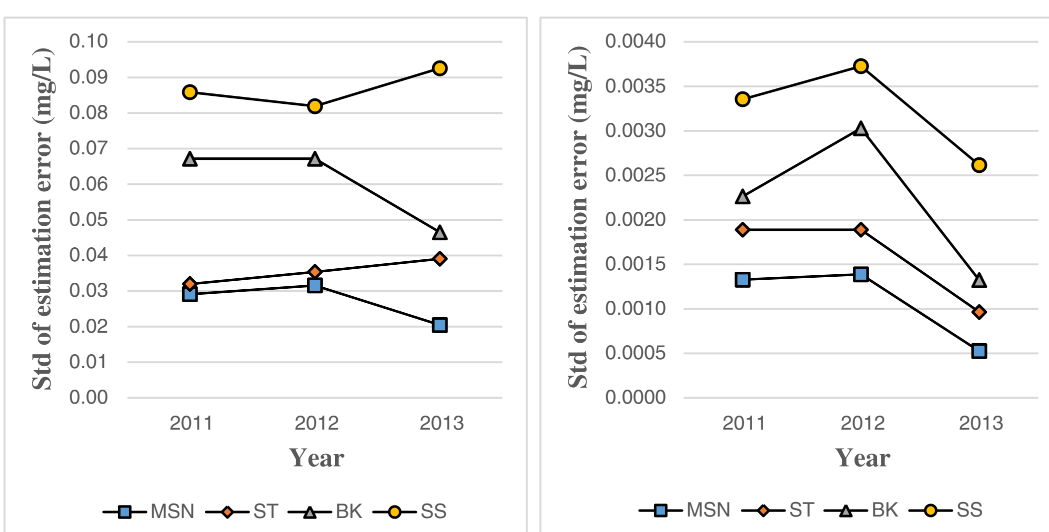
3. Results and analysis

A pollutant concentration distribution map obtained by the interpolation method is shown in Fig. 2. It is clear that the pollutant concentration near Chongming Island and Hangzhou Bay, where human activities are abundant, is higher than that of the open sea areas. The map also shows high spatial heterogeneity for both nitrate and phosphate. The statistical data are listed in Table 1, and the maximum concentrations for nitrate for years 2011 to 2013 are 2.95, 2.40, and 2.60 mg/L, respectively, which is far beyond the fourth grade of water quality according to GB3097-1997. Hangzhou Bay presents a similar situation to the Shanghai estuary. For phosphate, the maximum values are 0.11, 0.18, and 0.08 mg/L, respectively, all higher than the standard value of 0.045 mg/L. This imbalance may occur because the four sewage outlets of Shanghai are distributed on the southern passage of the Yangtze estuary, where the pollutant concentration is distinctly higher than in the other districts. Nantong in Jiangsu province is an industrially developed city, and there are six industrial parks in Hangzhou Bay, which results in perennial pollution from the intensive sewage outlets (Liu et al., 2003; Sun et al., 2009).

