AGU ADVANCING EARTH AND SPACE SCIENCES

Earth and Space Science

RESEARCH ARTICLE
10.1029/2023EA003144

Optimization of a Tsunami Gauge Configuration for Pseudo-Super-Resolution of Wave Height Distribution

Saneiki Fujita[1], Reika Nomura[2], Shuji Moriguchi[2], Yu Otake[1], Shunichi Koshimura[2], Randall J. LeVeque[2,3], and Kenjiro Terada[2]

[1]Department of Civil and Environmental Engineering, Tohoku University, Sendai, Japan, [2]International Research Institute of Disaster Science, Tohoku University, Sendai, Japan, [3]Department of Applied Mathematics, University of Washington, Seattle, WA, USA

Key Points:

- An optimization method implementing a greedy algorithm is presented to design a cost-effective observational network of tsunami gauges
- High-resolution information about tsunami wave height is constructed from sparse observations based on the concept of super-resolution
- The optimization process considering the existing observation network locations identifies their sufficiency or deficiency

Correspondence to:

S. Fujita,
saneiki.fujita.s8@dc.tohoku.ac.jp

Citation:

Fujita, S., Nomura, R., Moriguchi, S., Otake, Y., Koshimura, S., LeVeque, R. J., & Terada, K. (2024). Optimization of a tsunami gauge configuration for pseudo-super-resolution of wave height distribution. *Earth and Space Science*, *11*, e2023EA003144. https://doi.org/10.1029/2023EA003144

Received 10 SEP 2023
Accepted 11 JAN 2024

Author Contributions:

Conceptualization: Shuji Moriguchi, Yu Otake
Data curation: Shunichi Koshimura
Funding acquisition: Shuji Moriguchi, Kenjiro Terada
Methodology: Saneiki Fujita, Reika Nomura, Shuji Moriguchi, Yu Otake
Software: Saneiki Fujita, Reika Nomura
Supervision: Kenjiro Terada
Writing – original draft: Saneiki Fujita
Writing – review & editing: Reika Nomura, Randall J. LeVeque, Kenjiro Terada

Abstract In this study, we present an optimization method for determining a cost-effective sparse configuration for tsunami gauges to realize the reconstruction of high-resolution wave height distribution throughout the target region based on the concept of super-resolution. This optimization method consists of three procedures. First, we generate time series data of tsunami wave heights at synthetic gauges by conducting numerical simulations of various earthquake and tsunami scenarios at the target site. Next, we apply proper orthogonal decomposition to the synthetic tsunami data to extract the spatial features of the wave height distribution. Finally, according to these spatial features, an optimization process is performed to determine a sparse configuration of synthetic gauges. In the optimization, the optimal gauges are sequentially selected from the set of synthetic gauges to reconstruct the wave height distribution with the highest accuracy. Targeting hypothetical Nankai Trough earthquakes and tsunamis, we determine the optimal configuration near Shikoku and demonstrate the wave height reconstruction capability of the approach by comparing the performance of networks with optimally and randomly placed gauges. The results indicate that coastal gauges contribute more to improving the reconstruction accuracy and that a configuration with 21 optimal gauges has satisfactory performance. In addition, we assess the performance of the existing NOWPHAS network installed in the Shikoku region and find that the reconstruction performance of the existing network is equivalent to that of the optimal gauge network.

Plain Language Summary This study introduces a method of optimizing the sparse locations where actual tsunami gauges should be installed to obtain information on tsunami wave heights at any given point. By optimizing the locations of the observation points, it is possible to extend observations recorded at only a small number of points to obtain a good approximation to data at other points where the tsunami was not directly observed. First, numerical simulations are performed based on assumed earthquake and tsunami scenarios to generate synthetic time series data of tsunami waves. Then, by applying proper orthogonal decomposition to the obtained synthetic data, the characteristics of the tsunami wave height distribution are extracted. Finally, these characteristics are used to perform optimization by sequentially selecting the best gauges from among a set of candidate points to reconstruct wave height information for the entire target area, thereby determining the placements of a limited number of gauges. In a numerical demonstration example simulating a Nankai Trough earthquake and tsunami, the placement of gauges off the coast of Shikoku is optimized, and it is shown that the wave heights at arbitrary points over the entire area can be reproduced using data from at least 21 optimally placed gauges.

1. Introduction

On the Pacific coast of Japan, which has experienced many large earthquakes and tsunamis, various observational networks have been installed for real-time monitoring of seismic and tsunami information. The DONET (Dense Oceanfloor Network System for Earthquakes and Tsunamis) (Kaneda, 2010) and S-net (Seafloor Observation Network for Earthquake and Tsunami along the Japan Trench) (Kanazawa, 2013) systems consist of seismometers and water pressure gauges connected by submarine cables installed around the Nankai Trough and Japan Trench, respectively. In addition, N-net (Nankai Trough Seafloor Observation Network for Earthquakes and Tsunamis) (NIED, 2023) is currently being constructed on the southwest side of the Nankai Trough, where no other seafloor observation system presently exists, although the specific locations of its gauges have not yet been revealed to the public. Furthermore, NOWPHAS (Nationwide Ocean Wave information network for Ports and

HArbourS) (Ports and Harbours Bureau, 2022) has 78 wave gauges throughout Japan's coastal regions. Global Positioning System buoys, which are employed in the NOWPHAS network, have been reported to provide useful information for tsunami detection (Kato et al., 2005, 2010).

However, since the observational instruments require high installation and operation costs, monitoring nation-wide information with high resolution is difficult. Indeed, tsunami observation networks in Japan have been developed considering the installation and operational costs (Araki et al., 2008), as well as previous tsunami records, topographical conditions, and legal aspects (Abe & Imamura, 2013). Thus, standards of gauge locations to be installed are required to realize cost-effective observation networks consisting of as few gauges as possible.

To obtain sparse but effective observational networks, various studies have investigated the optimal configuration of gauges in tsunami observation networks. Schindelé et al. (2008) and Omira et al. (2009) determined possible gauge locations by considering different epicenter locations and the resulting tsunami travel times. In addition, there are some optimization objectives, such as maximizing the tsunami prediction accuracy (An et al., 2018; Hossen et al., 2018; Meza et al., 2020; Mulia et al., 2017, 2019) and minimizing the detection time (Ferrolino et al., 2020). Such optimization schemes are expected to be useful for recently developed early forecasting systems based on data assimilation and machine learning techniques (e.g., Heidarzadeh et al., 2019; Liu et al., 2021; Maeda et al., 2015; Makinoshima et al., 2021; Wang et al., 2020).

In tsunami risk assessment, it is essential not only to forecast tsunami arrival times as early as possible but also to capture information on wide-area tsunami dynamics. Indeed, as reported by Koshimura et al. (2020), many studies have been conducted to assess the impact of tsunamis over a wide area using recent advancements in remote sensing and image analysis techniques. In addition, various studies (e.g., Horspool et al., 2014; Kotani et al., 2020; Park et al., 2018) on probabilistic tsunami hazard assessment have also been conducted to quantitatively evaluate tsunami risks in wide coastal regions. Similarly, in the design of observation networks of wave gauges, it is desirable to ensure that the necessary information can be obtained at any given point in both coastal and offshore areas where tsunamis could threaten human beings and marine structures. Moreover, the tsunami height changes at locations without available observational instruments must be obtained to establish better evacuation and rescue operation standards. Thus, a sparse observation network consisting of a small number of informative observation gauges that can reproduce the necessary information at other locations throughout the region should be designed.

The demands on the gauge configuration can be addressed through sparse modeling, which is a known technique for representing the original high-dimensional data using their latent low-dimensional structure. In this context, Manohar et al. (2018) proposed a sparse sensor selection framework for signal reconstruction by combining machine learning and sparse sensing. The sparse sensors were optimally selected among the candidate points according to the low-dimensional spatial features extracted from high-dimensional data by applying proper orthogonal decomposition (POD). However, the application example of the optimization and reconstruction schemes for dynamic systems is limited. They validated the signal reconstruction performance based on the optimal sensor placement in dynamic systems, including fluid dynamics, and sea surface temperature changes. The sea surface temperature changes periodically and gradually in the spatial and temporal domains. In contrast, tsunamis occur suddenly and are size- and location-dependent phenomena, unlike sea surface temperature changes. Although Manohar et al. (2018) applied their scheme to the complex fluid dynamics of flow through a cylinder, the effectiveness of such an optimization scheme should be validated for the abovementioned tsunami-specific phenomena.

Inspired by the above background, this study presents a method of determining an optimal sparse arrangement of tsunami gauges to construct a virtually dense observation network. The proposed optimization method selects only p optimal gauges among n synthetic gauges ($p < n$) such that the information of the remaining $n - p$ gauges can be recovered, thus preventing the loss of significant information. In other words, the resulting virtually dense observation network consisting of p real and $n - p$ virtual gauges can provide observational data with higher spatial resolution than the data obtained at the p actually installed gauges. Since the spatial resolution enhancement using the observed wave heights is based on the concept of super-resolution, we refer to it as pseudo-super-resolution (PSR) in this paper. First, numerical simulations are carried out to generate synthetic tsunami wave data at synthetic gauges for various hypothetical fault rupture scenarios. Second, in line with Manohar et al. (2018), the POD method is applied to the obtained synthetic data to extract the spatial modes that characterize the high-dimensional and complex dynamics of the tsunami waves under consideration. Third, according to the extracted spatial modes, the optimal sparse arrangement of synthetic gauges is determined to realize

a virtually dense observation network that minimizes the PSR-based tsunami wave height estimation error for all n synthetic gauges. Thus, the present study is the first to apply a sparse sensor selection scheme (Manohar et al., 2018) to determine the optimal sparse arrangement of wave gauges for understanding the tsunami height distribution over a wide area.

Similar to this study, Mulia et al. (2017) proposed an optimization method for tsunami gauge configuration, targeting Nankai Trough earthquakes and tsunamis. The critical difference between their study and ours is the objective of optimization. While they focused on tsunami source inversion and determined an arrangement that can inversely estimate fault information with the highest accuracy, ours is based on the reconstruction of spatially dense wave heights from sparse observations. Additionally, Mulia et al. (2017) carried out empirical orthogonal function (EOF) analysis, a mathematically equivalent procedure to POD, for time series data of tsunami wave heights. At this point, EOF was used to determine candidate locations for subsequent optimization, while we directly search for the optimal configuration using the POD results based on pre-set candidates. Meanwhile, another novel contribution of this study is that uncertainty due to observation noise, which has been ignored in previous studies, is considered in the formulation. This uncertainty consideration is based on the use of Bayesian estimation to derive an objective function instead of the least squares estimation employed in the previous study (Manohar et al., 2018).

To test the performance of the presented optimization scheme, we demonstrate its ability in the offshore region of Shikoku Island (Japan), which has been threatened by large tsunami risks in the Nankai Trough. A tsunami database consisting of $n = 134$ candidate gauges and 1,564 hypothetical fault rupture scenarios is considered to construct virtually dense observation networks with $p = 7$, 21, and 42 optimal gauges. For validation, the performance of the optimal configuration realizing the virtually dense observation networks is compared to that of a relatively dense configuration with randomly arranged gauges. Additionally, we discuss the sufficiency and/or deficiency of the existing gauge network installed at the target site. It should be noted that early warning is not considered in the optimization of this study, and therefore, the optimized configuration may differ from the arrangement for early warning, where offshore observations are potentially important.

2. Methodology

This section explains the optimization method used to determine the sparse gauge configuration for realizing a virtually dense observation network.

2.1. Proper Orthogonal Decomposition (POD)

A series of tsunami simulations are carried out with l hypothetical fault rupture scenarios and n synthetic gauges measuring the synthetic wave data. Then, the time series wave data of all the synthetic gauges for all the scenarios are stored in one single data matrix $\boldsymbol{X}$, which is formulated as

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}^1 & \boldsymbol{X}^2 & \cdots & \boldsymbol{X}^l \end{bmatrix} \in \mathbb{R}^{n\times(m\times l)}, \tag{1}$$

where $\boldsymbol{X}^{j\in\{1,\ \ldots,\ l\}}$ includes the synthetic time series data of wave heights of the jth scenario and is defined as

$$\boldsymbol{X}^j = \begin{bmatrix} | & | & & | \\ \boldsymbol{x}^j_{t_1} & \boldsymbol{x}^j_{t_2} & \cdots & \boldsymbol{x}^j_{t_m} \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n\times m}. \tag{2}$$

Here, m denotes the total number of snapshots, and each column vector $\boldsymbol{x}^j_{t\in\{t_1,\ \ldots,\ t_m\}}$ is the snapshot vector at time t, which stores the wave heights at all synthetic gauges and is defined as

$$\begin{Bmatrix} | \\ \boldsymbol{x}^j_t \\ | \end{Bmatrix} = \begin{Bmatrix} x^j_{1,t} \\ \vdots \\ x^j_{n,t} \end{Bmatrix} \in \mathbb{R}^n. \tag{3}$$

The POD method is applied to the data matrix $\boldsymbol{X}$ with the singular value decomposition approach (Kerschen et al., 2005), and the dimension is reduced as

$$\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}} \approx \boldsymbol{\Phi}_r\boldsymbol{S}_r\boldsymbol{V}_r^{\mathrm{T}} = \boldsymbol{\Phi}_r\boldsymbol{A}_r, \tag{4}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{V} \in \mathbb{R}^{(m\times l)\times n}$ are matrices composed of left- and right-singular vectors arranged in the row direction, respectively, and $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ is a matrix with singular values of $\boldsymbol{X}$ aligned in its diagonal components in descending order. Here, $\boldsymbol{S}_r \in \mathbb{R}^{r\times r}$ is a diagonal matrix with the first r singular values, and $\boldsymbol{\Phi}_r \in \mathbb{R}^{n\times r}$ and $\boldsymbol{V}_r \in \mathbb{R}^{(m\times l)\times r}$ are the corresponding matrices with reduced dimensions. The last relationship is regarded as a reduced-order model (ROM). Additionally, $\boldsymbol{A}_r = \boldsymbol{S}_r\boldsymbol{V}_r^{\mathrm{T}} \in \mathbb{R}^{r\times(m\times l)}$ is a POD coefficient matrix that stores the r-dimensional state vectors; see, for example, Weiss (2019) and Tozato et al. (2022) for a detailed explanation of the POD method. Based on this expression, the wave snapshot for the jth scenario can be approximated as

$$\boldsymbol{x}_t^j \approx \boldsymbol{\Phi}_r\boldsymbol{a}_t^j, \tag{5}$$

where $\boldsymbol{a}_t^j \in \mathbb{R}^r$ is the corresponding column vector extracted from $\boldsymbol{A}_r$. In line with Nomura et al. (2022), we note that $\boldsymbol{a}_t$ contains scenario-specific and time-dependent information, which is a key factor for tsunami risk assessment in their framework. On the other hand, the mode matrix $\boldsymbol{\Phi}_r$ represents the common spatial features for all scenarios.

