



GENERATING ARTIFICIAL TSUNAMI WAVEFORMS USING STATISTICAL METHOD

Hideharu SUGINO¹ and Yuta ABE²

¹ Member, Dr. Eng., Senior Researcher, Nuclear Regulation Authority,
Tokyo, Japan, sugino_hideharu_r2w@nra.go.jp

² Dr. Sc., Itochu Techno-Solutions Corporation,
Tokyo, Japan, yuta.abe.150@ctc-g.co.jp

ABSTRACT: To assist the advancement of tsunami risk assessment methods for nuclear power plants, requirements for the generation of artificial tsunami waveforms crucial for close linkage between the probabilistic tsunami hazard analysis and tsunami fragility analysis were first listed. Then, basic studies were conducted to model the phase and amplitude spectra of waveforms observed during the 2011 Tohoku earthquake tsunami. Consequently, the average values of the group delay times and amplitude spectrum for each period band were confirmed to be effective for each modeling. Finally, based on these requirements and results of the basic studies, a technique for generating artificial tsunami waveforms using a statistical method was proposed.

Keywords: *Artificial tsunami waveform, Wavelet transform, Group delay time, Probabilistic tsunami hazard analysis, Tsunami fragility analysis, Dissimilarity between time history waveforms*

1. INTRODUCTION

Implementation Standards concerning the tsunami probabilistic risk assessment (PRA) of nuclear power plants¹⁾ (referred to as the implementation standard) published by the Atomic Energy Society of Japan is known as a method for the quantitative evaluation of tsunami risk of nuclear power plants. Tsunami PRA evaluates the frequencies of core damage and loss of containment vessel function by combining three analyses: (1) a probabilistic tsunami hazard analysis (PTHA) to evaluate the relationship between the intensity of tsunamis striking the facility (tsunami water level, etc.) and its exceedance frequency, (2) a tsunami fragility analysis to evaluate the probability of loss-of function of SSCs (structures, systems, and components), and (3) an accident sequence analysis to evaluate the probability of core damage based on the safety function of the plant systems and quantify the probability as the comprehensive tsunami risk of nuclear power plant.

Remarkably, the three above analysis methods constituting the tsunami PRA methods are interrelated. Since tsunami fragility analysis requires various information related to function loss, such as inundation depth, flow velocity, hydrodynamic force, etc., at each installation location of SSCs against arbitrary levels of tsunami hazards, having some waveform time histories as input conditions on sea surfaces simulating tsunamis at a point off the coast (called the “artificial tsunami waveform”) are necessary when obtaining similar information by tsunami run-up analysis, indicating that these artificial

tsunami waveforms be closely related to tsunami hazard information.

To this end, Sugino et al.²⁾ proposed an analytical method for obtaining tsunami fragility curves, combining the mean/standard deviation of inundation depths at each installation location of SSCs, the mean/standard deviation of the capacity values related to the function loss of SSCs, and the inundation depths calculated by the tsunami run-up analysis using multiple artificial tsunami waveforms. In this method, sine waves were used as the artificial tsunami waveforms, causing their amplitudes and wavelengths to be varied while accounting for waveform variabilities owing to the tsunami source diversities. However, the sine waves could not represent the complexity of real tsunami waveforms because they oversimplified them, making the rationale for varying this waveform unclear.

Conversely, Iwabuchi et al.³⁾ proposed a method to generate artificial tsunami waveforms by analyzing and classifying frequency characteristics on analytical waveforms from many scenario tsunamis. Here, the analytical waveforms for scenario tsunamis were classified into three categories by highest water level and further into three categories by a ratio of the long-period component to the highest water level, followed by a separation of the analytical waveforms with similar shapes from each other. Afterward, average waveforms were calculated for each of the nine categories and multiplied by arbitrary factor to obtain artificial tsunami waveforms. This method also attempted to incorporate regional characteristics related to tsunami generation, such as bathymetries and earthquake types in the sea area around a target site. However, since the average waveforms were multiplied by an arbitrary factor, this step did not sufficiently consider the diversity of tsunami sources.

Since the artificial tsunami waveforms generated by both methods of Sugino et al.²⁾ and Iwabuchi et al.³⁾ are input waveforms to tsunami run-up analysis for the sea area around a target site, they have been treated as forced water level fluctuation inputs at analysis domain boundaries. Hence, inputting such forced water level fluctuations is a method described in the implementation standard¹⁾.

However, in addition to the above methods, the implementation standard¹⁾ shows a method for generating artificial tsunami waveforms that can be handled continuously in tsunami propagation and run-up analysis without dividing the sea area from the tsunami source by the vicinity of a target site. In this method, the contribution of a scenario tsunami is first evaluated at each earthquake occurrence region by re-decomposing the tsunami hazard curve. Then, representative scenario tsunamis are selected based on this information, followed by acquiring artificial tsunami waveforms by adjusting the initial profile of tsunamis so that the maximum water level of the representative scenario tsunami at the hazard evaluation point matches the target water level. Still, in this method, the crustal deformation associated with fault movement is set to be the same as that of the base representative scenario tsunami, even when the initial profile is adjusted, causing the relationship between the crustal deformation and the initial profile of a tsunami to be unnatural and the rational explanations to be insufficient.

Due to these limitations, Kihara et al.⁴⁾ proposed another method, which was similar to the implementation standard¹⁾, for generating artificial tsunami waveforms based on hazard re-decomposition, where the adjustment target differs from that of the implementation standard¹⁾ when the maximum water level at the hazard assessment point for representative scenario tsunamis is matched with that of the target water level. As a result, while the implementation standard¹⁾ adjusted the initial profile of the representative scenario tsunami, the method by Kihara et al.⁴⁾ adjusted the fault slip of the representative scenario tsunami. Afterward, crustal deformation associated with fault movement was calculated from the adjusted fault slip once the correlation between the tsunami water level at the hazard assessment point and the uplift/subsidence of the target site had been confirmed.