3.1. Stratification result evaluation

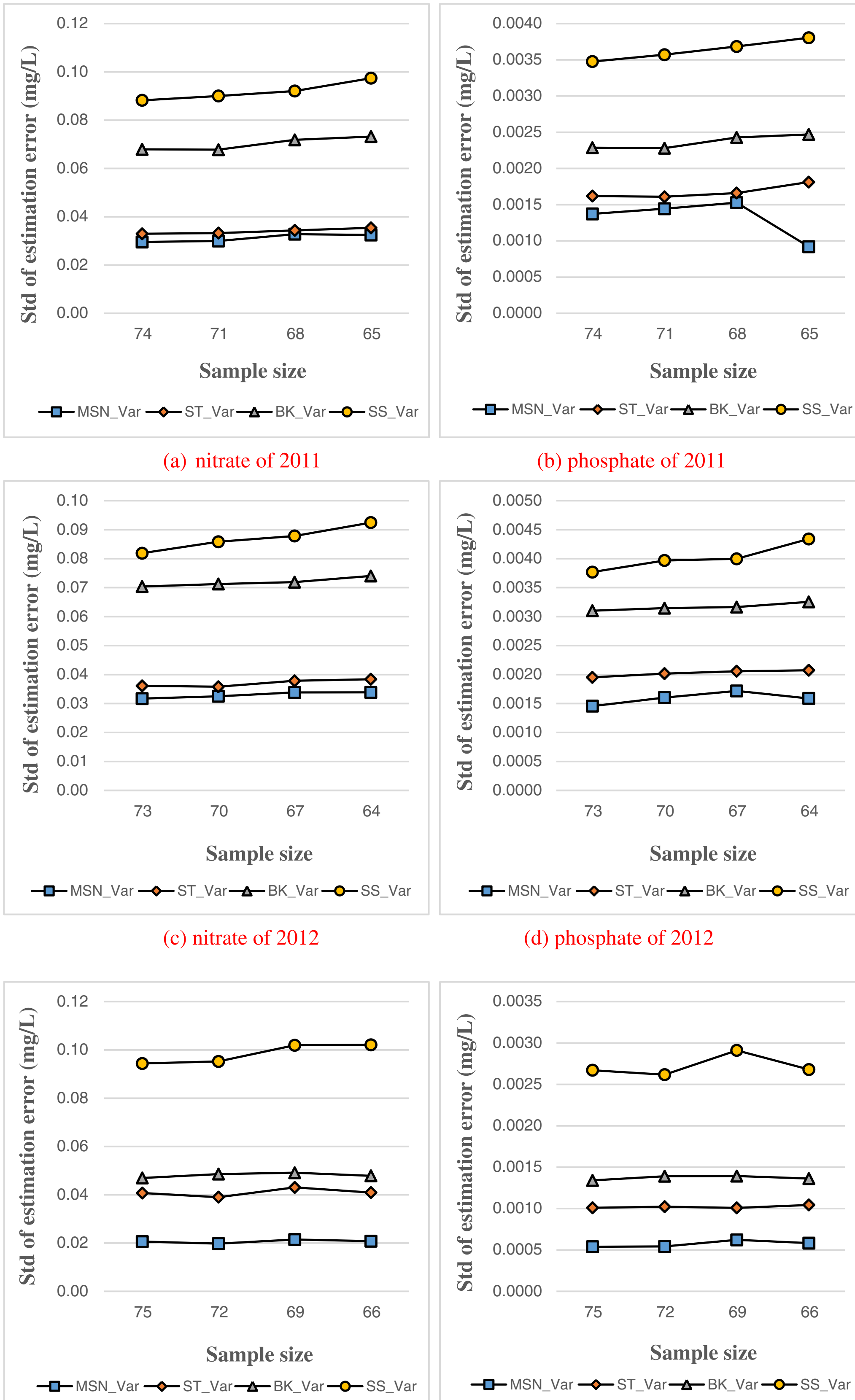
The pollutants drift and evolve in the sea, and the contamination situation changes every year. To accurately estimate the concentration, partition of the monitoring data is conducted every year. The SOFM process produces several point clusters, but Yangtze estuary is represented as a continuous surface. A Voronoi diagram can be used to obtain the zoning polygons (Jiao et al., 2011), but it does not perform well near the boundary; therefore, IDW is used here to obtain different zones according to the aggregate result. The entire process was performed using the software Matlab 7.8.0 and ArcGIS 10.1.

The area that contains all the sample data was divided into three strata according to the SOFM clustering result, as shown in Fig. 3. The yellow line shows the portion of the study area for which the mean concentrations were to be estimated. The monitoring data outside the study area are only employed for stratification and variogram estimation; they are not used for estimating the mean concentrations within the study area. The stratum near the river mouth has a high concentration, the stratum at the open sea has a low concentration, and the remaining part is a stratum with medium concentration. As shown in the legends of Fig. 3, darker color shades indicate more serious pollution. In general, the stratification boundary of nitrate is more stable than that of phosphate for the years 2011 to 2013. The stratification result can better represent the diffusion pattern of contamination every year. The most



(a) nitrate (b) phosphate

Fig. 5. Standard deviation of estimation errors for different methods.