2.2. Pseudo-Super-Resolution (PSR) Method for Realizing a Virtually Dense Observation Network

We consider a situation in which the wave heights have been observed at only p gauges among all n synthetic gauges. Then, with $\boldsymbol{y}_t \in \mathbb{R}^p$ denoting the wave snapshot vector storing the wave heights at these p gauges at time t, we expect that the expanded snapshot vector $\boldsymbol{x}_t \in \mathbb{R}^n$ containing the data of all n synthetic gauges can be defined as

$$\boldsymbol{x}_t \approx f(\boldsymbol{y}_t), \tag{6}$$

where $f(\boldsymbol{y}_t)$ is a nonlinear function and a PSR operator that is used to obtain the high-resolution wave height distribution. If this relationship is satisfied, the sparse observations $\boldsymbol{y}_t$ are considered to include sufficient information to cover all n gauges and realize an equivalent virtually dense observation network with n gauges.

To realize the virtually dense observation network, we consider the inverse relation in Equation 6. The observational data $\boldsymbol{y}_t$ can be represented by the following observation equation:

$$\boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{x}_t + \boldsymbol{C}\boldsymbol{w}, \tag{7}$$

where $\boldsymbol{w}$ is the observation noise, which should follow a Gaussian distribution. Here, $C \in \mathbb{R}^{p\times n}$ is a downsampling matrix defining the sparse p ($<n$) gauge locations, which is formulated as

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{e}_{\gamma_1} & \boldsymbol{e}_{\gamma_2} & \cdots & \boldsymbol{e}_{\gamma_p} \end{bmatrix}^{\mathrm{T}}, \tag{8}$$

where $\gamma = \{\gamma_1, \ldots, \gamma_p\} \subset \{1, \ldots, n\}$ is the index sequence of the selected synthetic gauges, and $\boldsymbol{e}_{\gamma_i} \in \mathbb{R}^n$ is the standard basis vector in the n-dimensional Euclidean space, which has a value of one for the γ_i-th component and values of zero for the other components.

In accordance with the observation equation in Equation 7 and the POD representation in Equation 5, we construct a state-space model based on the local level model (Commandeur & Koopman, 2007) as follows:

$$\boldsymbol{a}_t = \boldsymbol{a}_{t-1} + \boldsymbol{v}, \tag{9}$$

$$\boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{\Phi}_r\boldsymbol{a}_t + \boldsymbol{C}\boldsymbol{w}' = \boldsymbol{\Phi}_r^p\boldsymbol{a}_t + \boldsymbol{C}\boldsymbol{w}', \tag{10}$$

where $\boldsymbol{v}$ is the system noise and $\boldsymbol{w}'$ is the sum of the observation noise $\boldsymbol{w}$ and the ROM error ($=\Phi_{r+}\boldsymbol{a}_t$) caused by the truncation of the higher-order modes $\boldsymbol{\Phi}_{r+}$ in Equation 4. These noise vectors are assumed to follow Gaussian distributions, $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_v)$ and $\boldsymbol{w}' \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_w)$. Thus, the observation noise $\boldsymbol{C}\boldsymbol{w}'$ at the sparse gauge locations follows a Gaussian distribution with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}_w^p = \boldsymbol{C}\boldsymbol{\Sigma}_w\boldsymbol{C}^{\mathrm{T}}$.

To reconstruct high-resolution wave height distribution at all synthetic gauges, we estimate the POD coefficient vector a_t based on the observed wave heights y_t by solving the inverse problem of the state-space model represented by Equation 9 and Equation 10. For that purpose, we employ the Kalman filter (Welch & Bishop, 1995) while accounting for the effects of uncertainty due to observation and system noise. With the Kalman filter, the posterior estimates of the state vector $\hat{\boldsymbol{a}}_t$ and the error covariance matrix $\boldsymbol{P}_t$ are evaluated as follows (Yoshida et al., 2018, 2021):

$$\hat{\boldsymbol{a}}_t^- = \hat{\boldsymbol{a}}_{t-1}, \tag{11}$$

$$\boldsymbol{P}_t^- = \boldsymbol{P}_{t-1} + \boldsymbol{\Sigma}_v, \tag{12}$$

$$\hat{\boldsymbol{a}}_t = \hat{\boldsymbol{a}}_t^- + \boldsymbol{P}_t^-\left(\boldsymbol{\Phi}_r^p\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_w^p\right)^{-1}\left(\boldsymbol{y}_t - \boldsymbol{\Phi}_r^p\,\hat{\boldsymbol{a}}_t^-\right), \tag{13}$$

$$\boldsymbol{P}_t = \left[\left(\boldsymbol{P}_t^-\right)^{-1} + \left(\boldsymbol{\Phi}_r^p\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_w^p\right)^{-1}\boldsymbol{\Phi}_r^p\right]^{-1}. \tag{14}$$

Here, $\hat{\boldsymbol{a}}_t^-$ and $\boldsymbol{P}_t^-$ are the prior estimates of the state vector and error covariance matrix, respectively, at time t. Assuming that the estimated POD coefficients $\hat{\boldsymbol{a}}_t$ are reasonable, with sufficiently small errors relative to the true POD coefficients $\boldsymbol{a}_t$, we obtain the following approximation for the high-resolution wave height distribution:

$$\boldsymbol{x}_t \approx \boldsymbol{\Phi}_r\hat{\boldsymbol{a}}_t \tag{15}$$

$$= \boldsymbol{\Phi}_r\left[\hat{\boldsymbol{a}}_t^- + \boldsymbol{P}_t^-\left(\boldsymbol{\Phi}_r^p\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_w^p\right)^{-1}\left(\boldsymbol{y}_t - \boldsymbol{\Phi}_r^p\,\hat{\boldsymbol{a}}_t^-\right)\right]. \tag{16}$$

This formula is regarded as a substitute for the wave height PSR relation in Equation 6 and is expected to reproduce high-resolution wave height distribution from the available low-resolution distribution observed at p gauges.

2.3. Optimization of the Sparse Gauge Configuration

2.3.1. Optimization Problem

Highly accurate wave height PSR is realized by minimizing the estimation error $\boldsymbol{\epsilon} = |\boldsymbol{a}_t - \hat{\boldsymbol{a}}_t|$. This idea is based on previous studies (Manohar et al., 2018; Y. Saito et al., 2021) that proposed methods for optimizing sparse sensor arrangements for signal reconstruction. Specifically, Manohar et al. (2018) employed the determinant of the error covariance matrix $\det[\mathrm{Cov}(\boldsymbol{\epsilon})]$ as the objective function to be minimized, which is equivalent to the well-known D-optimal design (Papalambros & Wilde, 2000).

Following their idea, we pose the following combinatorial optimization problem:

$$\max_{\gamma}\ \det\left[\left(\boldsymbol{\Phi}_r^p\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_w^p\right)^{-1}\boldsymbol{\Phi}_r^p\right], \tag{17}$$

$$\text{where } \boldsymbol{\Phi}_r^p = \boldsymbol{C}\boldsymbol{\Phi}_r \ \text{ and } \ \boldsymbol{\Sigma}_w^p = \boldsymbol{C}\boldsymbol{\Sigma}_w\boldsymbol{C}^{\mathrm{T}}. \tag{18}$$

This objective function is based on minimizing the information entropy (Shannon, 1948); see Appendix A for more information. Here, we consider the noise covariance matrix $\boldsymbol{\Sigma}_w$ according to the uncertainty of each observation gauge. On the other hand, it should be noted that in a previous study (Manohar et al., 2018), the covariance matrix was assumed to be the unit matrix $\boldsymbol{I}$, that is, each observation gauge had the same uncertainty.

2.3.2. Greedy Algorithm With Column-Pivoting QR Decomposition

In this study, a greedy algorithm is employed to solve the above combinatorial optimization problem for the optimal set of indices γ. Greedy algorithms provide an approximation of the globally optimal solution with reasonable computational costs while sequentially considering local optima; see Appendix C for a comparison of greedy and genetic algorithms, where the latter are among the most popular evolutionary optimization schemes.

To enable greedy seeking for approximation of optimal solutions by QR decomposition with column pivoting, we expand the objective function in Equation 17 as

$$\det\left[\left(\boldsymbol{\Phi}_r^p\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_w^p\right)^{-1}\boldsymbol{\Phi}_r^p\right] = \det\left\{\left[\left(\boldsymbol{\Sigma}_w^p\right)^{-1/2}\boldsymbol{\Phi}_r^p\right]^{\mathrm{T}}\left(\boldsymbol{\Sigma}_w^p\right)^{-1/2}\boldsymbol{\Phi}_r^p\right\} \tag{19}$$

$$= \prod_{i=1}^{r}\sigma_i\left[\left(\boldsymbol{\Sigma}_w^p\right)^{-1/2}\boldsymbol{\Phi}_r^p\left(\boldsymbol{\Phi}_r^p\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_w^p\right)^{-1/2}\right]. \tag{20}$$

Here, $\sigma_i(\bullet)$ denotes the ith singular value of the matrix $\bullet$. Following Manohar et al. (2018), QR decomposition with column pivoting is applied for optimal gauge selection as

$$\boldsymbol{\Sigma}_w^{-1/2}\boldsymbol{\Phi}_r\boldsymbol{\Phi}_r^{\mathrm{T}}\boldsymbol{\Sigma}_w^{-1/2}\boldsymbol{C}^{\mathrm{T}} = \boldsymbol{QR}, \tag{21}$$

where $\boldsymbol{Q}\in\mathbb{R}^{n\times n}$ is an orthogonal matrix, and $\boldsymbol{R}\in\mathbb{R}^{n\times n}$ is an upper triangular matrix with r_{jj} in its diagonal components. Additionally, $\boldsymbol{C}^{\mathrm{T}}\in\mathbb{R}^{n\times n}$ can be regarded as a permutation matrix for $\boldsymbol{R}$ to obtain the diagonal components in descending order as $r_{11}\geq r_{22}\geq\cdots\geq r_{nn}$. Thus, the product of the singular values in the last expression of Equation 20 is replaced by the product of the diagonal entries r_{jj}. Thus, the determinant-based maximization process is performed by successively identifying the largest diagonal component. As a result, we can obtain the optimized index sequence γ composing the downsampling matrix $\boldsymbol{C}$; see Appendix B for more details on the process of QR decomposition with column pivoting.

In accordance with the greedy algorithm based on QR decomposition with column pivoting presented in Appendix B, we optimize gauge configuration by selecting the best gauge from the set of synthetic gauges corresponding to candidate points. Specifically, we calculate the diagonal component r_{jj} in $\boldsymbol{R}$ when each synthetic gauge is assumed to be adopted and select the gauge with the highest r_{jj} as the optimal one. Thus, the precomputed r_{jj} of each synthetic gauge is used as a quantitative measure of its degree of optimality, that is, its degree of effectiveness in increasing the objective function value. We determine the final gauge combination by sequentially selecting the locally optimal solution, which is determined as the most effective gauge in each optimization process. Therefore, the greedy algorithm employed in this study has an advantage over other optimization methods, such as direct search methods and heuristic algorithms, which evaluate the suitability of an entire combination of observation gauges at once.

It is worth considering the relationship between the number of modes r and the number of optimal gauges p. The sparse sensor optimization is equivalent to identifying the combination of p rows from the contracted POD mode matrix Φ_r. Here, the rank of the mode matrix after r mode reduction and p row selections $\boldsymbol{\Phi}_r^p$ is calculated as follows:

$$\mathrm{rank}\left(\boldsymbol{\Phi}_r^p\right) = \begin{cases} p & (p\leq r) \\ r & (p>r) \end{cases} \tag{22}$$

In other words, it depends on the number of optimal gauges for undersampling ($p<r$) and on the number of modes for oversampling ($p>r$). This is the reason why the oversampled optimal gauges ($p>r$) are less informative than those of the undersampling case, suggesting that the number of modes for dimensionality reduction is important to achieve a sparse arrangement. In this study, the number of modes is determined using the contribution rate of the spatial modes, as described in the next section. However, establishing a more reasonable number of modes is an important issue and is a subject for future research.

Table 1
Fault Rupture Parameters Depending on Mw

Mw	Length [km]	Width [km]	Dislocation [m]
7.6	115.1	57.6	1.17
7.9	162.6	81.3	1.65
8.2	229.7	114.9	2.33
8.5	324.5	162.3	3.29
8.8	458.4	229.2	4.64

3. Tsunami Data Set

Targeting the off/nearshore area of Shikoku Island, Japan, which has been threatened by the large tsunami triggered by the Nankai Trough subduction zone (Furumura et al., 2011; Ishibashi, 2004; S. Miyazaki & Heki, 2001; T. Saito et al., 2010), we carry out simulations for various tsunami scenarios and store the wave height data obtained by the synthetic gauges. Since this subduction zone is located along the long section of Japan's southeastern coast, there are various possible tsunami paths in the Shikoku region. Therefore, there is a specific need for a virtually dense observation network that can cover the entire offshore area.