The implementation standard¹⁾ and Kihara et al.⁴⁾ methods are effective in practice because necessary parameters, such as inundation depth in the onshore run-up area, could be obtained by adjusting the initial profile or the fault slip of the representative scenario tsunami. However, since this adjustment method is almost the same as simply multiplying the amplitude by a factor so that the highest water level of the representative scenario tsunami becomes the target water level, the phase characteristics remain the same as the waveforms of the representative scenario tsunamis themselves. Therefore, these results indicate no clear rationale for accepting the waveform obtained in this way as the target water level waveform.

Alternatively, a method was employed in generating artificial earthquake ground motions by calculating stationary waveforms using a design amplitude spectrum⁵⁾ and a phase spectrum given by

uniform random numbers, and multiplying the stationary waveforms by amplitude envelopes representing the earthquake ground motion characteristics in the seismic design of buildings and structures of nuclear facilities. In particular, using this method, Sato et al.⁶⁾ focused on the fact that the mean value and standard deviation of the group delay time obtained from the phase spectrum corresponds to the location of the centroid and spread (duration) of the seismic motion in a time domain. Hence, they proposed a modeling method for the phase spectrum using the wavelet transform⁷⁾ based on observed seismic waves. Consequently, this method was used in the seismic design of railroad structures⁸⁾. By subsequently adopting the seismic PRA method⁹⁾, a continuous evaluation system from probabilistic seismic hazard analysis to fragility analysis was established, combining this phase spectrum model with the uniform hazard spectrum obtained from probabilistic seismic hazard analysis.

With tsunami PRA, developing a method for generating artificial tsunami waveforms that can represent variabilities in these waveforms caused by tsunami source diversities with reasonable explanatory power while considering the regional characteristics of tsunami generation around a target site instead of the above methods of using sine waves or adjusting the initial profile/fault slip of the representative scenario tsunamis, is necessary. It is also desirable that the evaluation system from PTHA to tsunami fragility analysis is continuous and that the entire tsunami PRA method has close/organic linkage. Here, the regional characteristics related to tsunami generation represent the unique bathymetry that affects the tsunami waveforms and earthquake types (such as interplate or intraplate earthquakes, etc.), inducing tsunamis in the sea area surrounding the facility. The tsunami source diversities indicate not only different earthquake types but also different seismic moments, fault locations, fault areas, and slip distributions, even if the earthquake types are similar. Organic is a term that describes a state in which many parts are assembled to constitute a whole, like an organic body¹⁰⁾. In such cases, each part is closely linked and influences the other. Therefore, while this paper considers the tsunami PRA method as a whole, PTHA, tsunami fragility analysis, and accident sequence analysis were considered as each part. Then, we used the term “organic linkage” to refer to a state in which these parameters are more closely linked.

In this paper, we first identify the artificial tsunami waveform requirements for tsunami fragility analysis. Next, we conduct basic studies for modeling the phase and amplitude spectra of waveforms observed during the 2011 off the Pacific coast of the Tohoku Earthquake and Tsunami (subsequently referred to as the Tohoku Earthquake Tsunami). In particular, our study refers to the method of Sato et al.⁶⁾, which has a proven track record in generating artificial earthquake time histories, to create a phase spectrum model. Then, to closely/organically link PTHA and tsunami fragility analysis, we extract statistical properties from several analytical waveforms of scenario tsunamis obtained by PTHA, thereby proposing a method to generate artificial tsunami waveforms using the Monte Carlo method (subsequently referred to as the MC method). Finally, we present examples of the generation of such waveforms.

2. ARTIFICIAL TSUNAMI WAVEFORM REQUIREMENTS FOR TSUNAMI FRAGILITY ANALYSIS

We define the artificial tsunami waveform requirements to improve the reliability of tsunami fragility analysis as follows:

Requirement 1: The artificial tsunami waveform should be defined at the PTHA target point. The horizontal axis of the tsunami fragility curve for SSCs scattered around the facility site should be represented by a similar index to that for the tsunami hazard curve to provide uniform information associated with tsunami hazard information to an accident sequence analysis in the tsunami PRA method. Consequently, the artificial tsunami waveform should be defined at the PTHA target point; its maximum water level should correspond to that obtained from the tsunami hazard curve.

Requirement 2: The artificial tsunami waveform should be able to define any tsunami water level in the tsunami hazard curve. To account for the tsunami propagation model uncertainty in PTHA, the tsunami hazard curves are calculated assuming lognormal distributions, with maximum water levels of

the scenario tsunamis being the median values. Consequently, the horizontal axis range of the tsunami hazard curve exceeds the maximum water level of the scenario tsunamis, causing the artificial tsunami waveforms within the range of the maximum water level of the scenario tsunamis to be insufficient because the tsunami fragility analysis also need to consider the effects of the above uncertainties in the PTHA. Therefore, the maximum water level of artificial tsunami waveform should be able to define any water level exceeding the maximum water level of the scenario tsunamis.

Requirement 3: The variabilities in phase and amplitude owing to regional characteristics and tsunami source diversities should be statistically and rationally considered on the artificial tsunami waveforms. Even given that the maximum water levels are similar at an offshore point away from the coast where the facility is located, if the waveform shapes are different, differences should exist in the inundation depth at the installation site of the SSC, the inundation area at the site, and the hydrodynamic force on the SSC. Consequently, the artificial tsunami waveforms would need to be represented by multiple waveforms using the statistical and rational explanatory method so that variabilities in phase and amplitude characteristics due to the tsunami source diversity and the regional characteristic related to tsunami generation in the sea area around the facility can be considered.

