Article

Optimal design of the water quality monitoring network in the semi-enclosed estuary

Nam-Hoon Kim [1], Jin Hwan Hwang [1,2,*]

[1] Department of Civil and Environmental Engineering, Seoul National University, Republic of Korea; nhkim426@gmail.com

[2] Institute of Construction and Environmental Engineering, Seoul National University, Republic of Korea; jinhwang@snu.ac.kr

* Correspondence: jinhwang@snu.ac.kr

Abstract: The semi-enclosed estuary is very susceptible to changes in the physical and environmental characteristics of the inflow from the land. Therefore, continuous and comprehensive monitoring of such changes is necessary for managing the estuary. Nevertheless, the procedure or framework has not been proposed appropriately to determine how many instruments are necessary and where they need to be monitored and standardized to detect critical changes. The present work proposes a systematical strategy for the deployments of the monitoring array by using the combination of the graphical optimization with the objective mapping technique. In order to reflect the spatiotemporal characteristics of the bay, the representative variables and eigenvectors are determined by the Empirical Orthogonal Function (EOF), and the cosine angle among them are calculated and used as a design index of optimization. At the recommended locations, the sampled representative variables are interpolated to reconstruct their spatiotemporal distribution and compared with the true distribution. Analysis confirms that the selected locations, even with a minimal number of points, can be used for on-site monitoring. Also, the present framework suggests how to determine installable regions for real-time monitoring stations, which reflect the global and local characteristics of the semi-enclosed estuary.

Keywords: Optimal design procedure; monitoring network; water quality; graphical optimization; objective mapping

1. Introduction

Due to the fast urban-sprawl or urbanization near the coastal areas, the amounts of released pollutants are increasing vigorously and flowing a lot into the coastal seas through the river from the urbanized watershed. The coastal areas influenced by the freshwater are often semi-enclosed so that once contaminants originated from the watershed flow into the bay, they are possible to be accumulated continuously deteriorating water quality ([1-3]). Along with the deterioration of the water environment, the demands of water supply have soared up, and it ultimately requested the construction of many sea-dikes for the coastal reservoir to secure the water resources. However, such sea-dikes efficiently inhibit the tidal momentum of the offshore sea from advecting to the river and near coastal waters and so significantly affect the physical characteristics of the water column such as stratification, mixing, and circulation of flow ([4-6]). Also, the gates of the sea-dikes are discharging freshwater irregularly to maintain the water level of the upstream to be constant, and so the physical characteristics and water quality of the coastal sea change complicatedly and unexpectedly ([7-9]).

The estuaries of the West Sea of Korea are shallow with wide tidal-flat, considerable tidal variations of the macro level, and complex geomorphology. Moreover, they are exposed to the physical and environmental alterations caused by the irregular discharges of freshwater from the gates of the coastal reservoir ([7, 10, 11]). Inflows from the watershed lead to unexpected unbalance of the nutrients and often occur the red tides by eutrophication in summer (i.e., flood season) ([12,

13]). Sometimes, water quality worsened due to the leakage of the green algae from the upstream of the reservoir ([14]). Also, a decrease in numbers of freshwater discharge in the winter due to low rainfall (i.e., dry season) can affect severely salt-sensitive aquacultures ([15, 16]). Therefore, earlier detection of changes in the nutrients, temperature, and salinity are essential to mitigate the impact in advance, understand the natural process, and manage the environment soundly.

Generally, two approaches have been mainly used to monitor the coastal sea: on-site monitoring and continuous real-time monitoring ([17]). The on-site monitoring is a way of visiting the site of interest, carrying with instruments regularly to monitor the characteristics of the bay. The continuous real-time monitoring is a way of collecting data remotely by installing the unmanned observatory. In any case, an important question can be raised as to how many and where the sensors or stations should be installed to sufficiently represent the spatial and temporal characteristics of the region of interest. Therefore, strategies need to be established and standardized to deploy and operate the monitoring array for managing sound environment. Even though several strategies could be proposed previously such as the guideline of US EPA ([18]), the monitoring locations have been judged arbitrarily by the discussion of stakeholders, engineers, and decision-makers rather than based on the robust and reliable systematic protocol or design guideline ([17, 19, 20]). Therefore, the strategies for deploying and operating a monitoring network need to be provided for the sound management of the coastal and ocean environment since the scientifically solid and robust data are essential in preparing countermeasures for decision making ([17]).

In a design of the monitoring locations, some requirements should be specified. Since the variables of interest are better to be measured simultaneously over the whole domain within a predetermined time, the limited numbers of measuring points should be optimally selected, which still must be sufficient to represent the spatiotemporal characteristics of the target region. Several prior studies have been done to meet these requirements. [21] conducted a trial-and-error method to find the points to best reconstruct the so-called objective mapping for visualizing the data. After the introduction of the objective mapping technique, some researches have followed focusing mainly on mapping-based optimization, which can compensate for the limitation of the trial-end-error method (e.g., [22-24]). Such prior studies designed arrays that can best reconstruct the spatial distribution by applying optimization technique such as the simulated annealing ([22]) and genetic algorithm ([23]) to minimize the covariance function or spatial averaged quadratic error ([24]). Besides, many types of research have performed to find the best objective mapping for the applications to the diverse fields such as the mooring locations to measure the sea level altitudes ([25]), the sensor arrays to monitor the oceanic meridional overturning circulation ([26]), and the collection data for the modeling with the data assimilation ([27, 28]).

Such developments recently lead to redesign the existing monitoring network, that had been intuitively and arbitrarily designed in the past. For example, [29] constructed the objective function based on the principal component analysis and solved it with spatial sampling optimization to eliminate redundant points in the Yangtze River Estuary, China. [30] and [31] performed a similar study using the Kriging and spatially simulated annealing method in the Changjiang Estuary and Hangzhou Bay, China, respectively. Most recently, [19], which is a precedent work of the present study, also proposed a primitive version of the framework designing the monitoring system using a similar method to the previously researched.

Most of the prior studies performed analysis mainly on the ideal case (e.g., [21-23]) or for the large-scale ocean (>5000 km) (e.g., [24, 26]) rather than small-scale waters (< 50 km) such as the coastal bays or estuaries except [19]. In general, the spatial and temporal variabilities of hydrodynamic and water quality variables in the global or large scales seem to follow more natural variation. However, the coastal water must depend not only on the global or large-scale variability but also significantly on the process of the land through the river. Therefore, the spatial variability of the characteristics in the coastal domain cannot be adequately represented by the same technique for the large-scale area. In addition, most of the prior studies did not design an integrated monitoring network which can detect the diverse variables simultaneously, but focus on a single variable such as current ([22, 23, 26], salinity ([32]), or water quality variable ([29-31]) to find the design variable for optimization.

As a recent study, [33] used the objective function as the quantitative function (i.e., scalar function), such as Root-Mean-Square-Error (RMSE) or covariance, etc. They found that such quantitative objective function not only requires a long computation time to find the solutions, but also it is hard to prevent the results from falling into the local solutions. Also, the optimization techniques based on the quantitative objective function are only suitable for the solutions of "how many" points are to be arranged and hard to find the solutions of "where" they are to be located. Moreover, there is little research to determine the locations of the real-time monitoring station ([34-36]).

Therefore, the present study is to propose a well-organized framework for designing the water quality monitoring network in the small-scale estuarine area. In order to reconstruct the spatiotemporal distribution to represent the variabilities of the target variables in the small-scale area, a graphical optimization technique is applied to find the best locations for the representative monitoring array with constructing the objective function of the optimal mapping approach. Since the graphical optimization technique can directly select the arrays of monitoring points in the continuous field, the computation time is short, and there is no possibility of falling into the local problems. Moreover, this technique is excellent in application to the problems of the steep gradient of signals with significant spatial variations due to the freshwater discharges. We present the methodologies for setting up the experiment in Chapter 2, results and discussion of the design for water quality monitoring network based on the scenarios in Chapter 3, and conclusions in Chapter 4.

2. Materials and Methods

2.1. Characteristics of the study area

The Geumgang Estuary (hereafter GE) is located at the mouth of a river on the west coast of Korea (Figure 1). GE has substantial variabilities of salinity since the freshwater is released irregularly and artificially from the coastal reservoir ([37]). The amount of the artificially discharged freshwater depends on the water level of the reservoir, which is closely related to the rainfall on the upstream watershed (Table 1). The amount of the discharged freshwater also determines the physical and environmental characteristics of the coastal seawater such as water temperature and salinity along with the concentrations of dissolved materials (e.g., nitrogen, phosphorus, chlorophyll, dissolved oxygen, etc) and so the significant changes of the freshwater discharge cause the large variations of the water quality. Such changes in the water quality need to be aware in advance to sustain the sound environmental conditions, and so well-organized monitoring and sampling methods are necessary. However, the monitoring points of the GE are sporadically arranged now without any specific standard (Figure 1b). Therefore, GE may be monitored irrationally at this moment, and so it is somewhat difficult to analyze and find intrinsic characteristics of the GE, in particular, where is strongly affected by the freshwater discharged from the upstream.

2.2. Numerical model

Seamless spatiotemporal information should be used as input data to design a monitoring network appropriately, but it is hard and expensive to perform field measurements in enough durations and area. Because of this, the scattered data of field measurements have limitations to be directly used for the design since they are available only at several specific points and during certain periods. An alternative way could be using the data from the satellite images instead, but they depend too much on the daily weather. Therefore, it is tough to obtain continuous spatiotemporal information, so using the satellite data can potentially miss some critical points and certain periods. For these reasons, the present work hired sets of the spatiotemporally highly resolved and well-validated numerical simulation data. The advantage of using the results from the numerical simulations is that diverse physical and environmental variables can be extracted from the numerical model simultaneously, which can be considered together as input data for designing the monitoring network.

This study used results from the numerical simulation performed by [15], assuming as a set of real data to design the monitoring network. The simulation results were calibrated and validated by comparing to the hydrodynamic and water quality variables measured in the field (see more detail in [15, 19]). Figure 1 shows a conceptual diagram, domain, and grid of the numerical model. A three-dimensional hydrodynamic model of Delft-3D ([38]) simulated the hydrodynamics and water quality near the coast. The initial and boundary conditions were carefully downscaled from the large-scale model of the Yellow Sea regional model (Figure 1a). The model results corresponding to about 55 km in the x-direction and about 35 km in the y-direction were extracted to apply and analyze (Figure 1b). In the numerical simulation, the initial and boundary conditions of freshwater discharge from the upstream were generated by a watershed model, STREAM ([15]). This model is a squared uniform grid and quasi-distributed watershed model that can simulate flow, sediment, and water quality of the watershed (Figure 1c).