3.1. Hypothetical Earthquake/Tsunami Scenarios

Hypothetical fault rupture patterns are first devised as the primary sources of tsunamis. In this study, we focus on intraslab and interslab earthquakes in the Nankai Trough due to the high tsunami risk and the feasibility of fault rupture modeling (Utsu, 2001). Taking into account the surface distribution of the Philippine Sea Plate (Baba et al., 2002) and the slip processes of the 1944 Tonankai earthquake and 1946 Nankai earthquake (Baba & Cummins, 2005; Kanamori, 1972), we employ the rectangular fault model (Okada, 1985, 1992). The fault rupture parameters, such as the fault's depth, dip, and strike, which depend on the location, are considered and set to range from 1 to 189 km, 5–67°, and 207–254°, respectively. These location-dependent parameters were determined by the Philippine Sea Plate geometry (e.g., Bird, 2003; Hirose et al., 2008). Note that the Dip angles of the high-value range reflect the steep subduction of the Plate at eastern Kyushu Island (e.g., K. Miyazaki et al., 2023; Yoshioka et al., 2008). The considered magnitudes range from 7.6 to 8.8, and the length, width, and dislocation are calculated according to the scaling law (Utsu, 2001). The specific values of the parameters depending on Mw are summarized in Table 1. The rake is assumed to be a constant value of 90°. Thus, two uncertainties, magnitude and location, are considered to generate the earthquake scenarios, and the other parameters can be determined from these parameters, by targeting intraslab and interslab earthquakes, which occur on a tectonic plate. Figure 1 shows the locations of faults corresponding to the 1,564 earthquake scenarios considered in this study. Here, each circle indicates the northeast apex position of the rectangular fault, whose size is determined by the width and length depending on the magnitude.

To obtain six-hour propagations immediately after the fault rupture, the TUNAMI-N2 model (Goto et al., 1997; Imamura et al., 2006) was adopted to simulate the propagation of each tsunami wave by solving the nonlinear shallow water equation. Then, the initial surface deformation of the seafloor was calculated for each scenario

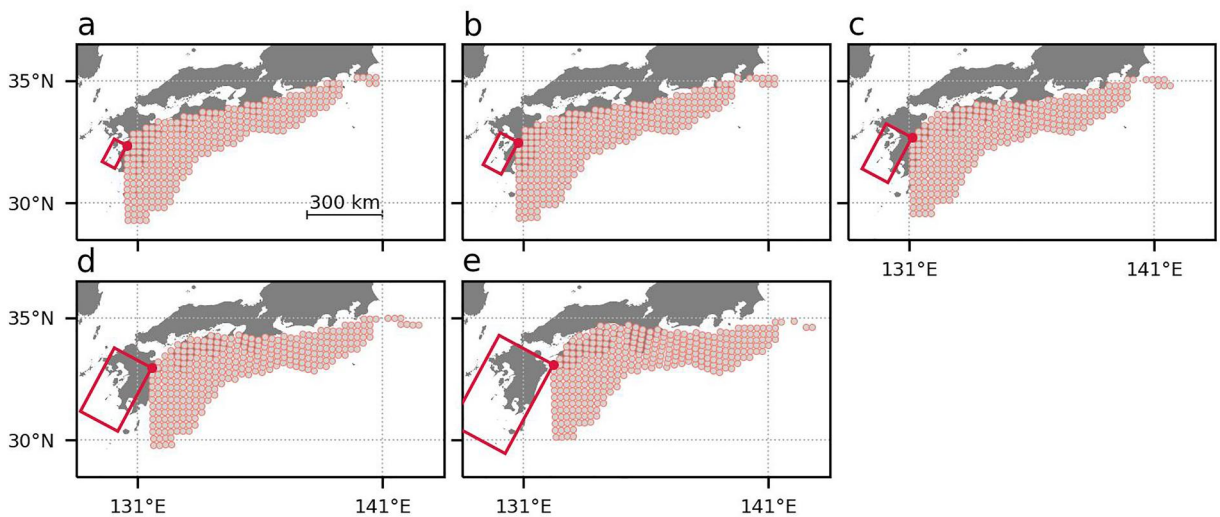

Figure 1. Locations of 1564 hypothetical faults. Panels (a–e) correspond to magnitudes 7.6, 7.9, 8.2, 8.5, and 8.8, respectively. The gray circles indicate the northeast apex of rectangular faults, and the red rectangle with a red circle in each panel is an example of fault geometry.

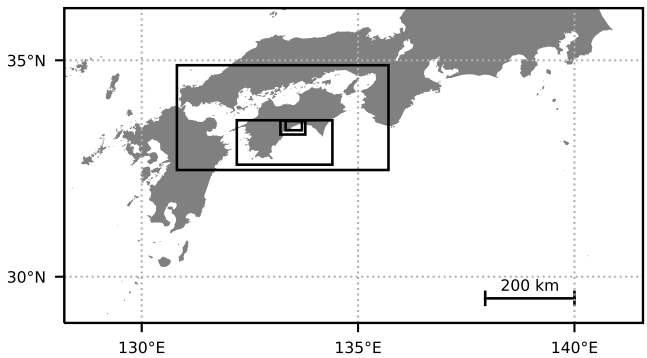

Figure 2. Tsunami simulation domain targeting the Shikoku regions, with five-level nesting.

using the Okada model (Okada, 1992). We carried out 1,564 tsunami simulations with the five-level nesting shown in Figure 2, and the wave heights were stored every five seconds. Additionally, the 1,564 scenario simulations are an extension of a data set of 666 Nankai Trough earthquake/tsunami scenarios (Koshimura & Nomura, 2022) under almost identical conditions.

All 1,564 scenarios were divided into test and training data sets for validation purposes. Specifically, 90% of the 1,564 scenarios, that is, 1,414 scenarios, were randomly selected as training data and used for the optimization procedure presented in the previous section to determine the optimal sparse arrangement of p gauges among the n synthetic gauges to realize the virtually dense observation network. Then, we examined the feasibility of the determined p gauges by considering whether the remaining $n - p$ gauge information can be reasonably reconstructed based on the PSR formula in Equation 16 for the 150 (=1,564 − 1,414) test scenarios.

3.2. Synthetic Gauge Locations

A total of $n = 134$ synthetic gauges, which are candidate points for the optimal gauge set, were placed in the simulation domain, as shown in Figure 3. The locations of some of these gauges were identical to those of the actual observation gauges, such as the gauges in the NOWPHAS (Ports & Harbours Bureau, 2022) and DONET2 (Kaneda, 2010) networks. Note that two of the NOWPHAS gauges off Kyushu are not included in the candidate points, while all the DONET2 gauges are considered. Thus, there are 1,564 (scenarios) × 134 (gauges) × 6 (hr) × 3,600 (sec/hr) × 1/5 (/sec) data points in total.

3.3. POD Representation for Time Series Data of Tsunami Wave Heights

The POD method is applied to the time series wave heights for all $n = 134$ gauges in the 1,414 training scenarios according to the formulation presented in Section 2.1. The spatial distributions of the first and second modes are shown in Figure 4. The results in panels (a) and (b) confirm that the data collected by the coastal gauges include prominent features for these lower-order modes. In general, the lower-order modes extracted by the POD method represent the global features of the spatial distributions, while higher-order modes reflect local distributions and noise effects. However, the lower-order spatial modes in tsunami wave heights data tend to reflect local effects in near-shore areas, which may be due to the inherent nature of tsunamis; that is, the tsunami wave heights near the coast are much higher than those offshore.

It should be noted that the optimization and wave height PSR capabilities of the present method are both affected by the number of modes used to create the ROM in Equation 4. In fact, as reported by Manohar et al. (2018), $p \geq r$ optimal sensors generally achieve high reconstruction performance with appropriately reduced-order data using the lower-order spatial modes extracted by the POD method. On the other hand, the use of higher-order modes leads to overfitting of the noise components, reducing the reconstruction accuracy. To determine the number of modes that should be employed, the relative importance of each mode is often measured based on the contribution rate.

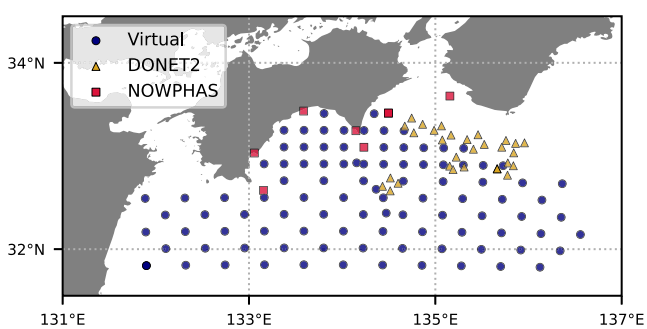

Figure 3. Locations of the $n = 134$ synthetic gauges. The gauges placed in identical locations to the gauges in the existing networks are represented with red squares (NOWPHAS) and yellow triangles (DONET2). In contrast, the other synthetic gauges are shown as virtual gauges with blue circles.

Figure 5 shows the contribution rates of the ith POD modes, which were calculated as

$$c(i) = \frac{\sigma_i}{\sum_{j=1}^{n} \sigma_j}. \tag{23}$$

The cumulative contribution rates are also shown in the figure. The figure shows that the lower-order modes have relatively high contribution ratios, but their values are not very large. Therefore, a large number of modes are needed to represent the original data; for example, the red lines in Figure 5 show that 81 modes are required to represent 90% of the original data. However, the

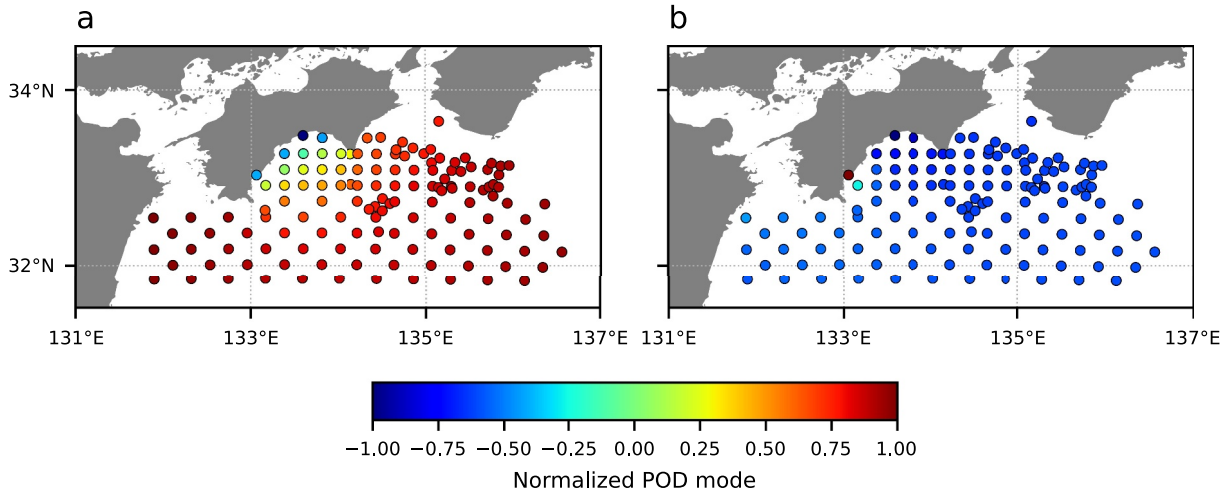

Figure 4. Distributions of the proper orthogonal decomposition (POD) spatial modes extracted from the training data of the time series tsunami wave heights. Panels (a, b) show the normalized values of the first and second POD modes, respectively.

22nd and subsequent modes have contributions of less than 1%. Therefore, we decided to employ $r = 21$ modes to create the ROM for the optimization and wave height PSR processes.

4. Capability Assessment: Results and Discussion

Under the conditions established in the previous section, we demonstrate the ability of the optimization method to determine the optimal sparse arrangement of actual tsunami wave gauges to realize a virtually dense observation network based on the PSR technique.

4.1. Optimization With the Greedy Algorithm

For comparison purposes, we consider sparse networks with three different numbers of optimal gauges: $p = 7$ ($=r/3$), 21 ($=r$), and 42 ($=2r$). The first case ($p = 7$) has the same number of gauges as the NOWPHAS network off

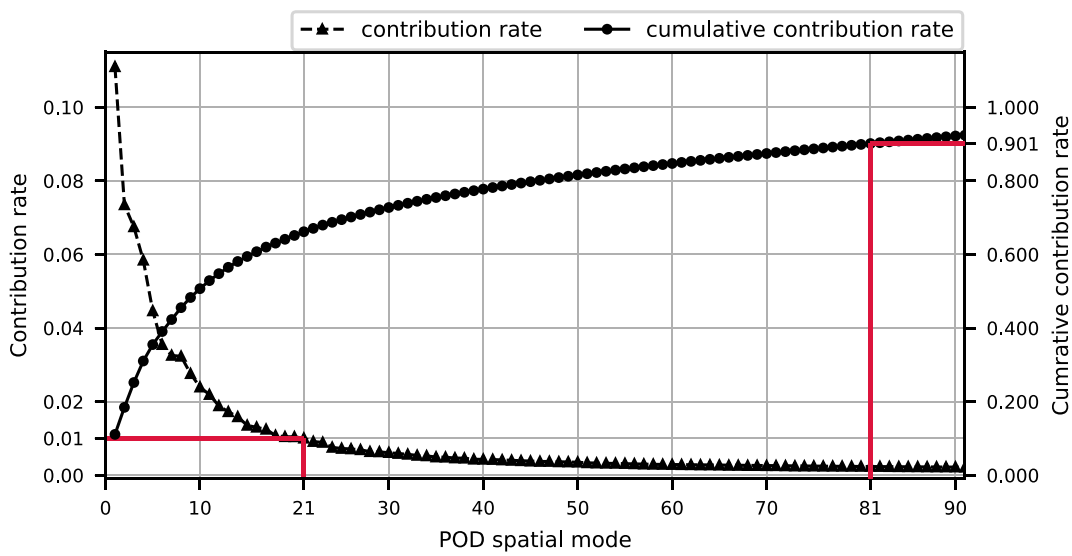

Figure 5. Contribution rates of each proper orthogonal decomposition mode calculated according to Equation 23. The triangles and left axis show the individual contribution rate, while the circles and right axis show the cumulative contribution rate.