In PTHA, multiple scenario tsunamis are first defined, after which information such as the maximum water levels from these tsunami height time histories are used. This procedure is to extract one representative parameter from the valuable information contained in the tsunami height time histories and to lose the phase and amplitude characteristic information due to regional characteristics related to tsunami generation and source diversities. To this end, it is crucial to recover information on these phase and amplitude characteristics lost in PTHA and reflect them in artificial tsunami waveforms.

Requirement 4: The artificial tsunami waveform should reflect the effects of crustal uplift or subsidence associated with fault movement. Elevation changes associated with uplift or subsidence of the crust due to fault movement have been directly reflected in the topographic model of a typical tsunami run-up analysis using a scenario tsunami. Even given that the two scenario tsunamis have similar maximum water levels at the offshore point, if the amount of subsidence at the sites differs, the tsunami with the more significant subsidence has a greater impact on the facility. Therefore, these effects need to be appropriately considered in the artificial tsunami waveforms to improve the reliability of tsunami fragility analysis.

In light of the above requirements, the aforementioned methods for generating artificial tsunami waveforms can be organized as follows:

First, all four methods¹⁾⁻⁴⁾ satisfy requirements 1 and 2. However, while the method of Sugino et al.²⁾ does not satisfy requirement 3 because it cannot represent the complexity of real tsunami waveforms, the method of Iwabuchi et al.³⁾ does not also satisfy requirement 3 because this method obtains the waveform with the target water level based on a multiplication of the average waveform by an arbitrary factor, making the rationale unclear and the shape of the artificial tsunami waveform limited. Interestingly, however, the method by Iwabuchi et al. indirectly satisfies requirement 4, using the analytical waveforms obtained from PTHA to make time-history waveforms of the relative water level by subtracting the uplift/subsidence (positive on the uplift side) at the target site from the amplitude of the analytical waveforms. Although the implementation standard¹⁾ and Kihara et al.⁴⁾ methods have also been considered to satisfy requirement 3 by extracting multiple representative scenario tsunamis with high contributions, they do not fully satisfy this requirement because, as mentioned above, no clear rationale exists that the tsunami waveform at the target water level can be created by multiplying the amplitude of the representative scenario tsunami by an arbitrary factor or adjusting the initial profile/fault slip with similar effects. Additionally, while the relationship between the adjusted initial profile and the crustal movement in the implementation standard¹⁾ is unnatural, even when the crustal movement calculated using the adjusted fault slip of Kihara et al.⁴⁾ is used, there is no clear rationale for allowing the fault slip to be adjusted, similar to the previous reason. Therefore, it is unlikely that either method provides a sufficiently reasonable explanation for requirement 4.

Based on the facts above, this paper propose a new method to solve the pending problems. We believe that satisfying the requirements will be a concrete measure to realizing the close/organic linkage between PTHA and tsunami fragility analysis. To this end, we notably follow the method of Iwabuchi et al.³⁾ to address requirement 4, after which we subject the analytical waveforms expressed based on

relative water levels to statistical analysis of phase and amplitude spectra. Additionally, we have defined the artificial tsunami waveform style as the forced water level fluctuation, which is an input waveform for the tsunami propagation and run-up analysis.

3. BASIC STUDY ON PHASE AND AMPLITUDE SPECTRA MODELING

3.1 The basic study's flow

We study the modeling of phase and amplitude spectra using observed waveforms to obtain fundamental data for constructing a method to generate artificial tsunami waveforms. This study's flow is described below.

Section 3.2 describes the procedure from the modeling of phase and amplitude spectra to the reconstruction of waveforms.

Section 3.3 estimates the phase/amplitude spectra models and reconstruction waveforms according to the procedure in Section 3.2, followed by an examination of the aforementioned method to determine whether Sato et al.'s method⁶⁾ can be applied to tsunami waveforms or whether there is a more appropriate method. First, in the study of phase spectra models, waveforms are reconstructed using various phase spectra models and an amplitude spectrum of the observed waveforms. Next, we compare the reconstructions with the observed waveforms. Then, we examine a method for creating an amplitude spectrum model that is harmonized with the method for creating a phase spectrum model, after which the waveforms reconstructed using various amplitude spectra models and a phase spectrum of the observed waveforms, are compared with the observed waveforms as before. Finally, the waveform differences are qualitatively shown by superimposing both the reconstruction and observed waveforms in the comparison.

In Section 3.4, furthermore, we perform a quantitative evaluation of the dissimilarity between the observed and reconstructed waveforms. Then, in contrast to the previous section's qualitative comparison results, we attempt a quantitative evaluation by applying the Jensen–Shannon divergence¹¹⁾.

Section 3.5 finally examines the impact of uncertainties on the initial integral value of a phase spectrum model. Although we obtain the phase spectrum model by setting the initial integral value and integrating the group delay time model, this value involve uncertainty. Hence, we first confirm the effect of the initial integral value's uncertainty and examine how to select appropriate initial integral values for generating artificial tsunami waveforms since these values are expected to affect the shape of the reconstructed waveform. Afterward, we create reconstructed waveforms using the phase and amplitude spectra models with various initial integral values, followed by an organization of the uncertainty effects in the initial integral values using a quantitative evaluation method for waveform dissimilarities.

3.2 Procedure from the creation of phase/amplitude spectra models to waveform reconstruction

As described above, this study first use an observed tsunami waveform to design a procedure for creating phase/amplitude spectra models. Then, we reconstruct a waveform using these models. Our procedure comprise four parts, each of which is described below.