(a) nitrate of 2011 (b) phosphate of 2011

(c) nitrate of 2012 (d) phosphate of 2012

(e) nitrate of 2013 (f) phosphate of 2013

Fig. 6. Standard deviations of estimation error of nitrate and phosphate by different methods after removing 3, 6, 9, or 12 sites from the monitoring data.

polluted district is the estuary near Chongming Island, and the contamination scope is different for nitrate and phosphate from 2011 to 2013.

In an optimum stratification, the variance within each stratum is as small as possible, and the difference between strata is as large as possible. Since the aim of stratification is to estimate the mean value more accurately, it is necessary to evaluate the stratification result. GeoDetector (Wang et al., 2010b) was initially proposed to detect the contributed environmental risk factor for a certain health effect (i.e., mortality rate for a particular disease). It also can be used to detect the stratification effect. In GeoDetector, the power of determinant (PD) value is defined as follows:

$$PD = 1-\frac{1}{n\sigma^2}\sum_{h=1}^{L} n_h\sigma_h^2 \tag{14}$$

where σ_h^2 represents the variance of each sub-stratum and σ^2 is the variance of the entire study area, n_h is the number of samples in the sub-stratum, and n is the total sample number. A satisfactory division is one in which the variance for each sub-stratum σ_h^2 is close to 0 and the PD value approaches 1. The calculations were performed using free software downloaded from www.sssampling.org. The results are listed in Table 2. The PD values from 2011 to 2013 are 0.87, 0.77, 0.89 for nitrate, and 0.87, 0.66, and 0.91 for phosphate, respectively. From the PD value, one can observe that the stratification carried out by SOFM is good. In general, 2013 is associated with the best stratification result, and the PD value for 2012 is relatively low. This result may arise because the dispersion variance and the weight for stratum1 is higher than for the other years for this stratum. The newly updated Geo-detector (Wang et al., 2016) can also test the significance of the stratification (P-values in q statistic). A P-value < 0.01 represents a statistical significance at the $\alpha = 0.01$ level. From Table 2, it can be observed that the P-values of all stratifications are equal to approximately 0, which demonstrates that the stratified heterogeneity is significant and that the stratifications are reasonable.

3.2. Experimental variograms

The theoretical basis for Kriging is the correlation within a certain distance; a semi-variogram is used in geostatistics to model this correlation. The semi-variogram is defined under the condition of second-order stationary behavior; it reflects the spatial variance of a regionalized variable within a distance h. Optimal (model with a least residual sum of squares RSS and maximum coefficient of determinant R^2) local and global experimental variograms (variogram models within and between strata) of nitrate and phosphate for each year were fitted using the Gaussian model and are listed in Tables 3 and 4.

The nugget of the semi-variogram is caused by random variation and the ratio of nugget to sill can reflect the degree of spatial correlation. A ratio larger than 75% suggests that variability of the target variable is caused by random factors and the spatial correlation is weak, while a ratio <25% indicates strong spatial correlation of the target variable. When the ratio is between 25% and 75%, moderate spatial correlation is suggested. The ratios for nitrate and phosphate are listed in Tables 3 and 4. The ratio of nugget to sill for nitrate in stratum 2 in 2011, for nitrate in strata 2 and 3 in 2012, and for phosphate in stratum 2 in 2012 are between 25% and 75%; i.e., they exhibit moderate spatial correlation. Except for the above strata, all the remaining strata have ratios below 25%, indicating strong spatial correlation.

The variances of nitrate and phosphate in the study area and the strata are also listed in Tables 3 and 4, respectively. The variance of each stratum is smaller than that of whole area, suggesting that the stratification is efficient.