The accuracy of the model results has been improved with calibration and validation step for each variable using the Index of Agreement (IOA) ([39]) and Relative Error (RE) ([40]), respectively. Overall, even though the water quality variables have slightly lower skill scores than the hydrodynamic, still both variables have strong correlations with the observation data (Table 2).
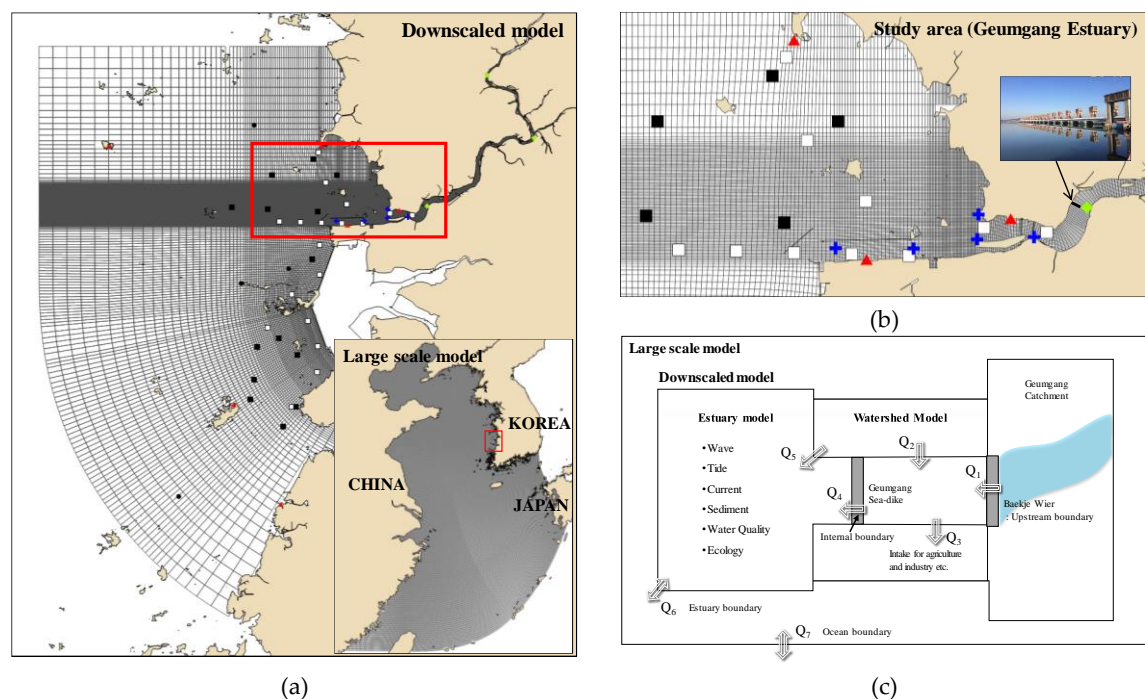

Figure 1. Monitoring status and computation grid of the (a) downscaled model, (b) Geumgang Estuary, and (c) concept of the integrated modeling.

Table 1. Monthly freshwater discharge in Geumgang Estuary (Sep. 1994 – Aug. 2017).

Month	1	2	3	4	5	6	7	8	9	10	11	12	Annual
Season	Winter		Spring			Summer			Autumn			Winter	
Discharge (10^6ton)	159	160	179	228	263	468	1202	1111	795	284	200	201	5250
Frequency	9	9	11	13	16	19	33	33	25	15	12	11	206
Total Time	22	23	28	35	42	59	131	127	93	38	28	28	654
Time/count	2.4	2.6	2.5	2.7	2.6	3.1	4.0	3.8	3.7	2.5	2.3	2.5	3.2

Table 2. Calibration and validation of the numerical model ([19]).

Variable	Parameter	Skill score		Skill index
		Calibration	Validation	
Wave	Hs	0.95	0.96	IOA
Tide	Semi-range	0.98	0.98	RE
	Phase-lag	1.00	0.99	
Tidal current	Amp.	0.82	0.87	RE
	Phase-lag	0.89	0.97	
SSC	-	0.65	0.64	RE
Water quality	Temperature	0.99	0.99	IOA
	Salinity	0.57	0.85	
	Chl-a	0.67	0.67	
	TN	0.95	0.95	
	TP	0.71	0.71	
	DO	0.85	0.65	

2.3. Design variables

Six variables were selected for analysis, water temperature (T), salinity (S), dissolved oxygen (DO), chlorophyll-a (Chl-a), total nitrogen (TN), total phosphorus (TP) which can commonly be obtained from the real field monitoring. The reason for considering multiple variables is to select representative variables among them to design the monitoring network. If the optimal location is determined by the representative variables and detect other variables with high reliability at that location, then we do not need to design the monitoring network complicatedly considering all the other variables. Therefore, it is imperative to select variables that can reflect the characteristics of other variables as a design variable. To reduce the number of variables and find a variable representing others, we used the Empirical Orthogonal Function (hereafter EOF) to compress extensive data set into a smaller number of independent pieces of information ([41, 42]), since it is hard and expensive to determine the convergence threshold of the objective function for each variable.

As the first step to use EOF, the eigenvalues corresponding to the series of a linear system need to be found, which can be expressed as follow:

$$\mathbf{C}\boldsymbol{\phi} - \lambda\mathbf{I}\boldsymbol{\phi} = 0; \tag{1}$$

where the covariance matrix, $\mathbf{C}$, consists of M elements of the data with the length of N ($M{\times}N$). $\mathbf{I}$ is the unity matrix, and $\boldsymbol{\phi}$ is the EOF. The EOF corresponding to the eigenvalue λ_M is the uncorrelated (i.e., orthogonal) mode of variability. If equation (1) is to have a nontrivial solution, the determinant of the coefficients must vanish and yield an Mth order polynomial, $\lambda^M + \alpha\lambda^{M-1} + \cdots$, whose M eigenvalues satisfy $\lambda_1 > \lambda_2 > \cdots > \lambda_M$ ([42]). Thus, the variances associated with each statistical mode are ordered according to their corresponding eigenvectors. The first mode, λ_1 contains the highest percentage of the total variance, and among the remaining variances, the greatest percentage is in the second mode, λ_2, and so on ([42]). This method can reduce the information of each variable to represent the variance concerning the eigenvectors.

The present work also chose a cosine angle between two eigenvectors of the representative variables in the three-dimensional Euclidean principal component (hereafter PC) space as a design variable for constructing the monitoring network. PC is constructed by the normalized six variables in this work, and two eigenvectors refer two most independent variables among six variables of T, S, DO, Chl-a, TN, and TP. The reason why we selected such a slightly complicate index as a design index is that if one variable is chosen for a design variable, other variables are hard to be monitored appropriately since each variable could have different spatiotemporal variabilities due to their different

sources. For example, water temperature is mainly determined by the local solar radiation and also water temperature from the open sea. However, salinity is mainly determined locally by the amount of the freshwater discharge from the river, so if the monitoring array is designed solely by salinity, this designed array is not likely adequate to detect the variations of water temperature (Figure 2a and c). Conversely, when the monitoring array is designed only by water temperature, the reconstructed distribution of salinity is totally different from the true distribution (Figure 2b and d).

Therefore, since the cosine angle that can represent the characteristics of variables having different origins, the use of that considers several variables simultaneously by the monitoring networks. The cosine angle between two vectors can be expressed as follow:

$$\cos(a,b) = \frac{a \cdot b}{\|a\| \cdot \|b\|} = \frac{\sum_{i=1}^{n} a_i b_i}{\left\{ \left(\sum_{i=1}^{n} a_i a_i \right) \times \left(\sum_{i=1}^{n} b_i b_i \right) \right\}^{1/2}} ; \tag{2}$$

where $a = (a_1, a_2, ..., a_n)$ and $b = (b_1, b_2, ..., b_n)$ are two eigenvectors of the representative variables. In the three-dimensional Euclidean PC spance, n must be three.
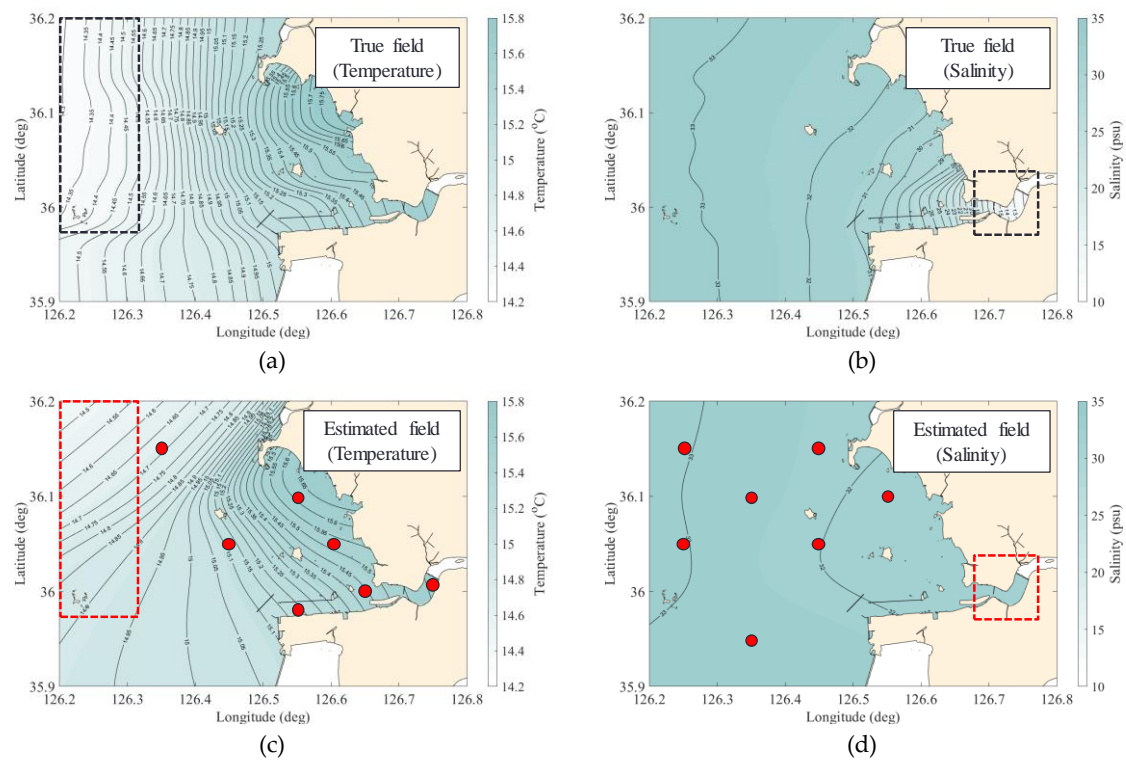

Figure 2. The true field of (a) water temperature and (b) salinity, and the example of the monitoring array (c) designed by considering only salinity and reconstruct the spatial distribution of the water temperature, and (d) designed by considering only water temperature and reconstruct the spatial distribution of the salinity.