3.2.1 Preprocessing of the observed tsunami waveform

After the Tohoku Earthquake Tsunami, we first conduct preprocessing to obtain necessary observation point information (mainly phase, amplitude, duration time, etc.), including waveform data¹²⁾ on GPS buoys installed 10–20 km offshore from the Pacific coast of the Tohoku region, by adopting the following procedures:

- 1) Set the time t_a (s) when the maximum water level appears, then shift the time to attain a similar maximum water level on the observed waveform to t_a .
- 2) Set the duration time T (s) so that the above t_a is included, then extract the part of the observed

waveform for T .

- 3) Conduct taper processing at the beginning and end of the extracted waveform. First, set a certain period T_t (s) at that time to conduct sine wave taper at the beginning so that the amplitude gradually increases from zero between $t = 0$ and T_t . Then, conduct cosine wave taper at the end so that the amplitude gradually decreases to zero between $t = T - T_t$ and T .
- 4) Add zero data to the observed waveform after T and finally revise the observed waveform with a total data number of $2^{j_{max}}$.

3.2.2 Creation of the phase spectrum model

Sato et al.⁶⁾ used wavelet transformation⁷⁾ to extract phase information from nonstationary waveforms, that is the observed seismic motion, using the scaling and shifting of function $\varphi(t)$, also called the mother wavelet. With $\varphi(t)$, wavelet transformation could extract analogous waveforms to the mother wavelet at various scales within the time-history data $x(t)$ without losing time information. Since the observed tsunami waveform is similar to a nonstationary waveform, we refer to this method while improving its parts. Based on these facts, this section first discusses a general outline of Sato et al.'s method⁶⁾, followed by an explanation of the improvements.

From the method by Sato et al.⁶⁾, the dispersion wavelet transformation and inverse transformation of time-history data $x(t)$ could be expressed as in Eqs. (1) and (2).

$$w_m^{(j)} = \int_{-\infty}^{\infty} x(t) \varphi_{j,m}^*(t) dt \quad (1)$$

$$x(t) = \sum_j g^{(j)}(t) = \sum_j \sum_m w_m^{(j)} \varphi_{j,m}(t) \quad (2)$$

where $\varphi(t)$ represents the mother wavelet, * represents the complex conjugate, suffix j represents the scale factor, suffix m represents the time position, $w_m^{(j)}$ represents the wavelet coefficient at j indicating the roughness (resolution) of the signal. $\varphi_{j,m}(t)$ is a wavelet of a different resolution. By adopting the WaveLab version.850 of Buckheit et al.¹³⁾ as the wavelet transformation and inverse transformation tool, the time-history data $g^{(j)}(t)$ at j could further be expressed in Eq. (3), called the wavelet component.

$$g^{(j)}(t) = \sum_m w_m^{(j)} \varphi_{j,m}(t) \quad (3)$$

Each component waveform was then Fourier transformed to obtain the phase spectrum $\phi^{(j)}(\omega)$, after which $\phi^{(j)}(\omega)$ is differentiated with the angular frequency ω to obtain the group delay time $t_{gr}^{(j)}(\omega)$, and the unwrap method of Sawada et al. was applied to calculate the average $\mu^{(j)}$ and standard deviation $\sigma^{(j)}$ of group delay time at the support section of $f^{(j)}$.

Subsequently, Sato et al.⁶⁾ applied Meyer¹⁵⁾ as the mother wavelet $\varphi(t)$. In such a case, $\varphi(t)$ becomes the compact support in the frequency domain. Hence, in this support section, they expressed $f^{(j)}$ (1/s) by Eq. (4).

$$\frac{2^j}{3T_d} \leq f^{(j)} \leq \frac{2^{j+2}}{3T_d} \quad (4)$$

where T_d represents the total time of the time-history data.

Since two values were observed in the support section $f^{(j)}$ expressed by Eq. (4) because of the overlap at the neighboring j of the phase spectrum and group delay time, Sato et al.⁶⁾ defined a support section $f^{(j)}$ expressed by Eq. (5) such that the phase and amplitude spectrum were paired.

Consequently, it was possible to calculate the $\mu^{(j)}$ and $\sigma^{(j)}$ of the group delay times using the data within this support section.

$$\frac{2^{j-1}}{T_d} < f^{(j)} \leq \frac{2^j}{T_d} \quad (5)$$

Next, they applied similar procedures to several observed seismic motions to obtain a regression model equation based on the data groups $\mu^{(j)}$ and $\sigma^{(j)}$. Then, the group delay time model $t'_{gr}{}^{(j)}(\omega)$ was calculated using the MC method, thereby obtaining the phase spectrum model $\phi'^{(j)}(\omega)$ by integrating this group delay time model $t'_{gr}{}^{(j)}(\omega)$ with ω . The above result is from the method proposed by Sato et al.⁶⁾

Since this study partially improve Sato et al.'s method⁶⁾, the improvements must be explained. First, we discuss the improvements in the unwrap method: Initially, Sawada et al.¹⁴⁾ obtained $t_{gr}{}^{(j)}(\omega)$, produced its histogram, and conducted a $\pm T$ time shift to the data outside the region $[(T_{mod}{}^{(j)} - T/2) \sim (T_{mod}{}^{(j)} + T/2)]$ centered around the mode $T_{mod}{}^{(j)}$ for the data to fit within this region. Investigations reveal that this time shift is equivalent to a $\pm 2\pi T/T_d$ (rad) shift, causing any shift other than $\pm 2\pi$ to overwrite the phase information originally contained in the tsunami waveform. In the following, we improve the unwrap method to prevent the original tsunami waveform's phase information from being rewritten by the unwrap method.