3.3. Mean value estimation and estimation result evaluation

Mean values for the study area (within the yellow lines) based on MSN theory were calculated using programs written in the R language. The estimation results of the mean values obtained by MSN, BK, SS and ST are shown in Fig. 4. Error bars were used to visualize the estimated mean values and the corresponding uncertainty. Lines of different colors represent different methods, and the points on the lines are the estimated means. The SS inference mean value is the highest compared with ST inference, BK and MSN methods every year. For nitrate, 2012 has the lowest and 2013 has the highest mean value in the three years, implying that the water quality improved from 2011 to 2012 and dropped, to a certain extent, in 2013. For phosphate, 2013 is associated with the lowest mean value. For different methods, the calculated mean values of phosphate do not show a significant difference except for the simple sampling inference, which still yields a relatively high mean value. Since the actual mean value is unknown, the performance of sea surface pollutant mean estimations by different methods was evaluated by the standard deviation of estimation error. The error bars in Fig. 4 span one standard deviation of estimation error above and below the mean point and indicate the annual degree of error. From the figure, it can be observed that the MSN method performs better than other methods.

To compare the performance of the different methods more clearly, the standard deviations of estimation error for the different methods are shown in Fig. 5. In general, for both nitrate and phosphate, the MSN mean estimation method performs better than the other methods, as indicated by a lower estimation error. The estimation error for ST of nitrate in 2011 and 2012 is close to that of the MSN method, followed by BK. The simple sampling inference method has the highest estimation error. Compared with ST inference, MSN reduced the standard deviation of estimation error for nitrate by 9%, 11%, and 48% from 2011 to 2013. For phosphate, MSN reduced the standard deviation estimation error by 30%, 27%,and 46% from 2011 to 2013. Compared with BK, MSN reduced the estimation error for nitrate by 57%, 53%, 56% and for phosphate by 41%, 54%, 60% from 2011 to 2013.

To test the performance of MSN further, 3, 6, 9, or 12 sites were randomly removed from the monitoring data for the three years, and the means and standard deviations of estimation errors were calculated for the remaining data by the different methods. The standard deviations of estimation error result are plotted in Fig. 6. For each method, a reduction of the number of monitoring sites is accompanied by an increase in the standard deviation of estimation error, although these increases were relatively small. The performances of the different methods are similar to the results with all monitoring data. The MSN is still superior to the other methods in reducing the standard deviation of the estimation errors. The results demonstrate that the MSN mean estimation method can reduce the estimation error effectively compared with traditional and frequently used methods, mainly because it takes the sea-surface non-homogeneity into account while simultaneously providing the best linear unbiased estimation.

4. Conclusion

In the present research, the sea-surface non-homogeneity and interdependence of observed data were studied thoroughly to estimate the nitrogen and phosphate concentrations of the study area of the Yangtze estuary. To make an accurate estimation, stratification was conducted each year for both pollutants. After stratification, experimental variogram models of the whole study area and each stratum were fitted. With the stratification result and experimental variogram models in hand, MSN theory could then be used to estimate the mean values of the pollutant concentrations.

To evaluate the performance of MSN, the estimation results were compared with those of traditional and frequently used methods. For both nitrate and phosphate, the MSN mean estimation method performed better than the other methods, with the lowest standard deviation of estimation error. The estimation ability of these methods, gauged by their order of estimation error, is MSN > ST > BK > SS. The standard deviations of estimation error for MSN were 9%, 11%, and 48% less than those for ST from 2011 to 2013 for nitrate. Likewise, MSN reduced the standard deviation estimation error for phosphate by 30%, 27%, and 46% from 2011 to 2013. Compared with BK, MSN reduced 57%, 53%, and 56% of the estimation error for nitrate and 41%, 54%, and 60% for the estimation error for phosphate from 2011 to 2013. The use of different sample sizes also supports this conclusion. Therefore, the MSN method is more reliable for estimating mean pollutant concentrations when the sea surface is non-homogeneously spatially stratified.