10.1029/2023EA003144

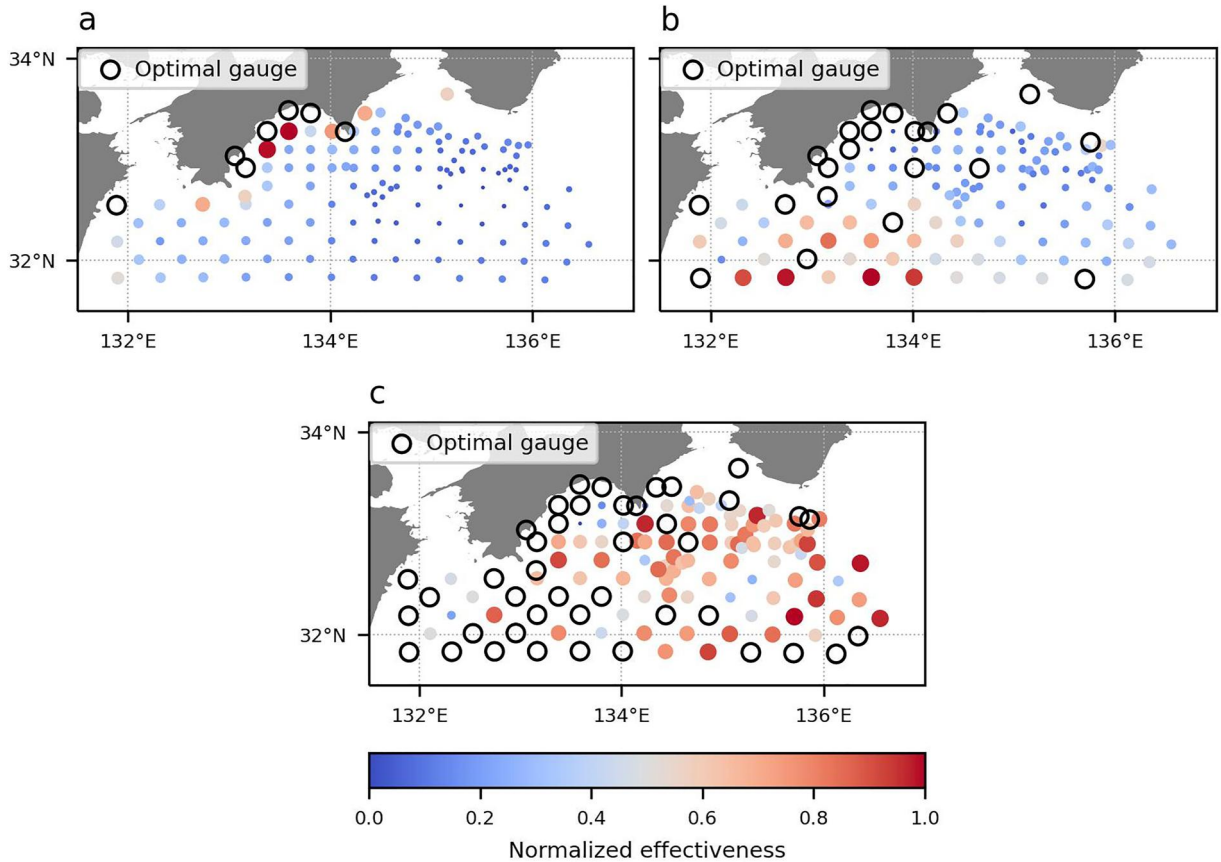

Figure 6. Optimization results determined by the greedy algorithm based on QR decomposition with column pivoting for three cases with different gauge numbers: $p = 7$, 21, and 42. The open circles in panels (a–c) show the locations of the optimal $p = 7$, 21, and 42 gauges selected from the $n = 134$ synthetic gauges, which are represented by the colored circles. The color and size represent the normalized value of the effectiveness, which is used to evaluate the optimality of each synthetic gauge, as presented in Appendix B.

Shikoku, as shown in Figure 3. We note that $r = 21$ was determined in the previous section considering the contribution rate of each spatial mode. To determine the optimal sparse arrangement of p actual wave gauges, the optimization problem formulated in Equation 17 is solved for each case by applying QR decomposition with column pivoting using the contracted mode matrix $\Phi_{r=21}$. Here, the covariance matrix of the noise $\boldsymbol{w}'$ is set to be $\boldsymbol{\Sigma}_w = \boldsymbol{I}$ because the noise in the system is unknown, while the uncertainty is considered in the verification analysis of the proposed optimization scheme in Appendix D.

The optimization was performed to determine the optimal configuration of candidate points consisting of all the synthetic gauges according to the degree of effectiveness, which is a measure of optimality calculated in the QR decomposition process, as explained in Appendix B. Figures 6a–6c show the optimal gauge placements in the cases with $p = 7$, 21, and 42 gauges, respectively. For more detailed information, the optimization process from zero to 20 gauges is shown in Figure B1 of Section Appendix B. Here, the black colored circles in each panel indicate the optimal gauges that show the highest effectiveness among the considered synthetic gauges in each process. The color and size in each plot indicate the normalized effectiveness of the other $n - p$ synthetic gauges. Figure 6a shows that the top seven gauges are preferentially selected near the shore of Shikoku, and some other gauges in the coastal areas exhibit high effectiveness. Additionally, Figure 6b indicates that the sparse network with 21 optimal gauges is designed in both the coastal and offshore areas, and then, effective synthetic gauges exist in the offshore regions. Furthermore, many offshore gauges are selected in the case with 42 optimal gauges, and a dense observation network is designed for both the coastal and offshore areas, as shown in Figure 6c. The tendency that the gauge locations that effectively increase the value of the objective function to be maximized are concentrated in the coastal area is consistent with the distributions of the lower-order POD modes, as shown in Figure 4.

Table 2
Fault Rupture Parameters in a Test Scenario to Validate the Optimization Results

Longitude	[Degrees]	135.12 E
Latitude	[Degrees]	33.12 N
Magnitude	[Mw]	8.8
Strike	[Degrees]	214.41
Dip	[Degrees]	8.42
Rake	[Degrees]	90
Length	[km]	458.39
Width	[km]	229.19
Depth	[km]	12.30
Dislocation	[m]	4.65

Notably, the sparse observation networks with $p = 7$, 21, and 42 gauges include three, five, and five gauges from the existing NOWPHAS network, respectively. Our results qualitatively suggest that the existing NOWPHAS observation network, which has a coastal neighbor configuration, is effective for grasping wide-area information on tsunamis based on the PSR technique. However, in this study, we focused on understanding the wide information based on the PSR technique in our formulation of the objective function in Equation 17. Therefore, if early tsunami detection or forecasting is considered, observations in offshore areas should be more important than those in non-offshore areas.

Additionally, since the most effective gauge is selected sequentially in the optimization process based on the greedy algorithm, each gauge's effectiveness in improving the PSR accuracy is quantified independently. Due to this feature, if some environmental conditions or installation costs make the installation of a selected gauge difficult, the next most effective gauge can be included in the set of optimal gauges. Therefore, the greedy algorithm is more advantageous than other approaches, such as direct search and heuristic methods, which determine the gauge combination in terms of network suitability.

4.2. PSR Approach for Creating a Virtually Dense Observation Network

To verify the capability of the PSR technique formulated in Equation 16 to create virtually dense observation networks, we test the adequacy of the optimal sparse arrangement of the synthetic gauges. Specifically, the optimal configurations determined above are used to construct virtually dense networks by estimating the high-resolution wave height distribution based on the Kalman filter as expressed in Equation 16. Here, we employ the initial state vector $\hat{\boldsymbol{a}}_0 = \boldsymbol{0}$, the covariance matrix $\boldsymbol{P}_0 = \boldsymbol{I}$, and the covariance matrix of the system noise $\boldsymbol{\Sigma}_v = 0.01\boldsymbol{I}$. The targeted test scenario for capability validation has the specific fault parameters summarized in Table 2. Figure 7 shows the wave height snapshot at $t = 30$ min simulated by TUNAMI-N2 for the test scenario. This snapshot is compared with the reconstruction results obtained by the PSR technique using observations at p optimized gauges. In the PSR relationships formulated in Equation 6 and Equation 16, the number of modes is set to $r = 21$, as in the optimization process.

Figure 8 shows the wave height snapshots obtained using the sparse observation networks with $p = 7$, 21, and 42 synthetic gauges, with the gauge locations determined by the above optimization process. In this figure, the results of a network with randomly selected gauges are also shown for reference. The top row shows the optimal $p = 7$, 21, and 42 gauge locations, the middle row shows the wave height distributions reconstructed by the PSR technique, and the bottom row shows the absolute error (AE) distributions between the reconstructed snapshot and the TUNAMI-N2 simulation result for the test scenario shown in Figure 7. The results in panels (b) and (c) show that the absolute errors are prominent near the coast when $p = 7$ gauges are used in the optimal

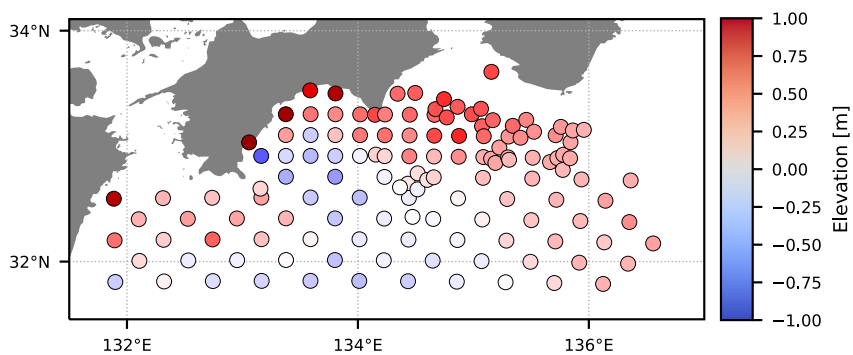

Figure 7. Tsunami wave height snapshot acquired by each synthetic gauge at 30 min. The snapshot is simulated by the TUNAMI-N2 model for a test scenario, with the fault parameters summarized in Table 2.

10.1029/2023EA003144

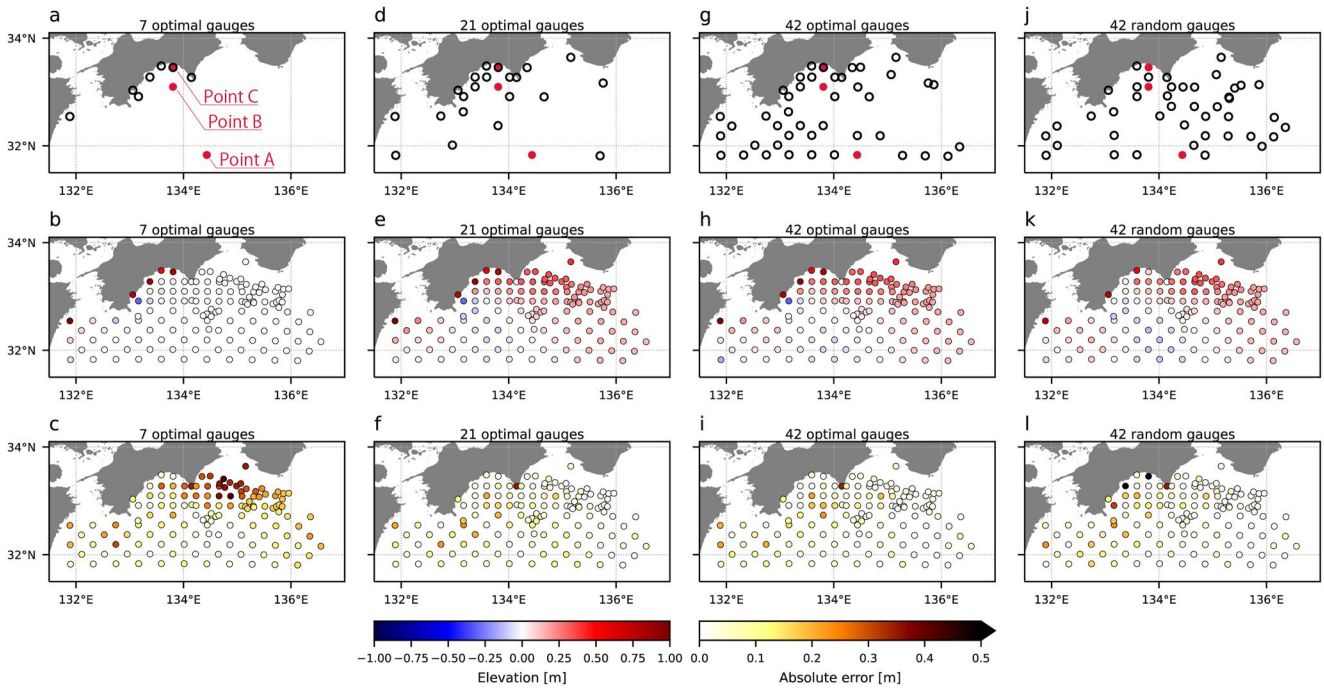

Figure 8. Comparison of wave height PSR results among the $p = 7$, 21, 42 optimal and $p = 42$ random gauge configurations. (a) Shows the seven optimal gauge locations. (b, c) show the wave height snapshot at 30 min obtained by the PSR technique, and the absolute error distribution between the wave height snapshots of the PSR result and the TUNAMI-N2 simulation, respectively, based on the seven optimal gauges. (d–f) are the results based on the optimal configuration with 21 gauges; (g–i) correspond to the optimal 42 gauges; and (j–l) correspond to the random configuration of 42 gauges. Points A, B, and C in panel (a) are points of interest for comparing the wave histories of TUNAMI-N2 and the PSR results based on the sparse configurations.

configuration. Thus, the PSR-based estimates do not provide sufficient high-resolution wave height distribution when only $p = 7$ gauges are used in the optimal configuration.

Since increasing the number of optimal gauges improves the wave height PSR accuracy, the reconstructed snapshot with $p = 21$ gauges in the optimal configuration shown in panel (d) has a smaller error than the previous case with $p = 7$ gauges, as shown in Figures 8f and 8c. Note that the optimal $p = 7$, 21, and 42 gauges are selected sequentially with the same optimization process; therefore, the positions of the $p = 7$ gauges remain fixed in the case with $p = 21$ gauges. Thus, the additional optimal gauges shown in panel (d) are selected after the $p = 7$ gauges are optimized, and the additional gauges are placed in the areas where relatively large errors occur with the virtually dense observation network consisting of $p = 7$ optimal gauges. As seen from panel (e), the reconstructed wave height snapshot is similar to that obtained by TUNAMI-N2, and the configuration with $p = 21$ optimal gauges provides accurate high-resolution wave height distribution for the entire space, allowing a reliable virtually dense observation network to be constructed with errors of less than 0.36 m. Additionally, we confirm that this configuration has an equivalent PSR ability to that of the configuration with $p = 42$ optimal gauges, as shown in panel (g), based on a comparison of the AE distributions shown in panels (f) and (i). Thus, the wave height information at the n gauge locations in the virtually dense observation network can be obtained from sparsely placed gauges. Although the network with 21 gauges includes more gauges than the existing NOWPHAS network off Shikoku, the number of gauges is somewhat reasonable compared to the 29 gauges included in the DONET2 network.