First, the mode $T_{mod}{}^{(j)}$ is changed to the median $T_{mid}{}^{(j)}$. The reason for changing from $T_{mod}{}^{(j)}$ to $T_{mid}{}^{(j)}$ is to avoid setting the mean at the end of the distribution because the number of data is small, within the range of about 2 to 4, causing the histogram to be biased. Then, for the data to fit within the region $[(T_{mid}{}^{(j)} - T_d/2) \sim (T_{mid}{}^{(j)} + T_d/2)]$ centered around the median $T_{mid}{}^{(j)}$, the shift conducted on the data outside this region to make them fit within also have to be changed to $\pm T_d$. Based on these modifications, the $\pm T_d$ shift become equivalent to the $\pm 2\pi$ (rad) shift, and the phase information of the observed tsunami waveform is saved as it is.

Finally, the group delay time model $t'_{gr}{}^{(j)}(\omega)$ by Sato et al.'s method⁶⁾ used the regression model equation obtained based on the data groups $\mu^{(j)}$ and $\sigma^{(j)}$ to calculate the group delay time from several observed seismic motions. Based on the relationship between the number k_{max} of divisions and the reproducibility of the reconstruction waveform, we define the $f^{(j,k)}$, as shown in Eq. (6), by dividing the $f^{(j)}$ expressed by Eq. (5) into multiple domains to consider an expression method for it with as little k_{max} as possible.

$$\left\{ 2^{j-1} + \frac{2^{j-1}}{k_{max}} \cdot (k-1) \right\} \cdot \frac{1}{T_d} < f^{(j,k)} \leq \left\{ 2^{j-1} + \frac{2^{j-1}}{k_{max}} \cdot k \right\} \cdot \frac{1}{T_d} \quad (6)$$

where k_{max} is the division number and k is a natural number from 1 to k_{max} . Note that when $k_{max} = 1$, Eq. (6) is equal to Eq. (5).

3.2.3 Creation of the amplitude spectrum model

Since no existing research on the amplitude spectrum model of the tsunami waveform is known, we propose an amplitude spectrum model that matches the aforementioned method of the phase spectrum model. First, the observed tsunami waveform is Fourier transformed to obtain the amplitude spectrum $A(\omega)$. Then, while $f^{(j,k)}$ is set similar to the group delay time model discussed above, the $\mu^{(j,k)}$ of each part is calculated using the amplitude spectrum and set as the amplitude spectrum model $A'^{(j,k)}(\omega)$. Although the method described here is relatively simple, the fact that a reconstructed waveform that uses such an amplitude spectrum model can reproduce originally observed tsunami waveforms to a certain extent will be discussed later in this paper.

3.2.4 Waveform reconstruction using the phase/amplitude spectra models

We finally conduct waveform reconstruction using the phase spectrum model $\phi^{(j)}(\omega)$ and the amplitude spectrum model $A^{(j)}(\omega)$ obtained through the methods discussed above in 3.2.2 and 3.2.3, respectively. Here, the phase spectrum model $\phi^{(j)}(\omega)$ and the amplitude spectrum model $A^{(j)}(\omega)$ with a similar scale factor j are used to conduct Fourier inverse transformation to obtain the component waveform $g^{(j)}(t)$ after modeling. Afterward, $g^{(j)}(t)$ in the wavelet inverse transformation of Eq. (2) is replaced with $g^{(j)}(t)$; the reconstructed waveform $x'(t)$ is obtained using Eq. (7).

$$x'(t) = \sum_j g^{(j)}(t) \quad (7)$$

3.3 Trial calculations for the phase/amplitude spectral models and the reconstructed waveforms

3.3.1 Observed tsunami waveforms and their preprocessing

Among the waveforms¹²⁾ observed by GPS buoys during the Tohoku Earthquake Tsunami, this paper use G802 (off the southern coast of Iwate). Figure 1 shows the location (depth 204 m¹²⁾ where the GPS buoy was installed. The observed waveform is divided into two types: The first half of the wave (subsequently referred to as the direct wave) that reached the observation point directly from the tsunami source, and the second half (subsequently referred to as the superimposed wave) that passed through the observation point, reached the coast and was reflected, reached the observation point again, and was superimposed on the direct wave. By considering the artificial tsunami waveform characteristics as input conditions for tsunami propagation and run-up analysis, targeting only the direct waves is appropriate. However, a strict distinction is not made here because the purpose is to conduct a basic study on modeling phase and amplitude spectra. Therefore, we set the parameters such as t_{max} mentioned in Section 3.2.1 under the condition that the direct and superimposed waves are mixed. Referring to the description of Sato et al.⁶⁾ that the data length is sufficiently long, and since the observed waveform G802 with a time interval of 5 s show a maximum water level rise approximately 25 min after the earthquake occurred, we also set the maximum water level generation time t_a to 1500 s, the duration

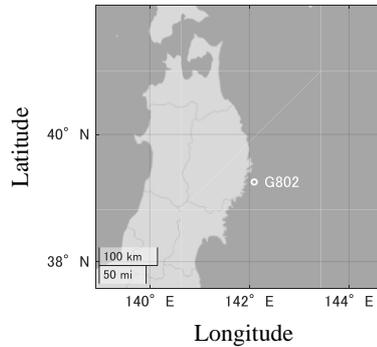


Fig. 1 Installation location of GPS buoy (G802, off the southern coast of Iwate)

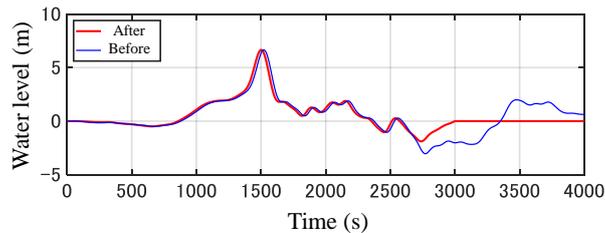


Fig. 2 Water level waveform observed at G802 (processing before and after)

time T is set to 3000 s (twice the time t_a), the taper processing time T_t is set to 600 s, and the j_{max} of the total number of data $2^{j_{max}}$ after adding zero data is set to 15.