2.4. Finding the optimal solutions

Once the design variable is determined, the optimization is performed to find solutions of the most appropriate numbers and locations for the monitoring in the domain of target. The general optimization problem is posed as follow:

$$
\begin{aligned}
\text{Minimize} \quad & f(\mathbf{X}) \\
\text{subject to} \quad & g_i(\mathbf{X}) \leq 0 \quad (i = 1, 2, ..., m); \\
& h_j(\mathbf{X}) = 0 \quad (j = 1, 2, ..., p); \\
& \mathbf{X}^{lower} \leq \mathbf{X} \leq \mathbf{X}^{upper};
\end{aligned}
\tag{3}
$$

where $f(\mathbf{X})$ is the objective function; $g_i(\mathbf{X})$ is the ith inequality constraint; m is the total number of inequality constraint functions; $h_j(\mathbf{X})$ is the jth equality constraint; p is the total number of equality constraints; $\mathbf{X}$ is the vector of design variables; and $\mathbf{X}^{lower}$ and $\mathbf{X}^{upper}$ are the lower and upper bounds of the design variables, respectively. To find the optimal solutions in a constrained optimization problem, it is necessary to construct feasible regions reflecting various constraint violations. Thus, the constrained optimization problem need to transformed into the unconstrained optimization problem by adding penalty terms for each constraint violation ([19, 43-46]). Finally, the objective function (i.e., augmented function) after transforming is solved by heuristic optimization such as genetic algorithm. This procedure is called an Augmented Lagrangian Genetic Algorithm (ALGA), which finds a set of stable solutions satisfying the Kuhn-Tucker conditions by mathematically handling a large number of constraint functions with less computational cost ([46, 47]).

To find the optimal solutions, we employed two optimization problems for comparison; one is a quantitative way, and another is a graphical one. The quantitative optimization finds an optimal arrangement, which can reconstruct the spatiotemporal distribution by constructing a quantitative objective function of the root-mean-square error (RMSE), which is expressed as follow:

$$
f(\mathbf{X}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} E_i - T_i}
\tag{4}
$$

where $E_i = (E_1, E_2, ..., E_N)$ and $T = (T_1, X_2, ..., T_N)$ are the estimated and true values, respectively. The quantitative objective function such as RMSE requires a long computation time to find a solution, and their result could easily fall into the location solution, and so there is no means to prove that whether a solution is the best or not. Therefore, even though this method has been applied to the selections of numbers of the observation points ([32]), the different arrays are possible to be found for each iteration due to arbitrary array selection of the locations.

The graphical optimization constructs the primary function of the optimal interpolation as an objective function and can find the optimal arrangement that represents the continuous spatiotemporal distribution ([19]). Currently, the most widely used optimal interpolation schemes in meteorological and oceanographic applications may be the statistical interpolation, also known as the Optimal Interpolation scheme (OI) ([48]), or the Barnes Objective Analysis (BOA) ([49]). Even though the OI is most prevalent to estimate the ocean data field (e.g., [21, 50-53]), it is not ideal to use in this study since the assumptions of spatial homogeneity and isotropy are not relevant to a small ($\leq$ 50 km) and highly dynamic area such as the small coastal seas.

The objective analysis is often referred to as a process of transforming data from observations at irregularly spaced points into data at the points on a regular space grid ([54]) for meteorological purposes. [49] modified this scheme to interpolate the whole complex region of interest by repeatedly applying a distance-dependent weighting ([55-57]). The objective function of the graphical optimization for designing the array of the monitoring points can be constructed by the loop function of BOA and can express as follow:

$$
f(\mathbf{X}) = \pm \left[\frac{\sum_{m=1}^{N} w_m R(\mathbf{X}_m)}{\sum_{m=1}^{N} w_m} + \frac{\sum_{m=1}^{N} w_m' \left(2R(\mathbf{X}_m) - E^1(\mathbf{X}_m) - E^2(\mathbf{X}_m)\right)}{\sum_{m=1}^{N} w_m'} \right];
\tag{5}
$$

where $R(\mathbf{X}_m)$ is the reference value (i.e., design variable) at the location m and $E^{1,2}(\mathbf{X}_m)$ are the estimates at each loop extracted at the same location to the reference value. The mth weights are given by:

$$w_m = \exp\left[-\left(\frac{d_{mx}^2}{c_x^2}+\frac{d_{my}^2}{c_y^2}\right)\right]; \tag{6}$$

$$w_m^{'} = \exp\left[-\left(\frac{d_{mx}^2}{\gamma c_x^2}+\frac{d_{my}^2}{\gamma c_y^2}\right)\right]; \tag{7}$$

where d_m is the distance between the grid point and the mth reference point and the length scales c_x and c_y control the fall-off rate of the weighting function in the different rates to x and y directions ([55, 56, 58]). The length scale could be solved by a nonlinear curve-fitting method of the Levenberg-Marquardt least square method ([59]). γ is a numerical convergence parameter that controls the difference between the weights on each step for the range of 0 to 1 ([60]).

In the graphical optimization, the objective function of equation (5) constructs the spatial distribution (i.e., domain) of the design variables to describe the target bay. The elliptic radius weighting function of equation (6) is the distance from the center to the border of an area described by the constrained function, which keeps the solution to be out of the radius of influence while finding the optimal solution. In addition, land and structures can be composed of several exterior nonlinear and graphical functions ([19]). Therefore, the solutions of maximum and minimum are located inside of the ellipse constraints (i.e., feasible region) refraining geomorphology from constraining them. To construct the objective function, BOA is selected due to its simplicity and applicability to the wide ranges of scales. It is also suitable for use in conjunction with the graphical optimization technique.

2.5. Methods of performance evaluation

Once the optimal solutions (i.e., monitoring array) are found, the spatiotemporal distribution of variables is reconstructed with the solutions and validated by comparing to the original data. The skills of the reconstruction can be evaluated by the statistical metrics ([61]), which tell the difference between true and estimated values. The present work used two types of skill metrics, one is the Taylor diagram ([62]), and another is the target diagram ([63]). These diagrams compile the statistical measures of the reconstruction skill into a single graph for allowing to compare and analyze the carious cases. Taylor diagram graphically summarizes and compares two sets of results regarding three statistics: correlation coefficient (COR), standard deviation (SD) of the true (subscription T) and estimated (subscription E) fields and centered (i.e., unbiased) root mean square difference (CRMSD), which have the following relationship:

$$\mathrm{CRMSD}^2 = \mathrm{SD}_T^2 + \mathrm{SD}_E^2 - 2\times\mathrm{SD}_T\times\mathrm{SD}_E\times\mathrm{COR}. \tag{8}$$

Another tool to evaluate the skills is a target diagram, which is derived from the relationship between the metrics of Bias, which means the difference of the mean values, CRMSD, and RMSD. This diagram used a Cartesian coordinate system where the x-axis represents the CRMSD, the y-axis represents the Bias, and the diagonal distance (radius) indicates the RMSD. CRMSD is an unbiased RMSD and removes any potentially biased information ([63]). The following relationship relates these three statistics:

$$\mathrm{RMSD}^2 = \mathrm{Bias}^2 + \mathrm{CRMSD}^2 \tag{9}$$

3. Results and discussion

3.1. Decomposition of the spatiotemporally dependent variable

At the first step to find the design variable, the monitoring points were distributed over the entire domain (Figure 3a), and the time-series of each variable were extracted at those points. In order to avoid dimensional heterogeneity, each variable was subtracted from the mean values of each variable, and we divided those differences by the standard deviations to have normalized values representing each viable (Figure 3b).

The most representative ones among the given variables were selected by the EOF analysis of the spatial distributions of the normalized six variables. Figure 4 shows the eigenvectors of the six variables, and Table 3 summarizes the results of the EOF analysis in detail. When the eigenvalues from the EOF are smaller than 1, those values are not significant, and so we used the first three ones among the six PCs of which eigenvalues are larger than 1 ([64]). The vectors in Figure 4 are constructed with the PCs representing the distribution of each variable. For example, the eigenvector representing salinity is composed of -0.5 of the first component, -0,36 of the second one, and 0.15 of the third one (Figure 4 and Table 1). In Figure 4, x-axis presents the contribution of the variable to the first PC, y-axis does to the second PC, and z-axis does to the third PC. When the magnitudes of the eigenvector are smaller than 0.5, the variables represented by that eigenvector are not significant ([64]), and so we do not consider to analysis.

Once the eigenvectors for each variable are calculated, six variables are categorized in the three groups depending on which PCs they contribute to. Among six variables, TN, TP, and S contribute to the first PC, which is 43% of the total variances, and T and DO contribute to the second PC of 32% (Figure 4a). Chl-a mainly contributes to the third PC, but only by 18% of the total variances so was excluded from consideration (Figure 4b). Therefore, the two groups contributing to the first and second PC are independent of each other since all PCs are orthogonal to each other. Of all the variables in two groups, S and T each contribute to the first and second PCs with a relatively larger magnitude than the other variables. Furthermore, they are also measured usually by one instrument simultaneously. Therefore, T and S were selected as representative variables to construct an index.
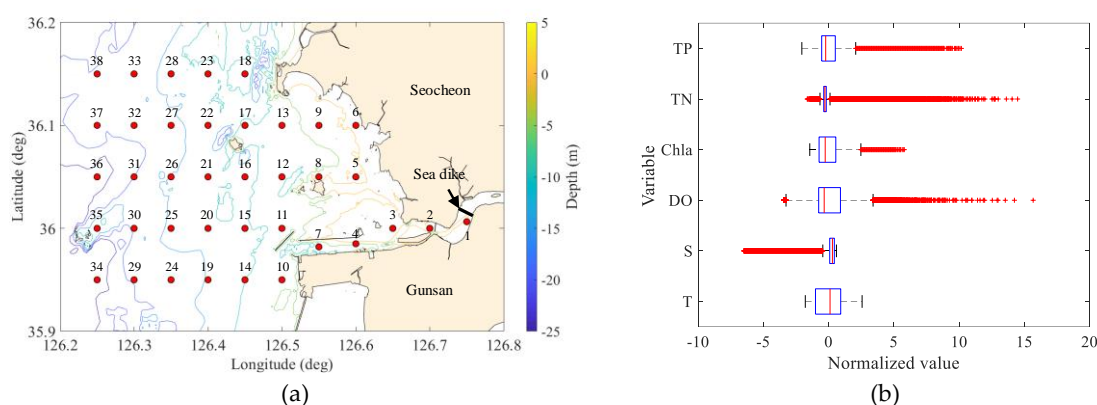