Acknowledgements

This study was supported by funding from the following grants: NSFC (41271404, 41601425), MOST (2012CB955503). The authors thank the East China Sea Environmental Monitoring Center of China for the marine environment monitoring data.

References

Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., Smaling, E.M.A., 2014. Soil organic carbon stocks in the Limpopo National Park, Mozambique: amount, spatial distribution and uncertainty. Geoderma 213, 46–56.

Chen, S.L., Yan, S.Z., Li, Y.Z., 2009. Characteristics of surface sediment distribution in the Yangtze estuary and its adjacent waters. Resour. Environ. Yangtze Basin 18, 152–156.

D., D., 2004. Ocean pollution from land-based sources: East China Sea, China. Ambio 33, 107–113.

Edmond, J.M., Spivack, A., Grant, B.C., Hu, M.H., Chen, Z., Chen, S., Zeng, X., 1985. Chemical dynamics of the Changjiang estuary. Cont. Shelf Res. 4, 17–36.

Floehr, T., Scholz-Starke, B.R., Xiao, H., Koch, J., Wu, L., Hou, J., Wolf, A., Bergmann, A., Bluhm, K., Yuan, X., 2015. Yangtze three gorges reservoir, China: a holistic assessment of organic pollution, mutagenic effects of sediments and genotoxic impacts on fish. J. Environ. Sci. 38, 63–82.

Gui lin, L.I., Jiang, L., Zhao, Y., 2012. Inter-annual variations and current situation of water quality in Lianyungang inshore. J. Huaihai Inst. Technol.

Haining, B.R., 1988. Estimating Spatial Means with an Appliscation to remotely Sensed Data, Communications in Statistics: Theory and Methods.

Hu, H., 1995. Water types and frontal surface in the Changjinag estuary. J. Fish. Chin.

Huang, B., Shi, X., Yu, D., Öborn, I., Blombäck, K., Pagella, T.F., Wang, H., Sun, W., Sinclair, F.L., 2006. Environmental assessment of small-scale vegetable farming systems in peri-urban areas of the Yangtze River Delta Region, China. Agric. Ecosyst. Environ. 112, 391–402.

J., V., 2002. Sampling and geostatistics for spatial data. Ecoscience 9, 152–161.

Jiao, L.M., Liu, Y.L., Zou, B., 2011. Self-organizing dual clustering considering spatial analysis and hybrid distance measures. Sci. Chin. Earth Sci. 54, 1268–1278.

Kern, J.W., Coyle, K.O., 2000. Global block kriging to estimate biomass from acoustic surveys for zooplankton in the western Aleutian Islands. Can. J. Fish. Aquat. Sci. 57, 2112–2121.

Kohonen, T., 1998. The self-organizing map. Neurocomputing 78, 1–6.

Li, H.M., Tang, H.J., Shi, X.Y., Zhang, C.S., Wang, X.L., 2014. Increased nutrient loads from the Changjiang (Yangtze) River have led to increased harmful algal blooms. Harmful Algae 39, 92–101.

Liu, C., He, Y., Lee, J.H.-w., Wei, H.P., 2003. Numerical simulation on pollutant tracking released from the sewerage outfalls in Shanghai. Shuili Xuebao 114–118.

Liu, L., Zheng, B., Meng, W., Tang, J., Cai, W., 2011. Sub-areas compartmentalization of Changjiang estuary based on the natural geographical characteristics. Acta Ecol. Sin. 31, 5042–5054.

Liu, R., Chen, Y., Sun, C., Zhang, P., Wang, J., Yu, W., Shen, Z., 2014. Uncertainty analysis of total phosphorus spatial-temporal variations in the Yangtze River Estuary using different interpolation methods. Mar. Pollut. Bull. 86, 68–75.

Matheron, G., 1963. Principles of geostatistics. Econ. Geol. 58, 1246–1266.