A solid red line in Fig. 2 shows the observed waveform after processing. For comparison, the figure shows a superimposition of the observed waveforms before processing on a solid blue line. We confirm from this schematic that the maximum water level generation time shifted to 1500 s, causing the start and end points of the waveform to be gradually increased and decreased by tapering processing. Subsequently, zero data is added after 3000 s, causing the total time to be 163840 ($= 5 \times 2^{15}$) s because the time interval is 5 s.

3.3.2 Waveform reconstruction with the phase spectrum model

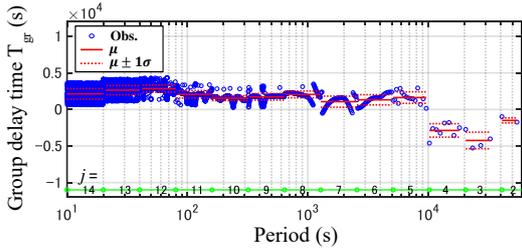
Here, before we consider creating a phase spectrum model, we first confirm the effectiveness of the improved unwrap method. To this end, the group delay times are obtained by combining the methods of Sawada et al.¹⁴⁾ and this paper. Then, we reconstruct waveforms using the amplitude spectrum obtained from the observed waveform, followed by a comparison with the originally observed waveform. Figure 3(a) shows the group delay time $t_{gr}^{(j)}$ when the unwrap method of Sawada et al.¹⁴⁾ is employed to the observed waveform in Fig. 2. Note in the figure that $t_{gr}^{(j)}$ is marked with a blue circle, the scale factor j and the support interval $f^{(j)}$ are indicated by green circles and solid lines, and the $\mu^{(j)}$ and $\sigma^{(j)}$ for each $f^{(j)}$ are indicated by solid red lines and dotted red lines, respectively. By adopting this unwrap method, investigations confirm that $t_{gr}^{(j)}$, which is scattered in $[-T_d \sim T_d]$, move to the region $[(T_{mod}^{(j)} - T/2) \sim (T_{mod}^{(j)} + T/2)]$ for each $f^{(j)}$.

Subsequently, we perform the shift of $\pm 2\pi T/T_d$ (rad) using this unwrap method, as described above, causing the phase information held originally by the observed waveform to be rewritten. Then, to confirm its effectiveness, we try to reconstruct the waveform using $t_{gr}^{(j)}$ of Fig. 3(a) and $A^{(j)}$ of the observed waveform. Figure 3(b) shows the reconstructed (red line) and observed (blue line) superimposed waveforms. Investigations reveal that although the maximum water level generation times coincide between the reconstructed and observed waveforms, the waveform shapes do not, proving that the phase information is indeed rewritten. We also observe that the difference between $\sigma^{(j)}$ for each j calculated based on $t_{gr}^{(j)}$ after the unwrap is small.

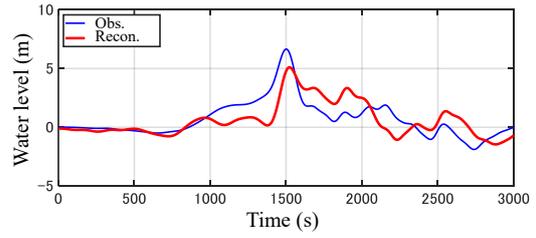
Next, Fig. 4(a) shows $t_{gr}^{(j)}$ when the improved unwrap method of this study is applied. The legend in the figure is the same as in Fig. 3(a). The result show that $\sigma^{(j)}$ greatly differ from that shown in Fig. 3(a), indicating a clear difference in j . Additionally, the distribution of $t_{gr}^{(j)}$ after the improved unwrap show that the distribution is not random but smooth and continuous for each $f^{(j)}$. Figure 4(b) shows the reconstructed waveform using $t_{gr}^{(j)}$ as before. Investigations reveal that although the reconstructed waveform almost match the observed waveform, this match is imperfect. The reason for this imperfect match is that the phase information is not that of the originally observed waveform after changing $f^{(j)}$ from Eq. (4) to Eq. (5). However, the extent of its impact is small. Thus, as a precaution, the reconstructed waveform using $t_{gr}^{(j)}$ before the improved unwrap is compared with the reconstructed waveform in Fig. 4(b), which makes them utterly consistent, confirming that the improved unwrap has no effect.

The above results show that it is practical to calculate $\mu^{(j)}$ and $\sigma^{(j)}$ from $t_{gr}^{(j)}$ based on the improved unwrap method, thereby obtaining intrinsic phase information on the observed waveform. Therefore, based on these results, we further consider how to create a phase spectrum model.