(a)   (b)

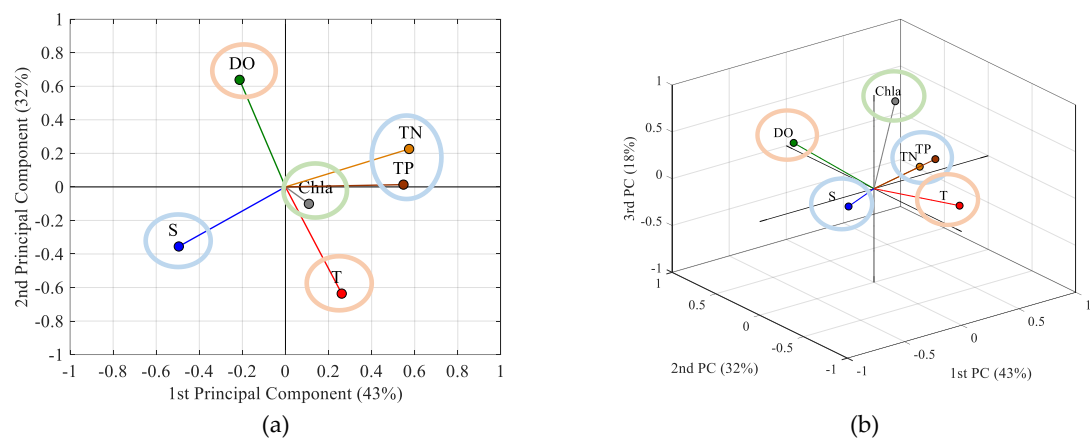**Figure 3.** (a) The initial placement of the monitoring points, and (b) boxplot of each variable.

Figure 4. The results of the EOF corresponding to the PCs for the spatial distribution.

Table 3. The results of the EOF corresponding to the PCs for the spatial distribution.

Category	Principal Component	Eigenvalue	Eigenvector					
			T	S	DO	Chla	TN	TP
Spatial	1st PC (43%)	2.56	0.26	-0.50	-0.21	0.11	0.57	0.55
(Entire	2nd PC (32%)	1.91	-0.64	-0.36	0.64	-0.10	0.23	0.01
domain)	3rd PC (18%)	1.06	0.02	0.15	0.27	0.94	-0.07	0.12

Once selecting the two most representative variables from the spatial PC analysis, the variables were analyzed again by the temporal EOF. Figure 5 shows an example of the time-series of the first, second, and third PCs at pt.1 and pt. 38. The first PC at pt.1 where is close to the gate of the sea-dike shows a sinusoidal tendency with irregular fluctuations, while the second and third PCs have very large irregular fluctuations compared to the first PC (Figure 5a). The irregular fluctuations appearing on the PCs have strong correlations of above 0.9 with those appearing on the real signals of salinity, which much be directly related to the releases of freshwater. Therefore, irregular fluctuations are probably due to the high frequency of artificial freshwater discharge. On the ocean side (pt.38), irregular fluctuations were not significantly observed on all PCs since this area is far from the gate, and the clearer sinusoidal time-series appears on the first PC (Figure 5b). The second and third PCs may have some tendencies, but we do not explain it here since the present work is aiming only to find an index rather than explain whole processes appearing in the area. Overall, sinusoidal characteristics are commonly decomposed in the first PC regardless of location, which is presumably due to seasonal variability.

In order to see how T and S contribute to each PC and relate them to the other four variables, the eigenvectors of six variables were calculated with the values measured at all 38 points. Among 38 sets of time series of PCs from the EOF, the results of pt.1 and pt.38 are presented in Table 4. On the closest location to the gate, pt.1, the first PC contributes 43% to the total variances, the second 32%, and the third 18%. The eigenvectors of T, DO, and TP appear to be greater than 0.5 on the first PC and S, Chl-a, and TN on the second PC. Chl-a and TN are mainly projected on the third PC, but the third PC does not contribute much to the total variances. On the ocean side, pt.38, the first PC contributes 47% to the total variances, the second 35%, and the third 11%. Here, T, S, and DO are projected onto the first PC with eigenvectors greater than 0.5, while TN and TP are on the second PC, and only Chl-a on the third PC.

The first PC is mainly affected by T and DO regardless of location and their sinusoidal trends are associated with seasonal variability. Also, the first PC near the gate shows irregular and highly frequent fluctuations, which is due to the gate operation. Those fluctuations seem to reflect the contribution of TP, which is originated from the upstream of the gate. While S contributes a lot to the second PC along with Chl-a and TN. On the ocean side, T, S, and DO show a large contribution to

the first PC reflecting the seasonality. After all, T and S show seasonal variability together in the ocean side, while near the gate, T still shows seasonal variability, but S appears close to the strong irregular variabilities. In other words, T mainly exhibits the seasonal variability along with DO, but S varies along with different variables depending on the locations (i.e., Chl-a and TN near the gate and T and DO on the ocean side). Therefore, T and S are selected as representative variables to be considered in the design of the monitoring network since they can reflect the effects of the seasonality and freshwater discharge, respectively and also help to deduce the changes of other variables.

As a next step, we calculated the cosine angle between T and S in a three-dimensional PC space, which can be a single design index representing the whole domain of the system. A low cosine angle (i.e., near zero) means that two variables representing the eigenvectors to construct that angle originate from different sources. On the other hand, a high cosine angle (i.e., near one) means that two variables are somewhat related and originated from similar sources.

Figure 6 shows the contour map of the cosine angles between the eigenvectors of T and S. The values of the cosine angle are increasing towards the open sea since T and S are simultaneously controlled by global open sea conditions such as current, solar radiation, and wind. Meanwhile, near the sea-dike, the values of cosine angles are low since while T still responds to global open sea conditions, S reflects not only to global conditions but also to a local condition such as artificially released freshwater. The cosine angle values near pt.6 are almost 0, which means that T and S have an orthogonal (independent) tendency. That is because this area has a shallow tidal flat, so the bottom surfaces are frequently exposed to the atmosphere during low tides (tidal amplitude is around 7.5 m, [4]). Such shallow tidal flats seem to be heated up and cooled down much faster than the deep southern navigation channel (near pt.4 and 7). In addition, this area is far from the gate and, therefore, may be less affected by the freshwater discharge.
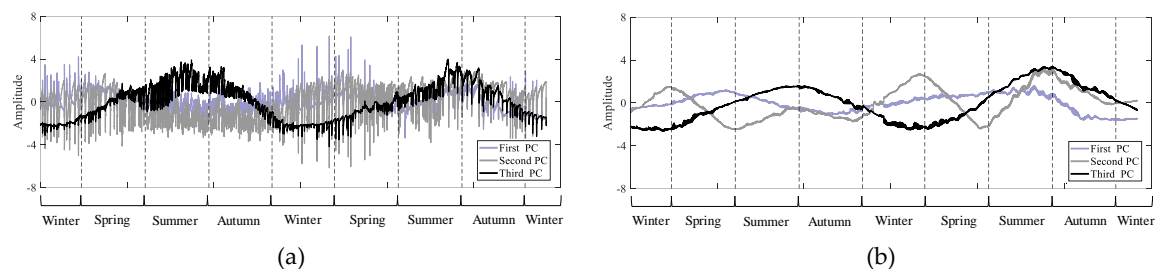

Figure 5. First, second, and third PC time-series of six decomposed variables extracted from (a) pt. 1(near the sea-dike) and (b) pt. 38 (ocean side).

Table 4. The results of the EOF corresponding to the PCs for the temporal distribution.

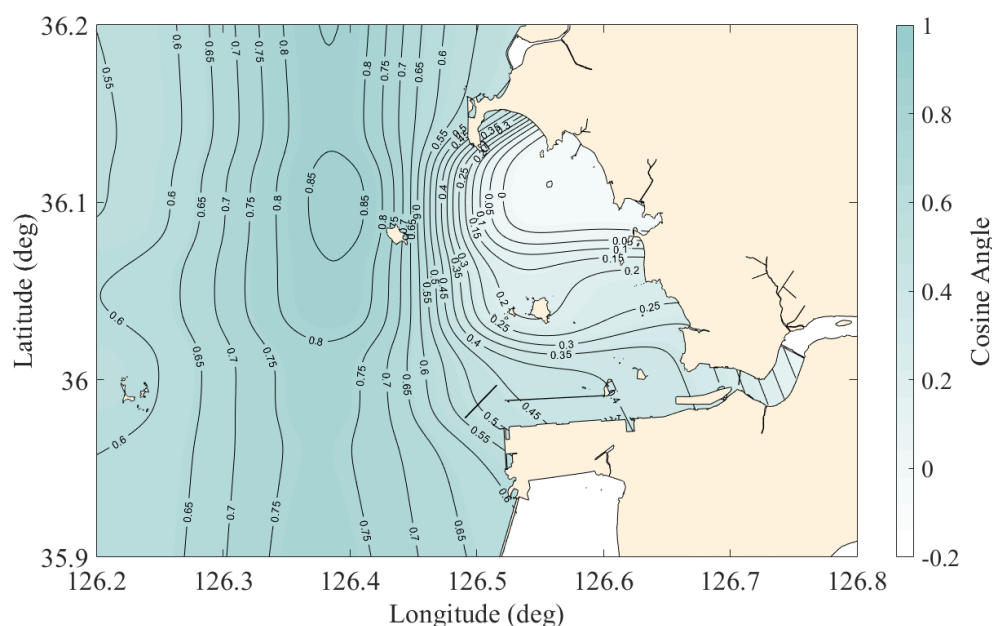| Category	Principal Component	Eigenvalue	Eigenvector					
			T	S	DO	Chl-a	TN	TP
Temporal (Pt.1 – near the sea-dike)	1st PC (43%)	2.59	0.58	0.10	-0.53	0.18	0.22	0.54
	2nd PC (32%)	2.20	-0.03	0.62	-0.20	0.51	-0.50	-0.25
	3rd PC (18%)	0.67	-0.14	0.10	0.39	0.66	0.62	0.05
Temporal (Pt.38 – ocean side)	1st PC (47%)	2.85	0.58	0.50	-0.52	0.31	0.04	0.23
	2nd PC (35%)	2.11	-0.10	-0.20	0.23	0.41	0.65	0.55
	3rd PC (11%)	0.67	-0.18	0.38	0.37	0.69	-0.09	-0.46

Figure 6. The spatial distribution of the true field of cosine angle composed of 38 points arranges at first.