The good performance of the optimization scheme is demonstrated by the relative advantage of the network with optimally selected gauges compared to that with randomly configured gauges. Figure 8j shows a configuration with 42 randomly selected gauges. By applying the PSR technique to this configuration, we obtain the wave heights at all synthetic gauges, as shown in panel (k), and the AE distribution with respect to the TUNAMI-N2 result, as shown in panel (l). These figures indicate that the wave height PSR accuracy of the configuration with $p = 42$ random gauges degrades along the coast, with errors of more than 0.5 m. Furthermore, the maximum AE

10.1029/2023EA003144

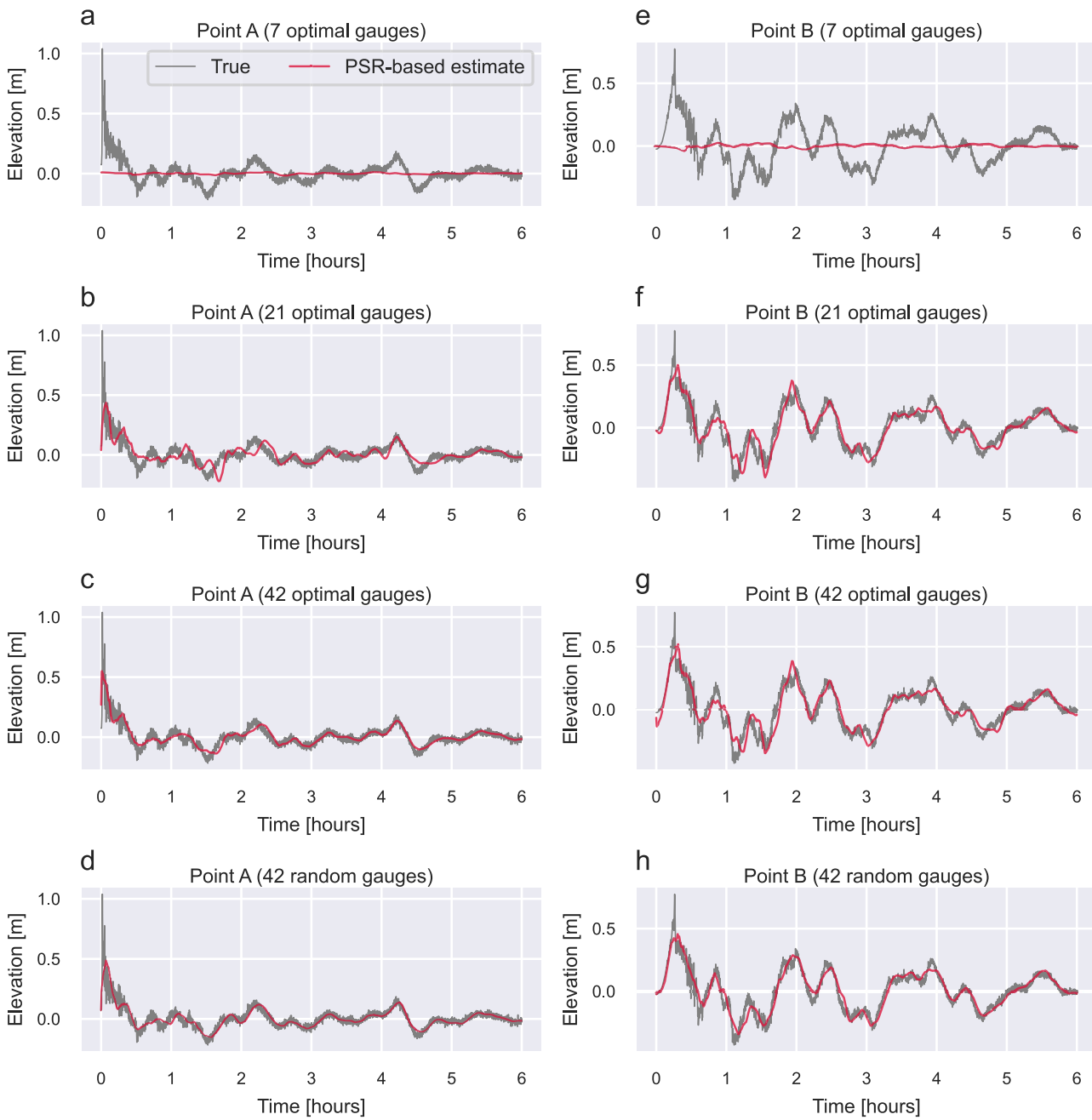

Figure 9. Comparison of time series wave heights obtained by TUNAMI-N2 and those reconstructed via the PSR technique. (a–d) show the wave height changes at Point A in Figure 8a based on the optimal gauge configuration with $p = 7$, 21, and 42 and the random gauge configuration with $p = 42$. (e–h) show those at Point B under the same conditions.

reaches 0.8 m and is more significant than the error of the optimal configurations obtained above. Therefore, the results demonstrate the superiority of the present optimization scheme over a random selection scheme.

To confirm the performance of the PSR method, the time series of the wave heights obtained at Points A and B indicated in the panels in the top row of Figure 8 are shown in Figure 9 in comparison with those obtained from the TUNAMI-N2 simulation. These evaluation points are located at synthetic gauges that were not selected in either the optimal gauge configurations with $p = 7$, 21, and 42 or the random gauge configurations with $p = 42$.

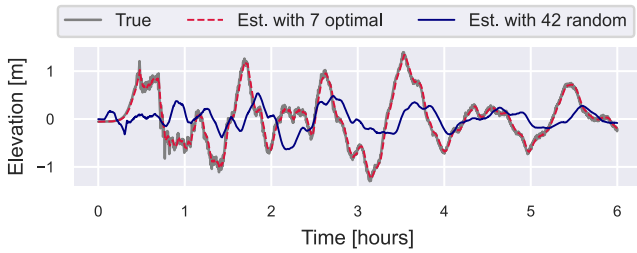

Figure 10. Comparison of wave heights obtained by TUNAMI-N2 and those reconstructed via the PSR technique based on the $p = 7$ optimal and $p = 42$ random gauge configurations. These waveforms are obtained at Point C shown in Figure 8a.

As seen from Figures 9a and 9e, the sparse configuration with seven optimal gauges has no PSR capability to reconstruct the wave heights because the wave heights at these points are almost zero the entire time. On the other hand, as shown in panels (b), (c), (f), and (g), the PSR method using 21 optimal gauges captures the reference waveform, and the predictions are comparable to those obtained using 42 optimal gauges. Thus, it is confirmed that the PSR method based on the optimized gauge configuration is capable of reconstructing the time series data of the wave heights, although its performance depends on the number of gauges selected. Meanwhile, as can be seen from panels (d) and (h), the configuration with 42 randomly selected gauges also has high performance, but not all gauges have good prediction accuracy. In fact, Figure 10, which illustrates the time series wave heights at Point C indicated in Figure 8a, shows that the PSR method with 42 randomly selected gauges fails to predict. This failure can be inferred from the fact that Point C corresponds to the coastal gauge that showed the large error shown in Figure 8l. We also include the result using seven optimal gauges with a red dashed line in Figure 10. Since the optimal gauge arrangement contains Point C, the resultant estimate naturally traces the true waveform.

To quantitatively compare the PSR results shown in Figure 8, we calculate the PSR errors of the wave heights obtained throughout the entire elapsed time window for all test scenarios and all synthetic gauges. The root mean square error (RMSE) is employed as the evaluation index:

$$\mathrm{RMSE} = \sqrt{\frac{1}{l \cdot m \cdot n} \sum_{j=1}^{l} \sum_{t=1}^{m} \sum_{i=1}^{n} \left| \boldsymbol{x}_{i,t}^{j} - \hat{\boldsymbol{x}}_{i,t}^{j} \right|^2}. \tag{24}$$

where l, m, and n are the total numbers of test scenarios, simulation time steps, and synthetic gauges, respectively. Additionally, $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ are the wave heights obtained in the TUNAMI-N2 simulation and those reconstructed using each of the four sparse configurations, respectively. Table 3 summarizes the RMSE of the PSR results obtained for the three optimal configurations and one random configuration discussed above. Here, the RMSE value of the random gauge configuration is the average value for 100 trials. This table shows that the PSR performance for constructing the virtually dense observation networks is improved by increasing the number of optimal gauges. In addition, an optimally designed sparse observation network with 21 gauges is more economical and more informative than an observation network with randomly selected gauges.

4.3. Potential Gauge Locations for Extending the NOWPHAS Network

In the previous subsection, we sequentially identified effective gauges in turn from the beginning, that is, without arranging the initial gauges, to optimize the gauge configurations. However, as shown in Figure 3, an existing gauge network known as NOWPHAS can be used as the initial configuration in the optimal design. Therefore, we fixed the seven NOWPHAS gauges installed in the target area as initial conditions to optimize additional gauge locations, starting with the eighth gauge. The optimized additional locations indicate potential locations of the existing NOWPHAS in terms of improving wave height PSR ability. Then, the resulting virtually dense observation network with optimal gauges, including the existing NOWPHAS stations, is compared with that constructed in the previous subsection.

Indexing the seven existing NOWPHAS gauges as γ_p ($p = 1, \ldots, 7$), we calculate the initial distribution of the effectiveness that evaluates the optimality of the remaining $n - p$ synthetic gauges as explained in Section 2.3.2. The results are shown in Figure 11. This figure shows that the red-colored coastal gauges are informative for the wave height PSR approach and can therefore be considered additional optimal gauges. This implies that although the existing NOWPHAS network is installed in coastal areas, the priority locations for adding the new observation instruments are also near

Table 3
Comparison of the Wave Height PSR Errors Between Each Gauge Configuration and the TUNAMI-N2 Simulation

Gauge configuration	RMSE
Optimal 7 gauges	8.61×10^{-2} [m]
Optimal 21 gauges	4.02×10^{-2} [m]
Optimal 42 gauges	3.51×10^{-2} [m]
Randomly selected 42 gauges	8.06×10^{-2} [m]

Note. The root mean square error (RMSE) is defined in Equation 24 and calculated for all test scenarios, time steps, and synthetic gauges.

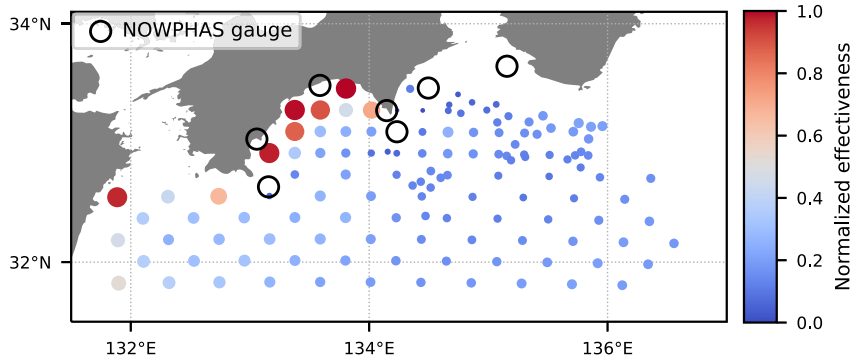

Figure 11. Normalized value of the effectiveness, which evaluates the synthetic gauge optimality, considering the existing NOWPHAS network as an initial condition. The open circles indicate the locations of the seven NOWPHAS gauges actually installed in the Shikoku region, and the color and size indicate the magnitude of the normalized effectiveness.

the coast, similar to the optimal locations of the seven gauges shown in Figure 8a. It is worthwhile to note that this kind of prediction prior to optimization can be made thanks to the use of the greedy algorithm based on QR decomposition with column pivoting.

Figure 12 shows the relationship between the number of optimal gauges, p, and the RMSE calculated using Equation 24 for all the test scenarios and observation times. This figure shows that the RMSE of the optimal configuration considering the existing NOWPHAS network decreases as the number of installed gauges increases. Interestingly, although the objective function in Equation 17 is intended to reduce the error with respect to the POD coefficients, the RMSE of the wave height estimations actually decreases as p increases. Additionally, the RMSE is comparable with that of the network without the initial seven gauges, and the RMSE does not significantly decrease after the 21st gauge is added. As described in Section 2.3.2, the figure shows that the decreasing trend in error changes before and after the 21 gauges, which is consistent with the number of modes used. Thus, the wave height PSR accuracy can be improved by adding the optimal gauges to the existing network, although it may be ideal to design the actual observation network using the optimization framework starting from the initial step. Moreover, the result of the random network shown in the same figure, which is the average of 100 random trials, has a larger RMSE than these two optimal configurations.

According to the trends of the errors in Figure 12, the optimal gauge network considering the NOWPHAS locations with $p = 21$ gauges (seven existing and 14 optimally added gauges) might construct a sufficient virtually dense observation network. Figure 13 shows the locations of the 21 gauges in this network. The figure shows that

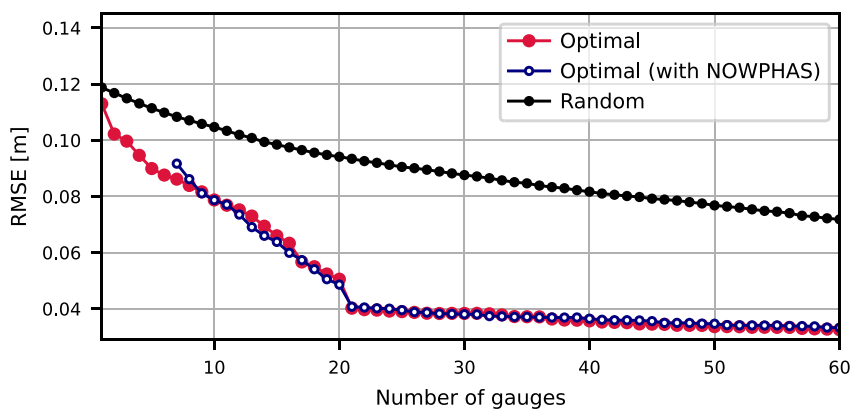

Figure 12. The relationship between the estimation error and the number of optimal gauges. The red line shows the root mean square error of the virtually dense network consisting of p optimal gauges, while the blue line shows that consisting of $p - 7$ optimal and seven existing NOWPHAS gauges. The black line shows the result of the configuration with randomly selected gauges.