Mishra, U., Lal, R., Liu, D.S., Meirvenne, M.V., 2010. Predicting the spatial variation of the soil organic carbon pool at a regional scale. Soil Sci. Soc. Am. J. 74, 906–914.

Murphy, R.R., Curriero, F.C., Ball, W.P., 2010. Comparison of spatial interpolation methods for water quality evaluation in the Chesapeake Bay. J. Environ. Eng. 136, 160–171.

Pan, S.J., Shen, Z.L., 2010. Distribution and variation of dissolved inorganic nitrogen in Changjiang estuary and adjacent waters. Mar. Environ. Sci. 29, 183–205.

Shen, Z.L., Liu, Q., Zhang, S.M., Miao, H., Zhang, P., 2001. The dominant controlling factors of high content inorganic in the Changjiang River and its mouth. Oceanol. Limnol. Sin.

Shimada, T.I., Taro, 2015. Study on application of kriging to evaluation of radioactivity concentration for ensuring compliance with the criterion of site release. J. Nucl. Sci. Technol.

Smith, V.H., 2003. Eutrophication of freshwater and coastal marine ecosystems. A global problem. Environ. Sci. Pollut. Res. 10, 126–139.

Su, Y., Feng, C., Li, Y., Yin, L., Shen, Z., 2015. Heavy metal pollution in the surface water of the Yangtze estuary: a 5-year follow-up study. Chemosphere 138, 718–725.

Sun, C.C., Shen, Z.Y., Xiong, M., Ma, F.B., Li, Y.Y., Chen, L., Liu, R.M., 2013. Trend of dissolved inorganic nitrogen at stations downstream from the three-gorges dam of Yangtze River. Environ. Pollut. 180, 13–18.

Sun, W.P., Yu, P.S., Ming, P.J., 2009. Assessment of dissolved trace metals in nearshore area of Changjiang estuary and Hangzhou Bay with gray cluster method. Acta Oceanol. Sin. 31, 79–84.

Thompson, J.A., Kolka, R.K., 2005. Soil carbon storage estimation in a forested watershed using quantitative soil-landscape modeling. Soil Sci. Soc. Am. J. 69, 1086–1093.

Wang, J., Chen, Z., Wang, D., Shiyuan, X.U., Chunjuan, B.I., Liu, J., 2006. Calculation of inorganic nitrogen fluxes in the Yangtze estuary tidal wetland. Acta Geograph. Sin. 61, 729–740.

Wang, J.F., Li, L.F., George, C., 2009. Sampling and kriging spatial means: efficiency and conditions. Sensors 9, 5224–5240.

Wang, J.F., Christakos, G., Hu, M.G., 2010a. Modeling spatial means of surfaces with stratified Nonhomogeneity. IEEE Trans. Geosci. Remote Sens. 47, 4167–4174.

Wang, J.F., Li, X.H., Christakos, G., Liao, Y.L., Zhang, T., Xue, G., Zheng, X.Y., 2010b. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. Int. J. Geogr. Inf. Sci. 24, 107–127.

Wang, J.F., Zhang, T.L., Fu, B.J., 2016. A measure of spatial stratified heterogeneity. Ecol. Indic. 67, 250–256.

Zhang, J., Liu, S.M., Ren, J.L., Wu, Y., Zhang, G.L., 2007. Nutrient gradients from the eutrophic Changjiang (Yangtze River) estuary to the oligotrophic Kuroshio waters and re-evaluation of budgets for the East China Sea Shelf. Prog. Oceanogr. 74, 449–478.

Zhang, Z.L., Shu, J., Zhang, Y., 2011. Temporal and spatial characteristics of inorganic nitrogen and phosphate in the Yangtze estuary adjacent to Yangshan habour. Mar. Sci. 35, 47–52.

Zhu, D.Y., Zheng, B.H., Lei, K., Wang, L.J., Qin, Y.W., Deng, Y.X., 2008. A nutrient-distribution-based partition method in the Yangtze estuary. Acta Sci. Circumst. 28, 1233–1240.