3.2. Solutions for the monitoring array

As described earlier, the objective function of quantitative optimization was composed of RMSE, and we solved it using a genetic algorithm until converging to an optimal solution. On the other hand, the graphical optimization was configured the objective function by BOA method, and the optimal array was graphically selected by using a genetic algorithm. Those quantitative and graphical methods were used for searching 4 to 10 monitoring points to compare those two methods and recommend a better one. Figure 7 compares the reconstructed spatial distribution using the design index of 4, 7, and 10 monitoring points selected by quantitative and graphical optimization with the true spatial distribution. The dotted lines are the contours of the true values, and the solid lines are those of the reconstructed estimates. The spatial distribution of the graphical optimization (the panels on the left columns of the figure) reconstructs the contours more similar to the true distribution than the quantitative optimization (the panels on the right columns of the figure). Also, no matter how many searching points we want to, the graphical optimization can find a consistent location of points (Figure 7b, d, and f). However, the quantitative optimization finds different locations for each desired number of points (Figure 7a, c, and e). For example, the location of 7 points searched by the graphical optimization is the same as the location of 7 points out of 10 points (Figure 7d and f). Nevertheless, 7 points found by quantitative optimization are arranged in different locations from the 10 points (Figure 7c and e).

The optimal array was evaluated by the skill metrics, which plot the statistical parameters between true and estimated spatial distribution (Figure 8). First, the statistical parameters are plotted on a Taylor diagram to figure out how similar the estimated spatial distribution is to the true distribution (Figure 8a). The spatial distribution reconstructed using 7 to 10 points selected by graphical optimization agrees to the true spatial distribution with a high correlation of about 0.95 or more and a very low CRMSD. On the other hand, if the points are selected to be 6 or less, the statistical points are getting farther from the origin, which means poor reconstruction performance. In order to confirm how well the reconstructed distribution reproduces the variabilities of the true ones, we have identified the bias and RMSD on the target diagram (Figure 8b). The variabilities of the spatial distribution of the graphically selected points are within 0.1 of CRMSD, RMSD, and bias, for the cases with the 7 to 10 points. In the case with less 6 points, the results do not reconstruct well the true values. On the other hand, even though 9 or 10 points are selected by quantitative optimization, the results are slightly worse than 7 points of graphical optimization. Therefore, quantitative optimization has a

relatively poor reconstruction performance compared to the graphical optimization except for the cases of less than 4 points.

In order to determine how many points should be selected to construct a monitoring network, RMSDs and CORs obtained by the quantitative and graphical solutions are presented in Figure 9. The box plots are the quantiles of the populations obtained by many iterations in the quantitative optimization, and the red circles are the single values found by the graphical optimization. Overall, the solutions found by the graphical optimization show better reconstruction performance even with fewer numbers than the quantitative optimization. Besides, the graphical solution reaches a certain threshold with 7 or 8 points and, after reaching the threshold, converges regardless of the number of points. The solutions of quantitative optimization are different from each other depending on the number of iterations without converging on a certain value. Therefore, quantitative optimization has some statistical distributions, but graphical optimization provides a single solution without statistical distribution since this method does not require iteration to find a solution. As a result, the graphical optimization finds the solution (i.e., representative monitoring array) with more stable converging to an optimal solution for less computation time.
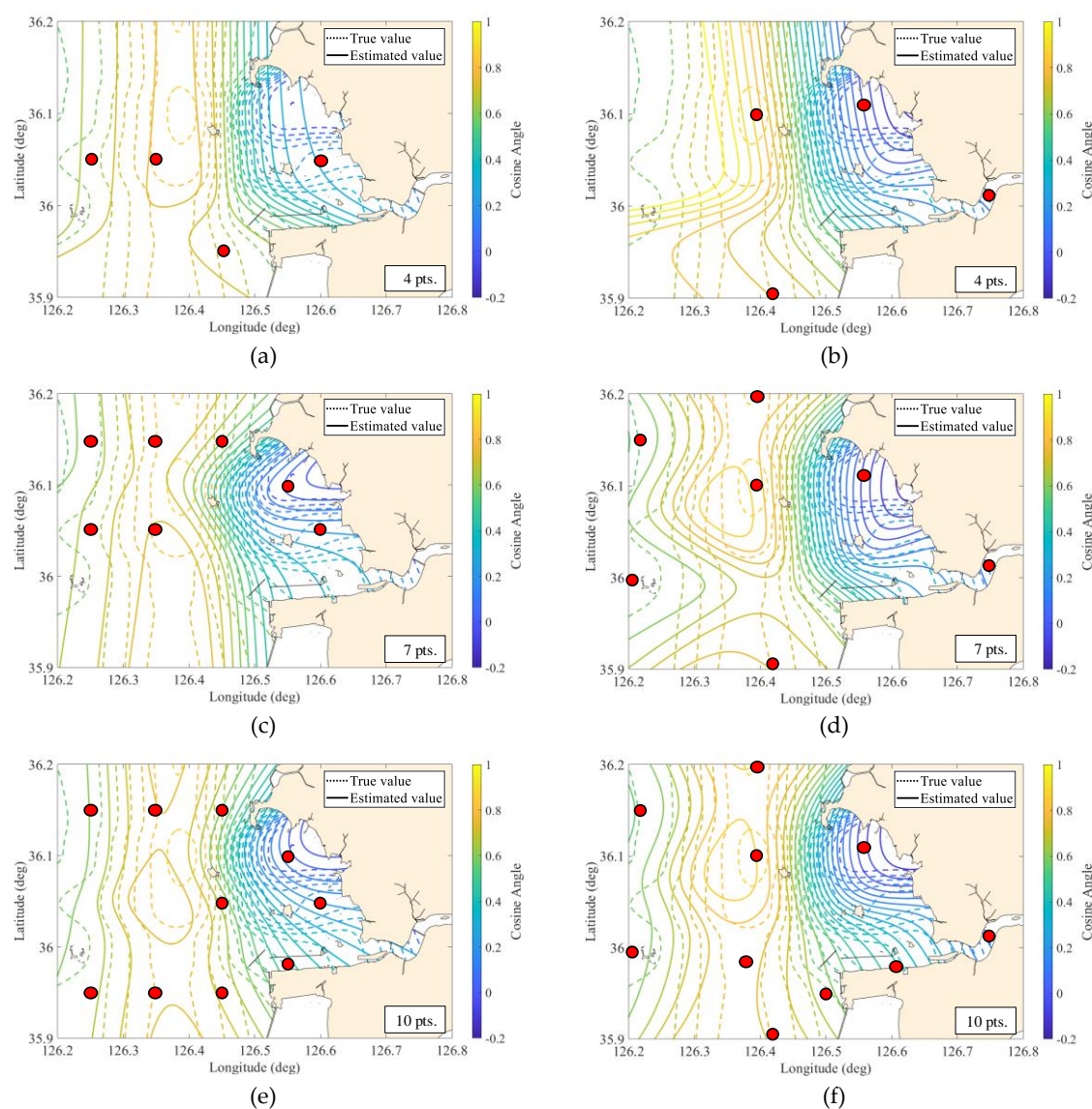

Figure 7. Comparison of the spatial distribution between true and estimated field reconstructed by using (a) and (b) 4, (c) and (d) 7, (e) and (f) 10 points of the monitoring array based on the quantitative optimization (left) and graphical optimization (right).
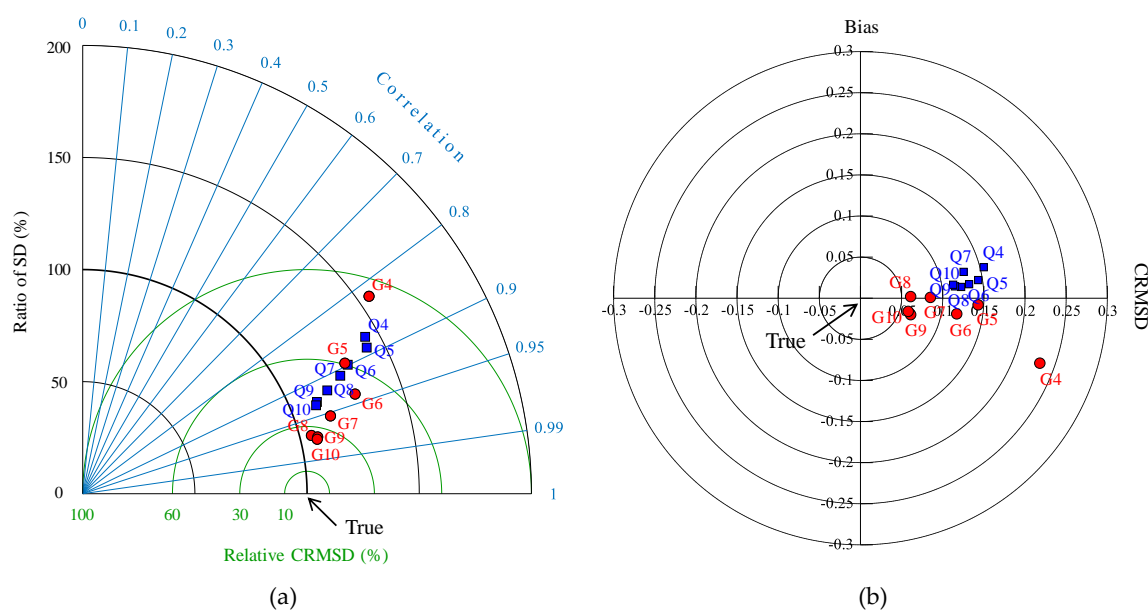
Figure 8. (a) Taylor diagram and (b) target diagram representing the statistics between the true and estimated spatial distribution. The abbreviation "Q" imply the quantitative and graphical optimization, respectively.
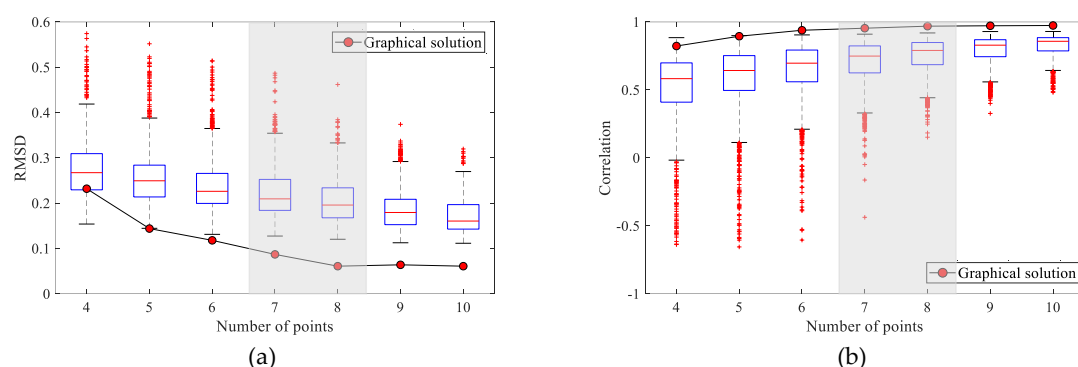

Figure 9. (a) RMSDs and (b) CORs of spatial distribution reconstructed by array of quantitative and graphical optimization.