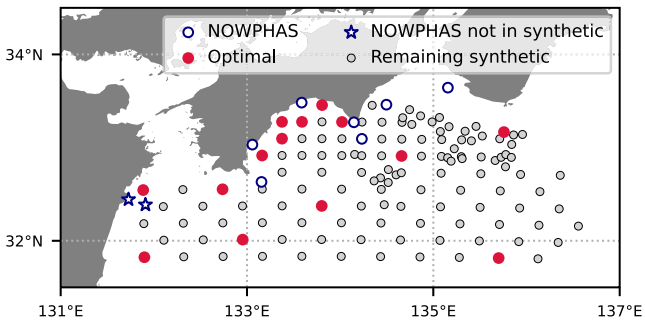

Figure 13. Configuration of the seven existing NOWPHAS and 14 optimally added gauges. The red circles show the 14 optimal gauge locations, and the blue circles represent the existing seven NOWPHAS gauges. The two blue stars indicate the existing NOWPHAS gauges that are not reflected in the synthetic gauge arrangement. The gray circles show the remaining synthetic gauges, which correspond to candidate points for the subsequent optimization processes.

these 21 gauges are configured similarly to those in the network obtained in the previous subsection, with 19 shared gauges, as shown in Figure 8d. Interestingly, the synthetic gauge off Kyushu selected by the optimization method is close to two existing NOWPHAS gauges (indicated with star-shaped markers). As mentioned in Section 3.2, these two NOWPHAS gauges off Kyushu were not considered candidate points in the optimization because this study focused on the Shikoku coast. Since the two existing gauges are proxies for the selected synthetic gauge off Kyushu, they should have been selected to obtain more accurate PSR results. In addition, if the optimal gauges are added to the existing NOWPHAS network, including the gauges indicated by star-shaped markers, they should be placed near the Shikoku coast rather than near the Kyushu coast and, in turn, several offshore locations.

Considering the several types of instruments available for tsunami observation, we discuss the feasibility of the sparse observation network with 21 optimal gauges, including the seven NOWPHAS gauges and 14 synthetic gauges obtained above. Among the 14 sites where the additional gauges should be installed, those near the Shikoku coast can be covered by extending the NOWPHAS network. In contrast, the wave heights at the offshore locations should be observed by networks other than NOWPHAS since the installation area of NOWPHAS gauges is limited, that is, they must be placed less than 20 km from the coast (Kato et al., 2018). The wave heights in such offshore locations can be obtained from the hydraulic pressure observations recorded by seafloor cable network systems, for example, DONET and N-net. In fact, the observational gauges in the DONET system are installed on the east side of the Shikoku offshore area, and the gauges of the N-net system are being deployed on the west side. Thus, a sparse network with highly accurate PSR capabilities for tsunami wave heights can be realized by combining observations from several networks: an extended NOWPHAS with additional gauges, DONET, and the under-development N-net system.

5. Conclusion

We explored an optimal sparse sensor selection method introduced in the literature (Manohar et al., 2018) to establish a framework to determine the optimal sparse configuration for tsunami gauges to reconstruct the high-resolution spatial distribution of tsunami wave heights throughout the target region. High-resolution wave height distribution is obtained by constructing virtually dense observation networks based on the PSR technique. The computational results confirm that the optimized sparse configuration can reconstruct the spatial information of tsunamis.

The established framework consists of three procedures. We first acquire tsunami wave heights at synthetic gauges by conducting numerical simulations for various earthquake and tsunami scenarios at a target site. Next, the features of the spatial distribution of tsunami heights are extracted by the POD method. Finally, we solve an optimization problem with a greedy algorithm to determine the optimal sparse configuration of the candidate points consisting of all synthetic gauges. To demonstrate the capability of the present approach, we constructed optimal gauge configurations off Shikoku, with the aim of addressing hypothetical Nankai Trough earthquakes. The results showed that the determined $p = 7$, 21, and 42 optimal gauge configurations were reasonable and superior to a randomly selected network in terms of the wave height PSR accuracy. In addition, 21 optimal observation gauges were sufficient for adequate PSR performance.

Furthermore, the performance of the existing NOWPHAS observation network was assessed by conducting another optimization by setting the actual locations of the seven gauges as initial conditions. The results show that the existing NOWPHAS network has equivalent PSR performance to the optimal gauge network determined by the present optimization framework. However, since the gauges in the optimal configuration tend to be concentrated in coastal areas in the present validation result, the configuration is unsuitable for early tsunami detection or forecasting. This is because the optimization objective function considers only the accuracy of the wave height PSR and does not consider the immediacy of the prediction, which should be considered in future work.

Appendix A: Optimization Objective Function in Terms of Information Entropy Minimization

According to the following equation expansion, the optimization objective function in Equation 17 can be derived in terms of minimizing the information entropy (Shannon, 1948), which is used to evaluate system uncertainty. The information entropy of a random variable $\boldsymbol{X}$ sampled according to the probability density function $p(\boldsymbol{x})$ is defined as

$$h(\boldsymbol{X}) = \int p(\boldsymbol{x}) \log p(\boldsymbol{x}) d\boldsymbol{x}. \tag{A1}$$

Here, assuming that the probability density function $p(\boldsymbol{x})$ is a multivariate normal distribution with a covariance matrix $\boldsymbol{P}_x$, we can expand this equation as follows:

$$h(\boldsymbol{x}) = \frac{1}{2} \log[.(2\pi e)^n \det \boldsymbol{P}_x]. \tag{A2}$$

Thus, the information entropy of a random variable that follows a Gaussian distribution depends on its covariance matrix $\boldsymbol{P}_x$.

As written in Section 2.2, the high-resolution wave height distribution can be obtained by estimating the POD coefficients using the wave data y_t observed at p gauges in a sparse observation network. Considering the result of the POD coefficient estimation based on the Kalman filter in Equation 13 and Equation 14, we can recognize that the posterior distribution of the POD coefficients follows a multivariate normal distribution with mean $\hat{\boldsymbol{a}}_t$ and covariance matrix $\boldsymbol{P}_t$.

Finally, because the estimated POD coefficient follows a Gaussian distribution, as mentioned above, its information entropy depends on the posterior covariance matrix $\boldsymbol{P}_t$, whose reduction rate from the prior estimation of the covariance matrix $\boldsymbol{P}_t^-$ is defined as follows from the Kalman filter relation in Equation 14:

$$\Delta \boldsymbol{P} = \boldsymbol{P}_t^{-1} - \left(\boldsymbol{P}_t^-\right)^{-1} = \left(\boldsymbol{\Phi}_r^p\right)^{\mathrm{T}} \left(\boldsymbol{\Sigma}_w^p\right)^{-1} \boldsymbol{\Phi}_r^p. \tag{A3}$$

The right-hand side of this equation coincides with the matrix to be maximized in the optimization objective function shown in Equation 17. Therefore, optimization with determinant-based maximization is equivalent to maximizing the reduction rate of the covariance matrix; that is, the posterior covariance matrix is minimized. In other words, the information entropy, which depends on the posterior covariance matrix, is minimized. As a result, we can obtain the optimal configuration while avoiding locations that are sensitive to observation noise.

Appendix B: Greedy Algorithm Based on QR Decomposition With Column Pivoting

The optimal gauges are selected sequentially in four steps using QR decomposition with column pivoting according to the following algorithm.

Step 1

For simplicity, we denote a part of the left-hand side of the QR decomposition equation Equation 21 for the optimization as follows:

$$\boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Phi}_r \boldsymbol{\Phi}_r{}^{\mathrm{T}} \boldsymbol{\Sigma}^{-\frac{1}{2}} = \boldsymbol{B}^{(1)} = \begin{bmatrix} | & | & & | \\ \boldsymbol{b}_1^{(1)} & \boldsymbol{b}_2^{(1)} & \cdots & \boldsymbol{b}_n^{(1)} \\ | & | & & | \end{bmatrix}, \tag{B1}$$

where $\boldsymbol{B}^{(1)}$ is the target matrix of the QR decomposition in the first iteration. Among the column vectors of $\boldsymbol{B}^{(1)}$, we extract the vector with the largest ℓ_2 norm as $\boldsymbol{b}_{\gamma_1}^{(1)}$ and calculate a householder matrix $\boldsymbol{H}^{(1)}$ that satisfies the following relations:

$$\boldsymbol{H}^{(1)}\boldsymbol{b}_{\gamma_1}^{(1)} = \boldsymbol{b}_{\gamma_1}^{(2)}, \tag{B2}$$

$$\boldsymbol{H}^{(1)}\boldsymbol{b}_{\gamma_1}^{(2)} = \boldsymbol{b}_{\gamma_1}^{(1)}. \tag{B3}$$

Here, the vectors $\boldsymbol{b}_{\gamma_1}^{(1)}$ and $\boldsymbol{b}_{\gamma_1}^{(2)}$ are defined as

$$\boldsymbol{b}_{\gamma_1}^{(1)} = \begin{Bmatrix} b_{1,\gamma_1}^{(1)} \\ b_{2,\gamma_1}^{(1)} \\ \vdots \\ b_{r,\gamma_1}^{(1)} \end{Bmatrix}, \quad \boldsymbol{b}_{\gamma_1}^{(2)} = \begin{Bmatrix} \left\| \boldsymbol{b}_{\gamma_1}^{(1)} \right\|_2 \\ 0 \\ \vdots \\ 0 \end{Bmatrix}. \tag{B4}$$

Step 2

The column vector $\boldsymbol{b}_{\gamma_1}^{(1)}$ that has the largest ℓ_2 norm among the column vectors in $\boldsymbol{B}^{(1)}$ in the first iteration is inserted into the first column of $\boldsymbol{B}^{(1)}$ by column pivoting. For the second iteration, the target matrix is derived from the column-pivoted matrix $B^{(1)}$ and the householder matrix $\boldsymbol{H}^{(1)}$ as follows:

$$\boldsymbol{B}^{(2)'} = \boldsymbol{H}^{(1)}\boldsymbol{B}^{(1)} = \begin{bmatrix} \left\| \boldsymbol{b}_{\gamma_1}^{(1)} \right\|_2 & b_{1,1}^{(2)'} & \cdots & b_{1,n}^{(2)'} \\ 0 & & & \\ \vdots & & \boldsymbol{B}^{(2)} & \\ 0 & & & \end{bmatrix}. \tag{B5}$$

Step 3

Then, $\boldsymbol{b}_{\gamma_2}^{(2)}$ is defined according to the magnitude of the ℓ_2 norm of the submatrix $\boldsymbol{B}^{(2)}$, and the householder matrix $\boldsymbol{H}^{(2)}$ is calculated as in Equation B1 and Equation B3. The householder matrix $\boldsymbol{H}^{(2)'}$ for the second iteration is calculated by replacing a submatrix of the unit matrix with $\boldsymbol{H}^{(2)}$ as follows:

$$\boldsymbol{H}^{(2)'} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boldsymbol{H}^{(2)} & \\ 0 & & & \end{bmatrix}. \tag{B6}$$

Step 4

The target matrix $\boldsymbol{B}^{(3)}$ in the third iteration is computed from $\boldsymbol{H}^{(2)}$ and $\boldsymbol{B}^{(2)}$ as

$$\boldsymbol{B}^{(3)'} = \boldsymbol{H}^{(2)'}\boldsymbol{B}^{(2)'} = \begin{bmatrix} \left\| \boldsymbol{b}_{\gamma_1}^{(1)} \right\|_1 & b_{1,\gamma_2}^{(2)} & b_{1,1}^{(2)} & \cdots & b_{1,n}^{(2)} \\ 0 & \left\| \boldsymbol{b}_{\gamma_2}^{(2)} \right\|_2 & b_{2,1}^{(3)'} & \cdots & b_{2,n}^{(3)'} \\ \vdots & 0 & & & \\ \vdots & \vdots & & \boldsymbol{B}^{(3)} & \\ 0 & 0 & & & \end{bmatrix}. \tag{B7}$$