First, as in Sato et al.⁶⁾, we assume a normal distribution using $\mu^{(j)}$ and $\sigma^{(j)}$ of the group delay times obtained for each $f^{(j)}$. Next, while the reconstructed waveform is calculated with the amplitude spectrum of the observed waveform after the group delay time model has been obtained by random sampling, $\mu^{(j)}$ and $\sigma^{(j)}$ are obtained from the group delay time after applying the improved unwrap method shown in Fig. 4(a). Then, normal random numbers are generated using MATLAB's function "randn¹⁶⁾" and the Mersenne Twister method¹⁷⁾. While Fig. 5(a) shows the group delay time model,

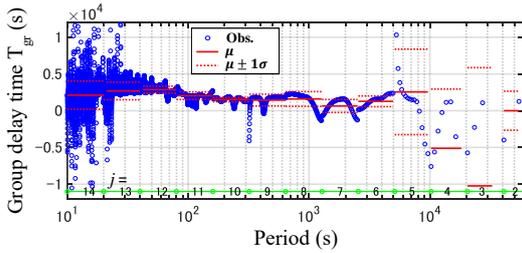


(a) Group delay time

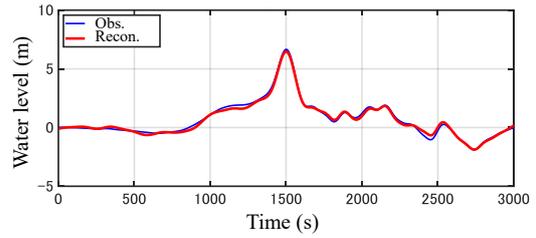


(b) Reconstructed and observed waveform

Fig. 3 Group delay time employed Sawada et al.'s unwrap¹⁴⁾ method and reconstructed waveform

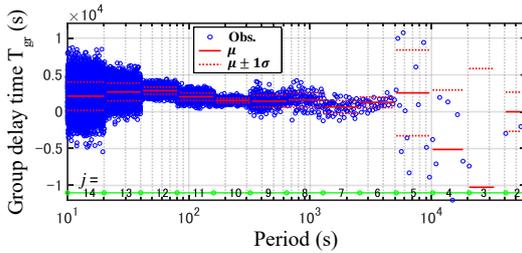


(a) Group delay time

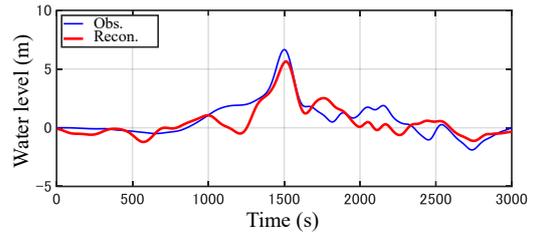


(b) Reconstructed and observed waveform

Fig. 4 Group delay time employed the improved unwrap method and reconstructed waveform

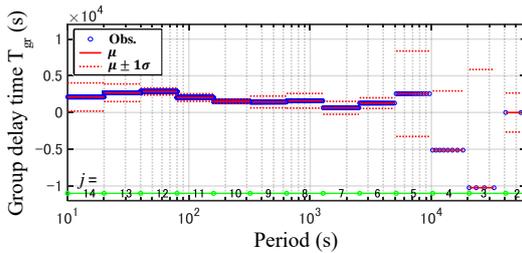


(a) Group delay time model

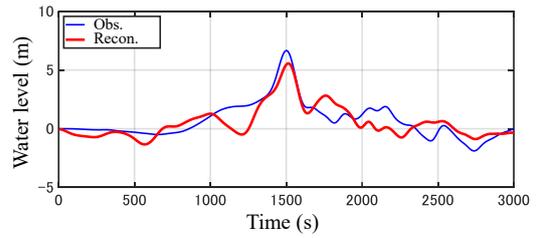


(b) Reconstructed and observed waveform

Fig. 5 Group delay time model using the normal random numbers and reconstructed waveform

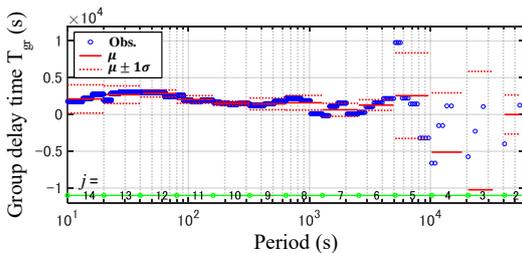


(a) Group delay time model

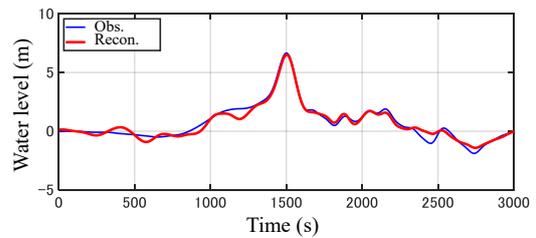


(b) Reconstructed and observed waveform

Fig. 6 Group delay time model using mean values and reconstructed waveform



(a) Group delay time model



(b) Reconstructed and observed waveform

Fig. 7 Group delay time model with four-division average and reconstructed waveform

Figure 5(b) shows the reconstructed waveform. We also examine the case where $\sigma^{(j)}$ is zero to confirm the effect of $\sigma^{(j)}$ on the group delay time model. The group delay time model and reconstructed waveform results are shown in Figs. 6(a) and 6(b), respectively. While the blue circles shown in Figs. 5(a) and 6(a) represent the group delay time model, the solid red and dashed green lines represent $\mu^{(j)}$ and $\sigma^{(j)}$ like in Fig. 4(a).

Subsequently, we compare the reconstructed waveforms, as shown in Figs. 5(b) and 6(b). Investigations show almost no difference in appearance. Although the reason for such a result has not been clarified, we observe that at least the effect of incorporating $\sigma^{(j)}$ into the creation of group delay time models is small. Additionally, comparing these reconstructed waveforms with the originally observed waveform, it is confirmed that these waveforms drop significantly around 1200 s, resulting in almost opposite phases between 1600 s and 2500 s. This finding indicates that the reproducibility is not good. Overall, this model is inappropriate since it does not sufficiently extract phase information from the observed waveform.