The GE has considerable spatial and temporal variabilities of water qualities due to the change of freshwater discharge, which can cause extreme situations. Therefore, in addition to the normal case discussed earlier and named as the scenario N here, three more scenarios were built and tested. The scenario 2N releases twice the amount of freshwater discharge of the scenario N, and the scenario 3N does three times. The scenario I reduces the amount of freshwater discharge to 50% and increases twice the frequency of releasing of the scenario N. The numerical simulations were performed based on the scenarios, and the same method was applied to design the monitoring network for extreme events. Even though the scenarios are the functions of the amount of freshwater discharge and frequency, the representative variables are T and S as in the scenario N. Also, the trend of the first and second PCs from the EOF represents seasonality and irregular freshwater discharge and contributes by about 90% to the total time-series variance. As a result, the number of points required to reconstruct the spatial distribution by graphical optimization for three extreme scenarios is the same as the scenario N, but their locations are slightly different from the scenario N (Figure 10).

3.3. Optimal design of the water quality monitoring network

Since the on-site monitoring points selected in the four scenarios are distributed at slightly different locations, it is necessary to find a way to determine a location representing them. The time-

series of data at the monitoring points of each scenario were analyzed, and the representative locations were expressed in the form of influence radius by grouping the points located near each other (Figure 10). The influence radius (black dotted ellipse) was determined by a distance-dependent weighting function of the time-series of the variable characteristics, and the center (red+) was determined by using the nonlinear least square method using the equation (7) with $\gamma = 1$ and e-folding value. Since the marked points in each ellipse are solutions of each scenario, the center of ellipsis could be regarded as a representative point that characterizes the elliptic region with the influence radius. Therefore, time-series of data acquired within the radius of influence are almost similar to the values corresponding to weight 1 from the center. Such a series of steps led to select seven representative points in GE.

In order to evaluate how well those representative points reconstruct the true distribution of all variables, the spatial distribution of each water quality variable were reconstructed and compared with the true values (Table 5). Overall, the CORs are much higher than 0.8, and the RMSD is very low in terms of their scale of mean and standard deviation. These statistical quantities mean that the representative points can reconstruct the spatial distribution very similar to the true distribution while expressing the spatial variability of each variable very well. Besides, if the locations are selected reasonably, then with deploying even the minimum number of representative monitoring points, the spatial distribution of the six water quality variables can be pretty well reconstructed individually. Therefore, if the on-site monitoring network is designed by the framework of this study, an array, a set of representative points having the influence radius, can be considered as an example of a good representation of GE's spatial characteristics.

In order to determine where to install the real-time monitoring station in the representative area (i.e., area within influence radius), the signals at each area were compared and analyzed based on the reference signals of an area with the high external force or variations. The signals of the representative area (hereafter RA) 1 is assigned as a reference which is closest to the sea-dike and compared with the signals of the remaining areas. Table 6 shows the comparison of time-series for six variables in each area with that of RA1 statistically. The CORs of time-series shows that all variables except T and DO decrease as the monitoring point gets farther away from the reference point of RA1. Also, RMSDs increase as the points get farther away from RA1, but T and DO do not increase much relative to the magnitudes of mean and standard deviation. This is because T and DO are strongly subject to seasonal variability rather than freshwater discharge, while other variables are more significantly affected by the amount of freshwater discharge. Besides, in the area close to the sea-dike (i.e., RA2 and 3), the time-series of irregular freshwater discharge is reflected more than the others (i.e., RA4-7). From these results, global signals, such as seasonal variability, can be obtained in any area, while local signals, such as freshwater discharge, can only be obtained in certain areas (e.g., RA1-3). Therefore, one station must be unconditionally installed close to RA1, and other stations should be deployed near RA2 and RA3 in order to obtain the local water quality characteristics of GE.

Once a station is chosen and installed on RA1, 2, and 3 to acquire local signals, it is necessary to determine whether to install the monitoring stations in RA4, 5, 6, and 7. This is because irregular signals due to freshwater discharge can be obtained from S, Chl-a, TN, TP in RA 1-3, while it is difficult to obtain their global signals originated from offshore characteristics. In order to determine whether to install the monitoring stations in RA4, 5, 6, and 7, the signals from the seven RAs were compared with the reference signals of an offshore observatory which is the same location currently operated by the Korean government (Figure 10). Table 7 shows the statistical comparison of the time-series for six variables at each location with the offshore signals as the reference. The signals in RA4, 5, 6, and 7 are highly correlated with the offshore signals for all variables. In addition, their RMSDs show relatively low values considering the magnitude of mean and standard deviation of them. However, the signals in RA1, 2, and 3 have a relatively low correlation with the signals in the offshore. In particular, S and TN cannot infer the global signals originated from the offshore with the signal in RA1, 2, and 3. Consequently, the global signals of six variables in RA4, 5, 6, and 7 are not necessarily monitored in this domain because they can be obtained from the outside of the domain and sufficiently infer all of them.
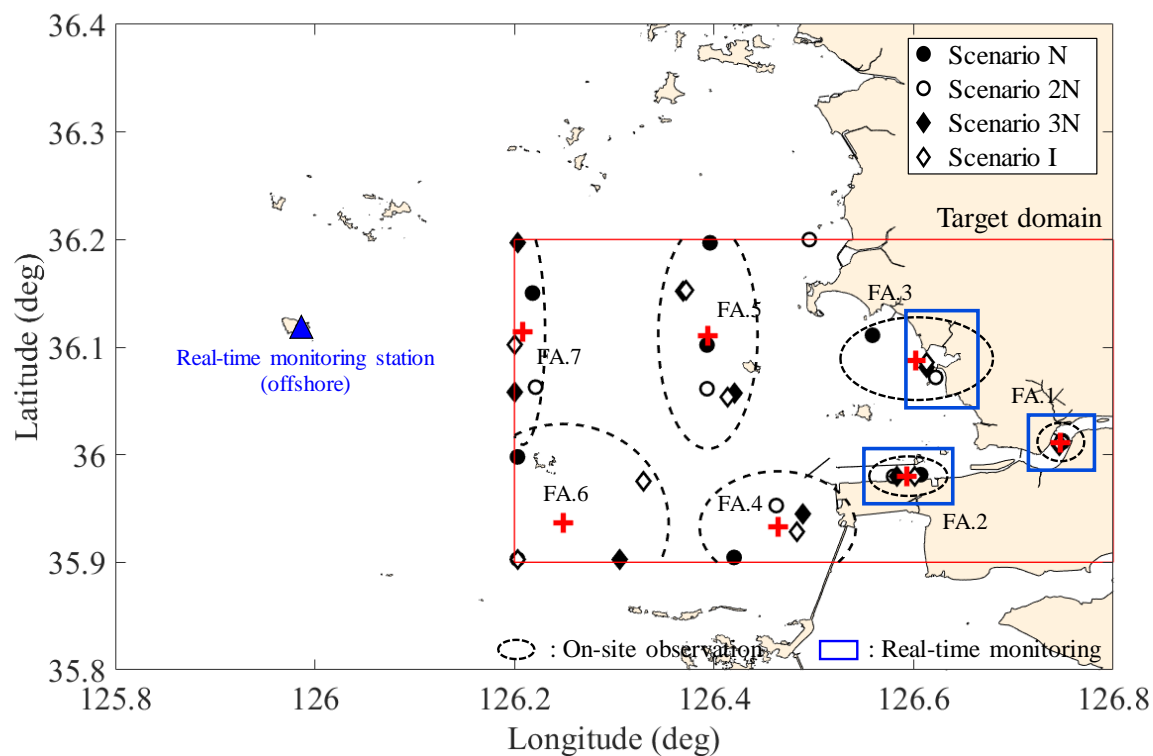
Figure 10. The selected points of the on-site monitoring (red+) and the installable area of real-time monitoring station (blue rectangular) in accordance with each scenario. The series of black dotted ellipse indicate maximum distances from the reference points (red+) corresponding to weight 1, and the blue rectangular regions are the installation area of the real-time monitoring station, which represent the temporal distribution of the local characteristic well. The blue triangle located in the outside of the target domain is the reference point of the offshore real-time monitoring station. The abbreviation "RA" imply the representative area.

Table 5. Statistical quantities of the reconstructed spatial distribution for six variables.

Statistics	Temperature	Salinity	Dissolved Oxygen	Chlorophyll-a	Total Nitrogen	Total Phosphorus
COR	0.99	0.99	0.80	0.93	0.98	0.96
RMSD	0.07	0.46	0.06	0.24	0.06	0.00
MEAN	15.48	31.64	8.43	4.39	0.52	0.05
STD	0.45	2.68	0.10	0.60	0.25	0.01

Table 6. Statistical quantities of the time-series distribution for six variables at each optimal point with RA1 as a reference point.

Statistics	Temperature						
	RA1	RA2	RA3	RA4	RA5	RA6	RA7
COR	1.00	0.99	1.00	0.96	0.95	0.88	0.90
RMSD	0.00	1.65	0.85	2.80	3.20	4.79	4.26
BIAS	0.00	0.50	-0.02	0.94	1.00	1.53	1.32
MEAN	16.38	15.88	16.40	15.44	15.38	14.85	15.06
STD	9.35	9.17	9.52	8.83	8.47	7.72	8.15
Salinity							
COR	1.00	0.38	0.45	0.35	0.37	0.22	0.26
RMSD	0.00	14.36	16.87	17.53	18.15	18.71	18.75
BIAS	0.00	-13.02	-15.73	-16.37	-17.01	-17.56	-17.61
MEAN	15.47	28.48	31.20	31.84	32.48	33.03	33.08
STD	6.55	2.50	1.28	1.01	0.67	0.51	0.53
Dissolved Oxygen							
COR	1.00	0.75	0.72	0.72	0.72	0.72	0.72
RMSD	0.00	1.94	2.01	2.05	2.10	2.12	2.11
BIAS	0.00	0.52	0.17	0.45	0.40	0.44	0.49
MEAN	8.82	8.31	8.65	8.38	8.43	8.38	8.33
STD	2.73	1.53	1.28	1.29	1.16	1.11	1.17
Chlorophyll-a							
COR	1.00	0.82	0.80	0.77	0.73	0.66	0.69
RMSD	0.00	1.57	2.81	1.79	1.90	2.32	2.03
BIAS	0.00	-0.20	-2.02	0.37	-0.13	-0.35	0.20
MEAN	4.07	4.27	6.08	3.70	4.20	4.41	3.87
STD	2.71	2.22	3.30	1.91	2.31	2.86	2.32
Total Nitrogen							
COR	1.00	0.54	0.30	0.27	0.16	0.20	0.15
RMSD	0.00	1.26	1.68	1.67	1.72	1.72	1.72
BIAS	0.00	1.10	1.51	1.50	1.56	1.56	1.56
MEAN	1.99	0.89	0.47	0.48	0.43	0.43	0.43
STD	0.74	0.27	0.05	0.05	0.04	0.04	0.04
Total Phosphorus							
COR	1.00	0.84	0.59	0.62	0.62	0.66	0.64
RMSD	0.00	0.03	0.04	0.04	0.04	0.04	0.04
BIAS	0.00	0.02	0.03	0.03	0.03	0.03	0.03
MEAN	0.07	0.06	0.04	0.05	0.04	0.05	0.05
STD	0.03	0.02	0.01	0.01	0.01	0.01	0.01

Table 7. Statistical quantities of the time-series distribution for six variables at each optimal point with the offshore as a reference point.