10.1029/2023EA003144

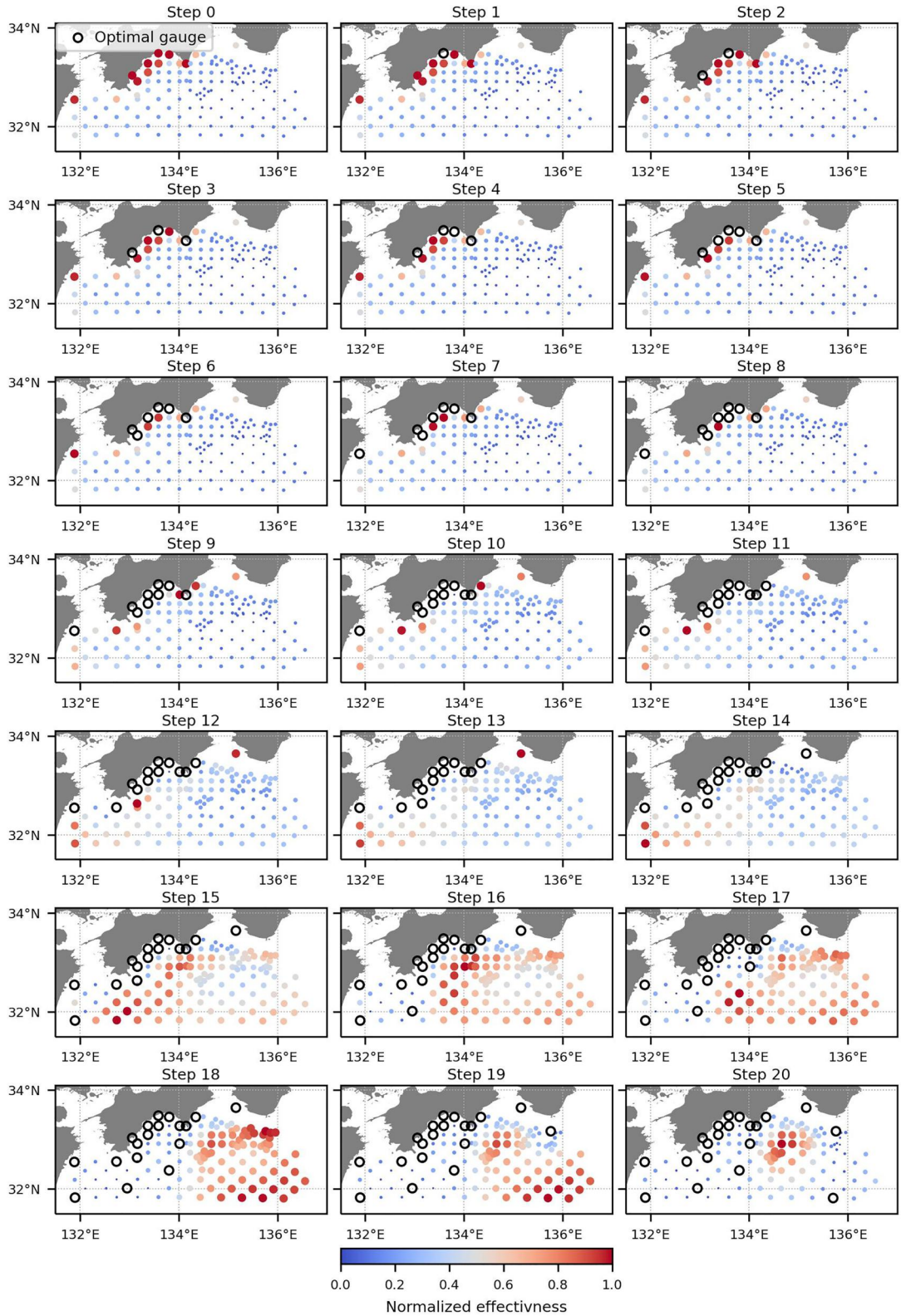

Figure B1. Optimization results determined by the greedy algorithm based on QR decomposition with column pivoting. Each figure shows the optimal gauge locations with the open circles from zero to 20 gauges. The color and size represent the normalized value of the effectiveness, which evaluates the optimality of each synthetic gauge.

By repeating Step 3 and Step 4 for the number of gauges to be optimized, the final target matrix $\boldsymbol{B}^{(p)'}$ becomes an upper triangular matrix that is equivalent to $\boldsymbol{R}$. Hence, the QR decomposition algorithm with column pivoting approximates the determinant-based optimization approach by sequentially maximizing the diagonal components of $\boldsymbol{R}$.

The objective function is defined as a product of the diagonal components r_{jj}, which are calculated as the maximum among the ℓ_2 norms of the column vectors. Thus, the candidate points are ranked in terms of their suitability for inclusion in the optimal configuration in each sequential optimal selection process in accordance with the ℓ_2 norms of the column vectors of the objective matrix $\boldsymbol{B}^{(i)}$. Therefore, the ℓ_2 norm is considered an indicator of the degree of effectiveness in increasing the objective function value and is used to quantitatively measure the optimality of adding a gauge.

The greedy algorithm based on QR decomposition with column pivoting sequentially selects the optimal gauge while ranking each gauge optimality, as described above. Figure B1 shows the optimal gauge selection process from the initial to 20 gauges. The color and size represent the normalized effectiveness of the remaining synthetic gauges. From all the results, representative results (7, 21, and 42 optimal gauge selection) are shown and discussed in Section 4.1. From the figure, we can confirm that some coastal gauges that are finally selected as the optimal set have high effectiveness from the beginning as discussed in Section 4.1.

Appendix C: Comparison Between the Greedy and Genetic Algorithms

To validate the optimization performance of the greedy algorithm based on QR decomposition with column pivoting employed in this study, we compare the optimization result of the greedy method with that of a genetic algorithm (Holland, 1992), which is a well-known type of heuristic algorithm for solving combinatorial optimization problems. As explained in Section 4.2, the greedy algorithm determines the optimal arrangement by sequentially adding a new optimal gauge to the configuration consisting of the previously optimized gauges. In contrast, a genetic algorithm would solve the optimization problem by determining the optimal combination of gauges among every evaluated set of p gauges. In other words, the optimal configuration of $p-1$ gauges obtained by solving the optimization problem with a genetic algorithm will not necessarily be contained within the corresponding optimal configuration of p gauges.

Genetic algorithms search for the global solution to an optimization problem with multiple individuals, which serve as solution candidates. Through generational updates, individuals with good fits are preferentially selected, and the algorithm is updated to search for solutions through repeating operations such as crossover and mutation. This study employs a basic genetic algorithm with tournament selection and uniform crossover, although various types of genetic algorithms have been reported (Katoch et al., 2021). After some trials to determine the settings of the genetic algorithm, we set the number of individuals to 100 and the probability of crossover and mutation to 90% and 10%, respectively. In addition, to maintain solution diversity, the elite conservation strategy, which preserves the 20 individuals with the top fits of each generation for the next generation, is employed in combination with tournament selection. The solution is determined to converge when the same individual is selected for 2,000 consecutive generations.

We compare the optimal configurations determined by the two algorithms in terms of the accuracy of the wave height PSR, which is defined by the relationship in Equation 16 based on the Kalman filter, using the contracted POD mode matrix $\boldsymbol{\Phi}_{r=21}$, as presented in Section 4.2. Here, the genetic algorithm directly evaluates the determinant of the objective function shown in Equation 17, while the greedy algorithm uses the expanded equation shown in Equation 21. Figure C1 shows the relationship between the number of gauges to be optimized by these two algorithms and the wave height PSR errors for all the test scenarios and observation times. Here, the RMSE, which is calculated as shown in Equation 24, is employed as the evaluation index for the error. The figure shows that the optimal configurations determined by these two algorithms have equivalent PSR capabilities since the RMSEs are comparable for each number of gauges. Coincidentally, the results seem to show nearly identical error trends. However, as explained earlier, different data sets will not necessarily yield the same results because the optimization methods are inherently different. We believe that this can be demonstrated on other occasions in the future.

Next, we compare the speeds of the greedy algorithm based on QR decomposition with column pivoting and the genetic algorithm. To solve optimization problems for all the considered numbers of optimal gauges ($p = 1$–133),

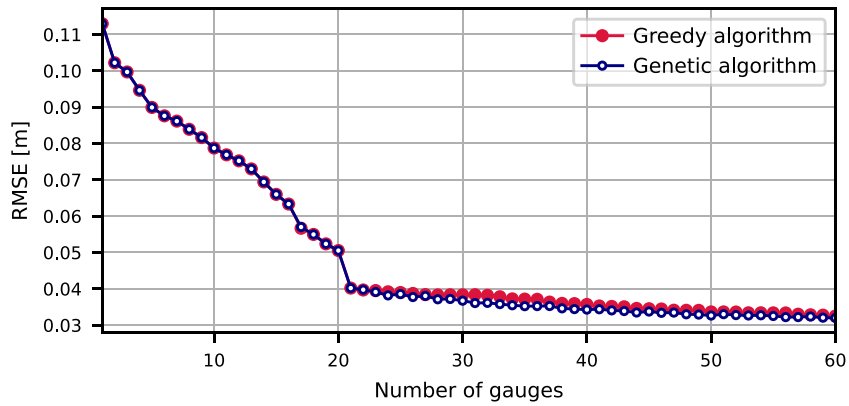

Figure C1. Comparison between the greedy and genetic algorithms. The two lines show the relationship between the number of optimal gauges and the root mean square error of the optimal gauges selected by the greedy algorithm (red) and the genetic algorithm (blue).

the greedy algorithm requires approximately 30 s using a computer (DeepLearning BOX/Alpha, CPU: AMD EPYC 7502P 32 cores/64 thread 2.50 GHz). On the other hand, the genetic algorithm-based optimization scheme requires approximately 7.5 hr to obtain the optimal solution under the same conditions. Therefore, the greedy algorithm can determine an optimal gauge configuration with equivalent PSR accuracy to the genetic algorithm in a shorter time.

Appendix D: Verification of Optimization Considering Virtual Observation Noise

To verify the performance of the optimization scheme accounting for the uncertainty caused by observation noise, the optimization and wave height PSR are carried out for synthetic tsunami data with virtual observation noise. We first generate the time series data of tsunami wave heights with noise by combining the original tsunami data simulated by TUNAMI-N2 and the virtual observation noise. Here, we assume that virtual observation noise follows a Gaussian distribution with mean $\mathbf{0}$ and a covariance matrix $\mathbf{\Sigma}$, which has diagonal components consisting of 0.01 m^2 or 1.0 m^2. Finally, the wave height PSR results obtained using these optimal configurations are compared.

Figure D1 shows the optimal 21 gauge locations determined by the greedy algorithm (panel (a)) and the genetic algorithm (panel (b)) considering the virtual observation noise effect in their objective functions. Each plot is color-coded according to the variance in the noise, equivalent to each diagonal component of the covariance matrix $\mathbf{\Sigma}$. This figure shows that the two optimization results are similar: the configurations have 17 gauges in common, and locations with low noise variance are preferentially selected.

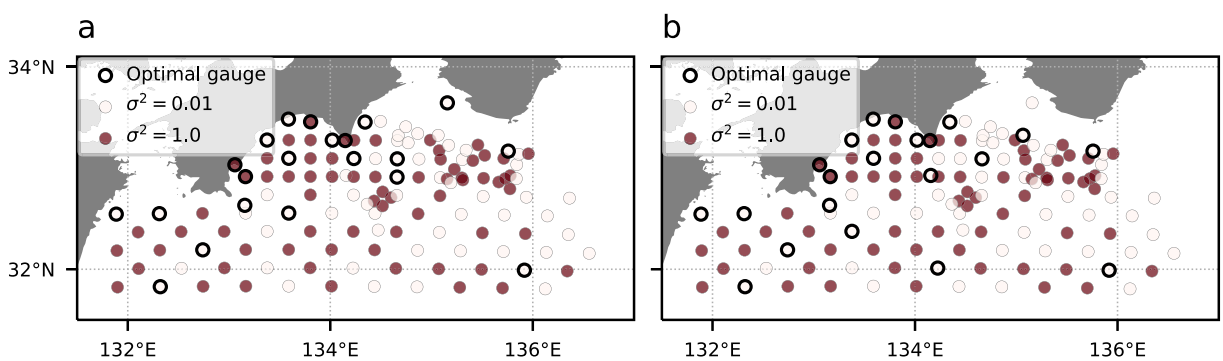

Figure D1. Optimization results considering the covariance matrix of the virtual observation noise in the objective function in Equation 17. (a, b) show the results of the greedy and genetic algorithms, respectively. The black-bordered circles indicate the optimal gauge locations, and each circle is color-coded according to the variance of the virtual observation noise at each gauge.

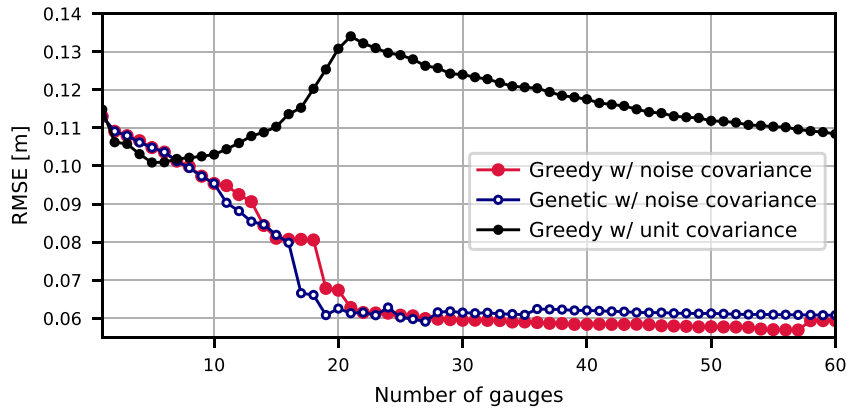

Figure D2. Relationship between the number of optimal gauges and root mean square error between the wave height PSR results using the observations with virtual noise and original simulation results. The red and blue lines show the PSR results using optimal gauges determined by the greedy and genetic algorithms considering the virtual observation noise effect, while the black line shows the result using optimization ignoring the noise effect.

The wave height PSR is conducted by handling the wave data with virtual noise obtained at the optimized gauges as the observation vector y_t in Equation 16. Additionally, the PSR accuracy is evaluated based on the RMSE, as formulated in Equation 24, between the PSR result using the wave data with virtual noise and the original wave heights in the TUNAMI-N2 simulation. Figure D2 shows the relationship between the number of gauges optimized considering the virtual noise covariance matrix and the resulting RMSE. This figure also shows the PSR result using the virtual noise-added wave height data observed at the 21 gauges determined considering a constant uncertainty ($\boldsymbol{\Sigma} = \boldsymbol{I}$), as in Section 4.1. This figure confirms that the RMSEs of the two configurations optimized considering the virtual observation noise effect are equivalent. In addition, these two configurations realize more accurate wave height PSR than that of the optimized configuration neglecting the uncertainty caused by the virtual observation noise. Therefore, the optimal configuration determined by the optimization scheme derived in terms of minimizing the information entropy achieves more accurate PSR results for noisy observation data.

Data Availability Statement

The python codes used for all numerical experiments in this work are available at https://doi.org/10.5281/zenodo.10373316 (Fujita, 2023). All synthetic wave sequences, the locations of the synthetic gauges, and the detail of fault parameters for 1564 tsunami simulations (TUNAMI-N2) are archived in https://doi.org/10.5281/zenodo.8287917 (Koshimura & Fujita, 2023).

- Fujita, S. (2023). Codes for optimization of tsunami gauges configuration (v1.0.0) [Software]. Zenodo. https://doi.org/10.5281/zenodo.10373316
- Koshimura, S., & Fujita, S. (2023). Data from 1564 earthquake/tsunami scenario simulations targeting the Nankai Trough subduction zone (1.0.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.8287917

Acknowledgments
This work was partially supported by JST, the establishment of university fellowships towards the creation of science technology innovation, Grant JPMJFS2102, by JSPS KAKENHI Grant JP23KJ0112, and by the Core Research Cluster of Disaster Science at Tohoku University (Designated National University).