Statistics	Temperature							
	RA1	RA2	RA3	RA4	RA5	RA6	RA7	Offshore
COR	0.76	0.84	0.79	0.90	0.92	0.97	0.96	1.00
RMSD	6.58	5.51	6.49	4.52	3.98	2.49	3.09	0.00
BIAS	-2.28	-1.78	-2.30	-1.34	-1.27	-0.75	-0.96	0.00
MEAN	16.38	15.88	16.40	15.44	15.38	14.85	15.06	14.10
STD	9.35	9.17	9.52	8.83	8.47	7.72	8.15	5.99
	Salinity							
COR	0.16	0.29	0.10	0.65	0.56	0.89	0.85	1.00
RMSD	18.79	5.21	2.30	1.50	0.83	0.26	0.29	0.00
BIAS	17.63	4.61	1.89	1.26	0.62	0.06	0.02	0.00
MEAN	15.47	28.48	31.20	31.84	32.48	33.03	33.08	33.10
STD	6.55	2.50	1.28	1.01	0.67	0.51	0.53	0.38
	Dissolved Oxygen							
COR	0.72	0.97	0.97	0.99	0.99	1.00	1.00	1.00
RMSD	2.13	0.59	0.41	0.31	0.17	0.13	0.21	0.00
BIAS	-0.35	0.17	-0.17	0.10	0.05	0.10	0.15	0.00
MEAN	8.82	8.31	8.65	8.38	8.43	8.38	8.33	8.48
STD	2.73	1.53	1.28	1.29	1.16	1.11	1.17	1.04
	Chlorophyll-a							
COR	0.62	0.76	0.80	0.82	0.94	0.99	0.96	1.00
RMSD	3.37	2.90	2.45	3.11	2.27	1.53	2.39	0.00
BIAS	1.32	1.13	-0.69	1.69	1.19	0.98	1.53	0.00
MEAN	4.07	4.27	6.08	3.70	4.20	4.41	3.87	5.39
STD	2.71	2.22	3.30	1.91	2.31	2.86	2.32	3.93
	Total Nitrogen							
COR	0.28	0.40	0.75	0.80	0.94	0.98	0.95	1.00
RMSD	1.71	0.52	0.05	0.06	0.02	0.01	0.02	0.00
BIAS	-1.55	-0.46	-0.04	-0.05	0.00	0.01	0.01	0.00
MEAN	1.99	0.89	0.47	0.48	0.43	0.43	0.43	0.43
STD	0.74	0.27	0.05	0.05	0.04	0.04	0.04	0.05
	Total Phosphorus							
COR	0.69	0.90	0.94	0.96	0.98	0.99	0.99	1.00
RMSD	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
BIAS	-0.03	-0.01	0.00	0.00	0.00	0.00	0.00	0.00
MEAN	0.07	0.06	0.04	0.05	0.04	0.05	0.05	0.05
STD	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01

4. Summary and Conclusions

The coastal monitoring system, which is composed of several different series of sensors, aims to provide reliable information to forecast sea weather, sustain sound water quality, and plan to decision-making. Therefore, monitoring has been carried out to understand the inherent characteristics of the bay ([19]) carefully, but still it is not schematically determined how the monitoring network is constructed, but rather arbitrarily chosen. Therefore, the present study proposed a way to design an optimal monitoring network to fully reflect the spatiotemporal variability of water quality in semi-enclosed estuaries such as GE, which is a complex coastal system connected to the upstream watershed.

For designing an optimal monitoring network, instead of using ground-truth data that is not available realistically, the results from the well-validated numerical model were used to secure high-resolution assuming as ground-truth data. Such highly resolved numerical models allowed us to design a comprehensive monitoring network. With the results from the simulation, design variables were chosen to reflect the spatiotemporal characteristics of the bay adequately. As a representative design variable, the present work selected the cosine angle between the two eigenvectors of the representative variables in the three-dimensional PC space, which is decomposed by using the EOF analysis. This approach analyzes the inherent characteristics of the representative variables with other variables so that even if the monitoring network is designed with only a variable, it can sufficiently represent the characteristics of the other variables.

The most challenging part of the present study is that we can consider "where" as well as "how many" monitoring points to be placed. Conventional quantitative optimization could determine "how many" monitoring points are needed, but the solutions converged locally so that at every trial, a consistent arrangement of solutions could not be achieved. Therefore, the graphical optimization is applied and results in a consistent array for each simulation once the target number of points is set without high computational cost. With the distance-dependent weighting, the interpolation functions are constrained for bounding a region to be feasible for converging the objective function to the optimal solution. After that, the array of the monitoring points can be found on the interpolated space by applying the ALGA.

Finally, the spatiotemporal distribution reconstructed by using the selected optimal array is compared to the true distribution. The estimated spatial distribution is statistically evaluated by the skill metrics, on which to design an array of the on-site monitoring network. Moreover, the installable region of the real-time monitoring station can be determined by time-series comparison based on the reference point from which the bay's global and local signals can be acquired. As a result, GE requires a total of seven on-site monitoring points to fully represent the spatial distribution of water quality variables, and three real-time monitoring stations within the installable regions to simultaneously acquire global and local time-series characteristics.

Such a design method for finding the optimal estuarine monitoring network could be useful as a tool for strategically supporting decision-making. Besides, it is more meaningful in that the method can help not only designing the on-site monitoring array but also finding the installable regions of the real-time monitoring station that has been rarely studied so far. Such a monitoring network can reduce the cost, time, and effort for operating and managing the coastal monitoring and increase the reliability of the monitoring data ([17]). Also, the design procedure of this study can strategically organize the standard framework to determine the monitoring network in a semi-enclosed estuary, as well as the lake, bay, and open ocean. Moreover, an appropriate monitoring network can secure additional advantages in improving the accuracy of hydrodynamic models for data assimilation ([65, 66]).

Author Contributions: N.H. proposed the original idea, designed the research and wrote the article and J.H. supervise the present research and edited the manuscript.

References

1. Jang, D.; Hwang, J. H., Estuary classification method for considering climate change effects in South Korea. *Journal of Coastal Research* **2013,** SI65, 962-967.

2. Newton, A.; Icely, J.; Cristina, S.; Brito, A.; Cardoso, A. C.; Colijn, F.; Riva, S. D.; Gertz, F.; Hansen, J. W.; Holmer, M.; Ivanova, K.; Leppakoski, E.; Canu, D. M.; Mocenni, C.; Mudge, S.; Murray, N.; Pejrup, M.; Razinkovas, A.; Reizopoulou, S.; Perez-Ruzafa, A.; Schernewski, G.; Schubert, H.; Carr, L.; Solidoro, C.; PierluigiViaroli; Zaldivar, J. M., An overview of ecological status, vulnerability and future perspectives of European large shallow, semi-enclosed coastal systems, lagoons and transitional waters. *Estuarine Coastal and Shelf Science* **2014,** 140, 95-122.

3. Yoon, S. J.; Hong, S.; Kwon, B. O.; Ryu, J.; Lee, C. H.; Nam, J.; Khim, J. S., Distributions of persistent organic contaminants in sediments and their potential impact on macrobenthic faunal community of the Geum River Estuary and Saemangeum Coast, Korea. *Chemosphere* **2017,** 173, 216-226.

4. Kim, N. H.; Hwang, J. H.; Hyeon, K. D., Evaluation of Mixing and Stratification in an Estuary of Korea. *Journal of Coastal Research* **2018**, 96-100.

5. Kim, N. H.; Hwang, J. H.; Ku, H., Stratification of tidal influenced navigation channel. *Journal of Coastal Research* **2016**, 63-67.

6. Masunaga, E.; Yamazaki, H., A new tow-yo instrument to observe high-resolution coastal phenomena. *Journal of Marine Systems* **2014,** 129, 425-436.

7. Hwang, J. H.; Van, S. P.; Choi, B. J.; Chang, Y. S.; Kim, Y. H., The physical processes in the Yellow Sea. *Ocean & Coastal Management* **2014,** 102, 449-457.

8. Kim, H. C.; Son, S.; Kim, Y. H.; Khim, J. S.; Nam, J.; Chang, W. K.; Lee, J. H.; Lee, C. H.; Ryu, J., Remote sensing and water quality indicators in the Korean West coast: Spatio-temporal structures of MODIS-derived chlorophyll-a and total suspended solids. *Marine Pollution Bulletin* **2017,** 121, (1-2), 425-434.

9. Ostrander, C. E.; McManus, M. A.; DeCarlo, E. H.; Mackenzie, F. T., Temporal and spatial variability of freshwater plumes in a semienclosed estuarine-bay system. *Estuaries and Coasts* **2008,** 31, (1), 192-203.

10. Figueroa, S. M.; Lee, G. H.; Shin, H. J., The effect of periodic stratification on floc size distribution and its tidal and vertical variability: Geum Estuary, South Korea. *Marine Geology* **2019,** 412, 187-198.

11. Koh, C. H.; Khim, J. S., The Korean tidal flat of the Yellow Sea: Physical setting, ecosystem and management. *Ocean & Coastal Management* **2014,** 102, 398-414.

12. Lie, H. J.; Cho, C. H.; Lee, S.; Kim, E. S.; Koo, B. J.; Noh, J. H., Changes in marine environment by a large coastal development of the Saemangeum Reclamation Project in Korea. *Ocean and Polar Research* **2008,** 30, (4), 475-784.