References

Abe, I., & Imamura, F. (2013). Problems and effects of a tsunami inundation forecast system during the 2011 Tohoku earthquake. *Journal of JSCE*, *1*(1), 516–520. https://doi.org/10.2208/journalofjsce.1.1_516

An, C., Liu, P. L., & Meng, L. (2018). A sensitivity analysis of tsunami inversions on the number of stations. *Geophysical Journal International*, *214*(2), 1313–1323. https://doi.org/10.1093/gji/ggy212

Araki, E., Kawaguchi, K., Kaneko, S., & Kaneda, Y. (2008). Design of deep ocean submarine cable observation network for earthquakes and tsunamis. In *Oceans 2008-MTS/IEEE Kobe techno-ocean* (pp. 1–4).

Baba, T., & Cummins, P. R. (2005). Contiguous rupture areas of two Nankai trough earthquakes revealed by high-resolution tsunami waveform inversion. *Geophysical Research Letters*, *32*(8). https://doi.org/10.1029/2004gl022320

Baba, T., Tanioka, Y., Cummins, P. R., & Uhira, K. (2002). The slip distribution of the 1946 Nankai earthquake estimated from tsunami inversion using a new plate model. *Physics of the Earth and Planetary Interiors*, *132*(1–3), 59–73. https://doi.org/10.1016/s0031-9201(02)00044-4

Bird, P. (2003). An updated digital model of plate boundaries. *Geochemistry, Geophysics, Geosystems*, *4*(3). https://doi.org/10.1029/2001gc000252

Commandeur, J. J., & Koopman, S. J. (2007). *An introduction to state space time series analysis*. Oxford University Press.

Ferrolino, A., Mendoza, R., Magdalena, I., & Lope, J. E. (2020). Application of particle swarm optimization in optimal placement of tsunami sensors. *PeerJ Computer Science*, *6*, e333. https://doi.org/10.7717/peerj-cs.333

Fujita, S. (2023). Codes for optimization of tsunami gauges configuration (version 1.0.0) [Software]. Zenodo. https://doi.org/10.5281/zenodo.10373316

Furumura, T., Imai, K., & Maeda, T. (2011). A revised tsunami source model for the 1707 Hoei earthquake and simulation of tsunami inundation of Ryujin Lake, Kyushu, Japan. *Journal of Geophysical Research*, *116*(B2), B02308. https://doi.org/10.1029/2010jb007918

Goto, C., Ogawa, Y., Shuto, N., & Imamura, F. (1997). Numerical method of tsunami simulation with the leap-frog scheme. *IOC Manuals and Guides*, *35*, 130.

Heidarzadeh, M., Wang, Y., Satake, K., & Mulia, I. E. (2019). Potential deployment of offshore bottom pressure gauges and adoption of data assimilation for tsunami warning system in the western Mediterranean Sea. *Geoscience Letters*, *6*(1), 1–12. https://doi.org/10.1186/s40562-019-0149-8

Hirose, F., Nakajima, J., & Hasegawa, A. (2008). Three-dimensional seismic velocity structure and configuration of the Philippine Sea slab in southwestern Japan estimated by double-difference tomography. *Journal of Geophysical Research*, *113*(B9). https://doi.org/10.1029/2007jb005274

Holland, J. H. (1992). Genetic algorithms. *Scientific American*, *267*(1), 66–73. https://doi.org/10.1038/scientificamerican0792-66

Horspool, N., Pranantyo, I., Griffin, J., Latief, H., Natawidjaja, D., Kongko, W., et al. (2014). A probabilistic tsunami hazard assessment for Indonesia. *Natural Hazards and Earth System Sciences*, *14*(11), 3105–3122. https://doi.org/10.5194/nhess-14-3105-2014

Hossen, M. J., Gusman, A., Satake, K., & Cummins, P. R. (2018). An adjoint sensitivity method applied to time reverse imaging of tsunami source for the 2009 Samoa earthquake. *Geophysical Research Letters*, *45*(2), 627–636. https://doi.org/10.1002/2017gl076031

Imamura, F., Yalciner, A. C., & Ozyurt, G. (2006). Tsunami modelling manual. *UNESCO IOC international training course on Tsunami Numerical Modelling*.

Ishibashi, K. (2004). Status of historical seismology in Japan. *Annals of Geophysics*, *47*(2–3). Retrieved from http://hdl.handle.net/2122/763

Kanamori, H. (1972). Tectonic implications of the 1944 Tonankai and the 1946 Nankaido earthquakes. *Physics of the Earth and Planetary Interiors*, *5*, 129–139. https://doi.org/10.1016/0031-9201(72)90082-9

Kanazawa, T. (2013). Japan Trench earthquake and tsunami monitoring network of cable-linked 150 ocean bottom observatories and its impact to earth disaster science. In *2013 IEEE international underwater technology symposium (UT)* (pp. 1–5).

Kaneda, Y. (2010). The advanced ocean floor real time monitoring system for mega thrust earthquakes and tsunamis-application of DONET and DONET2 data to seismological research and disaster mitigation. In *Oceans 2010 MTS/IEEE seattle* (pp. 1–6).

Kato, T., Terada, Y., Ito, K., Hattori, R., Abe, T., Miyake, T., et al. (2005). Tsunami due to the 2004 September 5th off the Kii peninsula earthquake, Japan, recorded by a new GPS buoy. *Earth Planets and Space*, *57*(4), 297–301. https://doi.org/10.1186/bf03352566

Kato, T., Terada, Y., Nagai, T., & Koshimura, S. (2010). Tsunami monitoring system using GPS buoy-present status and outlook. In *2010 IEEE international geoscience and remote sensing symposium* (pp. 3043–3046).

Kato, T., Terada, Y., Tadokoro, K., Kinugasa, N., Futamura, A., Toyoshima, M., et al. (2018). Development of GNSS buoy for a synthetic geohazard monitoring system. *Journal of Disaster Research*, *13*(3), 460–471. https://doi.org/10.20965/jdr.2018.p0460

Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: Past, present, and future. *Multimedia Tools and Applications*, *80*(5), 8091–8126. https://doi.org/10.1007/s11042-020-10139-6

Kerschen, G., Golinval, J.-c., Vakakis, A. F., & Bergman, L. A. (2005). The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview. *Nonlinear Dynamics*, *41*(1), 147–169. https://doi.org/10.1007/s11071-005-2803-2

Koshimura, S., & Fujita, S. (2023). Data from 1564 earthquake/tsunami scenario simulations targeting the Nankai trough subduction zone (version 1.0.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.8287917

Koshimura, S., Moya, L., Mas, E., & Bai, Y. (2020). Tsunami damage detection with remote sensing: A review. *Geosciences*, *10*(5), 177. https://doi.org/10.3390/geosciences10050177

Koshimura, S., & Nomura, R. (2022). Data from 666 earthquake/tsunami scenario simulations targeting Nankai subduction. *Zenodo*. https://doi.org/10.5281/zenodo.6785643

Kotani, T., Tozato, K., Takase, S., Moriguchi, S., Terada, K., Fukutani, Y., et al. (2020). Probabilistic tsunami hazard assessment with simulation-based response surfaces. *Coastal Engineering*, *160*, 103719. https://doi.org/10.1016/j.coastaleng.2020.103719

Liu, C. M., Rim, D., Baraldi, R., & LeVeque, R. J. (2021). Comparison of machine learning approaches for tsunami forecasting from sparse observations. *Pure and Applied Geophysics*, *178*(12), 5129–5153. https://doi.org/10.1007/s00024-021-02841-9

Maeda, T., Obara, K., Shinohara, M., Kanazawa, T., & Uehira, K. (2015). Successive estimation of a tsunami wavefield without earthquake source data: A data assimilation approach toward real-time tsunami forecasting. *Geophysical Research Letters*, *42*(19), 7923–7932. https://doi.org/10.1002/2015gl065588

Makinoshima, F., Oishi, Y., Yamazaki, T., Furumura, T., & Imamura, F. (2021). Early forecasting of tsunami inundation from tsunami and geodetic observation data with convolutional neural networks. *Nature Communications*, *12*(1), 1–10. https://doi.org/10.1038/s41467-021-22348-0

Manohar, K., Brunton, B. W., Kutz, J. N., & Brunton, S. L. (2018). Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems Magazine*, *38*(3), 63–86.

Meza, J., Catalán, P. A., & Tsushima, H. (2020). A multiple-parameter methodology for placement of tsunami sensor networks. *Pure and Applied Geophysics*, *177*(3), 1451–1470. https://doi.org/10.1007/s00024-019-02381-3

Miyazaki, K., Nakajima, J., Suenaga, N., & Yoshioka, S. (2023). Deep subduction of the Philippine Sea slab and formation of slab window beneath central Japan. *Earth Planets and Space*, *75*(1), 1–17. https://doi.org/10.1186/s40623-023-01846-z

Miyazaki, S., & Heki, K. (2001). Crustal velocity field of southwest Japan: Subduction and arc-arc collision. *Journal of Geophysical Research*, *106*(B3), 4305–4326. https://doi.org/10.1029/2000jb900312

Mulia, I. E., Gusman, A. R., & Satake, K. (2017). Optimal design for placements of tsunami observing systems to accurately characterize the inducing earthquake. *Geophysical Research Letters*, *44*(24), 12–106. https://doi.org/10.1002/2017gl075791

Mulia, I. E., Gusman, A. R., Williamson, A. L., & Satake, K. (2019). An optimized array configuration of tsunami observation network off Southern Java, Indonesia. *Journal of Geophysical Research: Solid Earth*, *124*(9), 9622–9637. https://doi.org/10.1029/2019jb017600

NIED. (2023). N-net: Nankai trough seafloor observation network for earthquakes and tsunamis. Retrieved from https://jpgu-agu2020.ipostersessions.com/default.aspx?s=9F-A6-DB-F9-05-97-89-A2-DF-C7-0F-73-71-91-AF-98

Nomura, R., Fujita, S., Galbreath, J. M., Otake, Y., Moriguchi, S., Koshimura, S., et al. (2022). Sequential Bayesian update to detect the most likely tsunami scenario using observational wave sequences. *Journal of Geophysical Research: Oceans*, *127*(10), e2021JC018324. https://doi.org/10.1029/2021jc018324

Okada, Y. (1985). Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America*, *75*(4), 1135–1154. https://doi.org/10.1785/bssa0750041135

Okada, Y. (1992). Internal deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America*, *82*(2), 1018–1040. https://doi.org/10.1785/bssa0820021018

Omira, R., Baptista, M. A., Matias, L., Miranda, J. M., Catita, C., Carrilho, F., & Toto, E. (2009). Design of a sea-level tsunami detection network for the gulf of Cadiz. *Natural Hazards and Earth System Sciences*, *9*(4), 1327–1338. https://doi.org/10.5194/nhess-9-1327-2009

Papalambros, P. Y., & Wilde, D. J. (2000). *Principles of optimal design: Modeling and computation*. Cambridge University Press.

Park, H., Cox, D. T., & Barbosa, A. R. (2018). Probabilistic tsunami hazard assessment (PTHA) for resilience assessment of a coastal community. *Natural Hazards*, *94*(3), 1117–1139. https://doi.org/10.1007/s11069-018-3460-3

Ports and Harbours Bureau. (2022). Real-time NOWPHAS (the nationwide Ocean Wave information network for Ports and HArbourS). Retrieved from https://www.mlit.go.jp/kowan/nowphas/index_eng.html

Saito, T., Satake, K., & Furumura, T. (2010). Tsunami waveform inversion including dispersive waves: The 2004 earthquake off Kii Peninsula, Japan. *Journal of Geophysical Research*, *115*(B6). https://doi.org/10.1029/2009jb006884

Saito, Y., Nonomura, T., Yamada, K., Nakai, K., Nagata, T., Asai, K., et al. (2021). Determinant-based fast greedy sensor selection algorithm. *IEEE Access*, *9*, 68535–68551. https://doi.org/10.1109/access.2021.3076186

Schindelé, F., Loevenbruck, A., & Hébert, H. (2008). Strategy to design the sea-level monitoring networks for small tsunamigenic oceanic basins: The Western Mediterranean case. *Natural Hazards and Earth System Sciences*, *8*(5), 1019–1027. https://doi.org/10.5194/nhess-8-1019-2008

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Tozato, K., Takase, S., Moriguchi, S., Terada, K., Otake, Y., Fukutani, Y., et al. (2022). Rapid tsunami force prediction by mode-decomposition-based surrogate modeling. *Natural Hazards and Earth System Sciences*, *22*(4), 1267–1285. https://doi.org/10.5194/nhess-22-1267-2022

Utsu, T. (2001). *Seismology*. Kyoritsu. (in Japanese).

Wang, Y., Heidarzadeh, M., Satake, K., Mulia, I. E., & Yamada, M. (2020). A tsunami warning system based on offshore bottom pressure gauges and data assimilation for Crete Island in the Eastern Mediterranean Basin. *Journal of Geophysical Research: Solid Earth*, *125*(10), e2020JB020293. https://doi.org/10.1029/2020jb020293

Weiss, J. (2019). A tutorial on the proper orthogonal decomposition. In *AIAA Aviation 2019 forum* (p. 3333).

Welch, G., & Bishop, G. (1995). An introduction to the Kalman filter.

Yoshida, I., Tasaki, Y., Otake, Y., & Wu, S. (2018). Optimal sampling placement in a Gaussian random field based on value of information. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *4*(3), 04018018. https://doi.org/10.1061/ajrua6.0000970

Yoshida, I., Tomizawa, Y., & Otake, Y. (2021). Estimation of trend and random components of conditional random field using Gaussian process regression. *Computers and Geotechnics*, *136*, 104179. https://doi.org/10.1016/j.compgeo.2021.104179

Yoshioka, S., Toda, M., & Nakajima, J. (2008). Regionality of deep low-frequency earthquakes associated with subduction of the Philippine Sea plate along the Nankai trough, southwest Japan. *Earth and Planetary Science Letters*, *272*(1–2), 189–198. https://doi.org/10.1016/j.epsl.2008.04.039