13. Yih, W.; Kim, H. S.; Myung, G.; Park, J. W.; Yoo, Y. D.; Jeong, H. J., The red-tide ciliate Mesodinium rubrum in Korean coastal waters. *Harmful Algae* **2013,** 30, S53-S61.

14. Bellinger, E. G.; Sigee, D. C., *Freshwater Algae: Identification and use as bioindicators*. Wiley-Blackwell: Chichester, West Sussex, UK, 2015.

15.   Kim, H. C.; Song, Y. S.; Kim, Y. H.; Son, S.; Cho, J. G.; Chang, W. K.; Lee, C. H.; Nam, J.; Ryu, J., Implications of Estuarine and Coastal Management in the Growth of Porphyra sp in the Geum River Estuary, South Korea: A Modeling Study. *Journal of Coastal Research* **2018**, 396-400.

16.   Nishikawa, T.; Tarutani, K.; Yamamoto, T., Nitrate and phosphate uptake kinetics of the harmful diatom Eucampia zodiacus Ehrenberg, a causative organism in the bleaching of aquacultured Porphyra thalli. *Harmful Algae* **2009,** 8, (3), 513-517.

17.   Karydis, M.; Kitsiou, D., Marine water quality monitoring: A review. *Marine Pollution Bulletin* **2013,** 77, (1-2), 23-36.

18.   USEPA *National coastal condition assessment: Site evaluation guidelines*; EPA 843-10-004; United States Environmental Protection Agency: Washington, DC, USA, 2015.

19.   Kim, N. H.; Hwang, J. H.; Cho, J.; Kim, J. S., A framework to determine the locations of the environmental monitoring in an estuary of the Yellow Sea. *Environmental Pollution* **2018,** 241, 576-585.

20.   Kitsiou, D.; Tsirtsis, G.; Karydis, M., Developing an optimal sampling design. A case study in a coastal marine ecosystem. *Environmental Monitoring and Assessment* **2001,** 71, (1), 1-12.

21.   Bretherton, F. P.; Davis, R. E.; Fandry, C. B., A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep Sea Research and Oceanographic Abstracts* **1976,** 23, (7), 559-582.

22.   Barth, N.; Wunsch, C., Oceanographic experiment design by simulated annealing. *Journal of Physical Oceanography* **1990,** 20, (9), 1249-1263.

23.   Barth, N. H., Oceanographic experiment design .2. Genetic algorithms. *Journal of Atmospheric and Oceanic Technology* **1992,** 9, (4), 434-443.

24.   Hernandez, F.; Letraon, P. Y.; Barth, N. H., Optimizing a drifter cast strategy with a genetic algorithm. *Journal of Atmospheric and Oceanic Technology* **1995,** 12, (2), 330-345.

25.   Hackert, E. C.; Miller, R. N.; Busalacchi, A. J., An optimized design for a moored instrument array in the tropical Atlantic Ocean. *Journal of Geophysical Research-Oceans* **1998,** 103, (C4), 7491-7509.

26.   Baehr, J.; McInerney, D.; Keller, K.; Marotzke, J., Optimization of an observing system design for the North Atlantic meridional overturning circulation. *Journal of Atmospheric and Oceanic Technology* **2008,** 25, (4), 625-634.

27.   Bennett, A. F., Array design by inverse methods. *Progress in Oceanography* **1985,** 15, (2), 129-156.

28.   McIntosh, P. C., Systematic design of observational arrays. *Journal of Physical Oceanography* **1987,** 17, (7), 885-902.

29.   Gao, B.-B.; Wang, J.-F.; Fan, H.-M.; Xu, K.; Hu, M.-G.; Chen, Z.-Y., A stratified optimization method for a multivariate marine environmental monitoring network in the Yangtze River estuary and its adjacent sea. *International Journal of Geographical Information Science* **2015,** 29, (8), 1332-1349.

30.   Fan, H. M.; Gao, B. B.; Xu, R.; Wang, J. F., Optimization of Shanghai marine environment monitoring sites by integrating spatial correlation and stratified heterogeneity. *Acta Oceanologica Sinica* **2017,** 36, (2), 111-121.

31.   Bian, X. L.; Li, X. M.; Qi, P.; Chi, Z. H.; Ye, R.; Lu, S. W.; Cai, Y. H., Quantitative design and analysis of marine environmental monitoring networks in coastal waters of China. *Marine Pollution Bulletin* **2019,** 143, 144-151.

32.   Rogowski, P. A technique for optimizing the placement of oceanographic sensors with example case studies for the New York Harbor region. ProQuest Dissertations Publishing, 2009.

33.   Kim, N. H.; Hwang, J. H., Reconstruction of TS spatial distribution using minimum points in Geumgang Estuary. *Journal of Korean Society for Marine Environment & Energy* **2018,** 21, (4), 351-360.

34.   de Jonge, V. N.; Elliott, M.; Brauer, V. S., Marine monitoring: Its shortcomings and mismatch with the EU water framework directive's objectives. *Marine Pollution Bulletin* **2006,** 53, (1-4), 5-19.

35. Karydis, M.; Kitsiou, D., Eutrophication and environmental policy in the Mediterranean Sea: a review. *Environmental Monitoring and Assessment* **2012,** 184, (8), 4931-4984.

36. Wang, J. C.; Wang, Z. Q.; Wang, Y. M.; Liu, S. X.; Li, Y. Z., Current situation and trend of marine data buoy and monitoring network technology of China. *Acta Oceanologica Sinica* **2016,** 35, (2), 1-10.

37. Jeong, S.; Yeon, K.; Hur, Y.; Oh, K., Salinity intrusion characteristics analysis using EFDC model in the downstream of Geum River. *Journal of Environmental Sciences* **2010,** 22, (6), 934-939.

38. Deltares *Delft3D-FLOW user manual*; 2014; p 684.

39. Willmott, C. J., On the validation of models. *Physical Geography* **1981,** 2, 184-194.

40. Abramowitz, M.; Stegun, I. A., *Handbook of mathematical functions with formulas, graphs and mathematical tables*. 10th ed.; Dover publications: Dover, New York, USA, 1972.

41. Abdi, H.; Williams, L. J., Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* **2010,** 2, (4), 433-459.

42. Thomson, R. E.; Emery, W. J., *Data analysis methods in physical oceanography*. 3rd ed.; Elsevier: Miami, USA, 2014.

43. Conn, A. R.; Gould, N. I. M.; Toint, P. L., A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *Siam Journal on Numerical Analysis* **1991,** 28, (2), 545-572.

44. Conn, A. R.; Gould, N.; Toint, P. L., A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. *Mathematics of Computation* **1997,** 66, (217), 261-288.

45. Deb, K., *Optimization for engineering design: Algorithms and examples*. 2nd ed.; PHI Learning Private Limited: New Delhi, India, 2012.

46. Venkataraman, P., *Applied optimization with MATLAB programming*. 2nd ed.; John Wiley & Sons, Inc.: Hoboken, New Jersey, USA, 2009.

47. Deb, K., An efficient constraint handling method for genetic algorithms. *Computer Methods in Applied Mechanics and Engineering* **2000,** 186, (2-4), 311-338.

48. Gandin, L. S. *The objective analysis of meteorological field, Israel program for scientific translations*; Jerusalem, 1965; p 240.

49. Barnes, S. L., A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology* **1964,** 3, 396-409.

50. Reynolds, R. W.; Smith, T. M., Improved global sea-surface temperature analyses using optimum interpolation. *Journal of Climate* **1994,** 7, (6), 929-948.

51. Zhang, M. H.; Lin, J. L.; Cederwall, R. T.; Yio, J. J.; Xie, S. C., Objective analysis of ARM IOP data: Method and sensitivity. *Monthly Weather Review* **2001,** 129, (2), 295-311.

52. Guinehut, S.; Larnicol, G.; Le Traon, P. Y., Design of an array of profiling floats in the North Atlantic from model simulations. *Journal of Marine Systems* **2002,** 35, (1-2), 1-9.

53. Hoyer, J. L.; She, J., Optimal interpolation of sea surface temperature for the North Sea and Baltic Sea. *Journal of Marine Systems* **2007,** 65, (1-4), 176-189.

54. Cressman, C. P., An operational objective analysis system. *Monthly Weather Review* **1959,** 87, 367-374.

55. Koch, S. E.; Desjardins, M.; Kocin, P. J., An interactive Barnes objective map analysis scheme for use with satellite and conventional data. *Journal of Climate and Applied Meteorology* **1983,** 22, (9), 1487-1503.

56. Spencer, P. L.; Janish, P. R.; Doswell, C. A., A four-dimensional objective analysis scheme and multitriangle technique for wind profiler data. *Monthly Weather Review* **1999,** 127, (3), 279-291.

57. Sinha, S. K.; Narkhedakar, S. G.; Mitra, A. K., Barnes objective analysis scheme of daily rainfall over Maharashtra (India) on a mesoscale grid. *Atmosfera* **2006,** 19, (2), 109-126.

58. Carr, F. H.; Spencer, P. L.; Doswell, C. A.; Powell, J. D., A comparison of 2 objective analysis techniques for profiler time-height data. *Monthly Weather Review* **1995,** 123, (7), 2165-2180.

59. Marquardt, D. W., An algorithm for the least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* **1963,** 11, (2), 431-441.

60. Barnes, S. L., Applications of the Barnes objective analysis scheme .3. Tuning for minimum error. *Journal of Atmospheric and Oceanic Technology* **1994,** 11, (6), 1459-1479.

61. Zhang, A.; Hess, K. W.; Aikman, F., User-based skill assessment techniques for operational hydrodynamic forecast systems. *Journal of Operational Oceanography* **2010,** 3, (2), 11-24.

62. Taylor, K. E., Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research-Atmospheres* **2001,** 106, (D7), 7183-7192.

63. Jolliff, J. K.; Kindle, J. C.; Shulman, I.; Penta, B.; Friedrichs, M. A. M.; Helber, R.; Arnone, R. A., Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems* **2009,** 76, (1-2), 64-82.

64. Bengraine, K.; Marhaba, T. F., Using principal component analysis to monitor spatial and temporal changes in water quality. *Journal of Hazardous Materials* **2003,** 100, (1-3), 179-195.

65. Pham, V. S.; Hwang, J. H.; Ku, H., Optimizing dynamic downscaling in one-way nesting using a regional ocean model. *Ocean Modelling* **2016,** 106, 104-120.

66. Kim, N. H.; Pham, V. S.; Hwang, J. H.; Won, N. I.; Ha, H. K.; Im, J.; Kim, Y., Effects of seasonal variations on sediment-plume streaks from dredging operations. *Marine Pollution Bulletin* **2018,** 129, (1), 26-34.