Figure 6 also shows poor reproducibility by just using the mean value in the support interval $f^{(j)}$. Conversely, Fig. 5 shows the inability of the observed waveform to be reproduced even if $t'_{gr}{}^{(j)}$ is simply randomly varied. Nevertheless, as seen in Fig. 4(a), $t'_{gr}{}^{(j)}$ is significantly and continuously distributed in the support interval $f^{(j)}$, making it necessary to simulate this distribution when reproducing the observed waveform. Therefore, as a method of simulating a continuous $t_{gr}{}^{(j)}$ distribution, we first adopt a simple method to obtain the mean value by dividing the support interval $f^{(j)}$ into several intervals though modeling it with some functions is possible. Then, using $f^{(j,k)}$ in Eq. (6), we confirm how the reproducibility of the reconstructed waveform improves when the resolution of the mean value is increased by dividing it so that the number of data is similar in the support interval $f^{(j)}$. Figure 7(a) shows the group delay time model when the division number is four. In the figure, we also compare similar $\mu^{(j)}$ and $\sigma^{(j)}$ as in Fig. 4(a). Consequently, the group delay time model approximates the group delay time of the observed waveform in Fig. 4(a) by increasing the division number from the long-period side where j is small. Similarly, Fig. 7(b) shows the reconstructed waveform when the division number is four. In this reconstructed waveform, compared with that in Fig. 6(b), the reproducibility improves, causing the drop at approximately 1200 s to be reduced and the opposite phase after 1600 s to be almost eliminated.

From the above results, it suggests that increasing the resolution of the mean values is more effective than using the group delay time standard deviations to create a phase spectrum model. Moreover, mean values of about four divisions should be used to reproduce the originally observed waveform.

3.3.3 Waveform reconstruction using the amplitude spectrum model

Subsequently, we reconstruct waveforms using various amplitude spectra models and the phase spectrum obtained from the observed waveform to design a method for creating amplitude spectra models. Then, we compare the reconstructed waveform with an initially observed one to confirm the effects of the amplitude spectrum model.

Figures 8 and 9 show the amplitude spectra models and reconstructed waveforms as the division number of $f^{(j,k)}$ is gradually increased in the phase spectrum model described above. Specifically, while Fig. 8(a) shows the amplitude spectrum model of one-division, Fig. 9(a) shows that of four-division. These figures show the amplitude spectrum model by red circles and the amplitude spectrum obtained from the observed waveform by blue circles. Subsequently, the amplitude spectrum model approximates the amplitude spectrum of the observed waveform by increasing the division number at the long-period side, where j is small. Conversely, Figs. 8(b) and 9(b) show the reconstructed waveforms using these amplitude spectra models by solid red lines and the observed superimposed waveforms by solid blue lines. Even with the one-division average shown in Fig. 8(b), the shape of the observed waveform is well reproduced. Investigations also reveal that the four-division average shown in Fig. 9(b) almost match the observed waveform, indicating that although reproducibility is improved by increasing the division number, the effect is small.

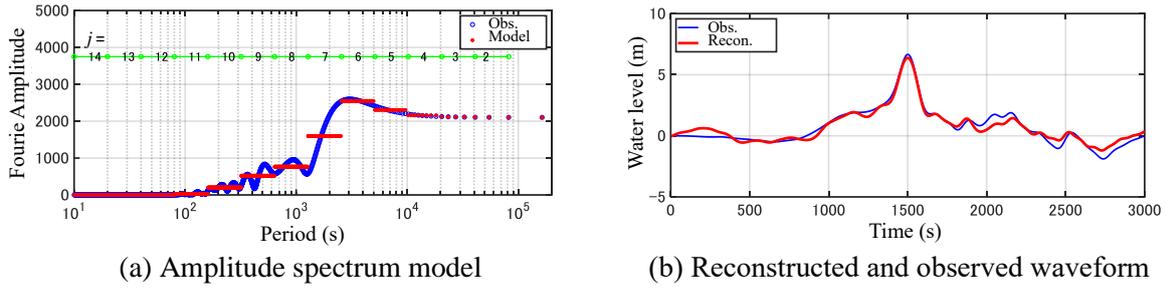


Fig. 8 Amplitude spectrum model with one-division average and reconstructed waveform

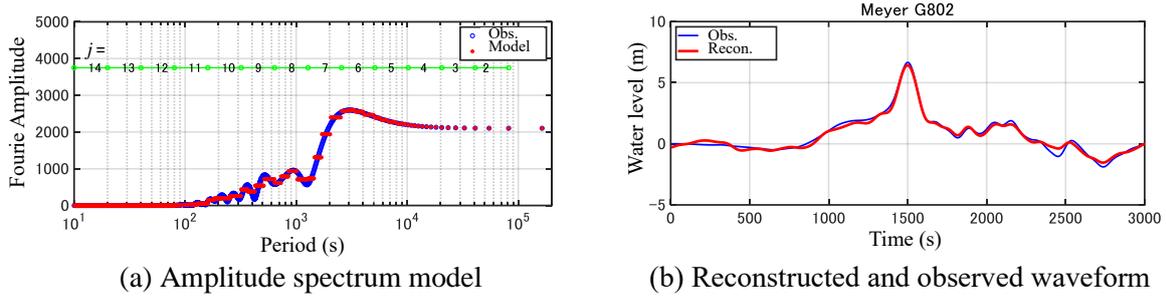


Fig. 9 Amplitude spectrum model with four-division average and reconstructed waveform

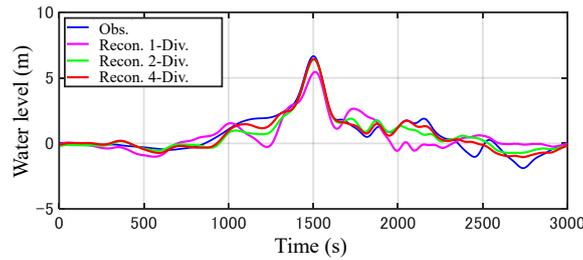


Fig. 10 Reconstructed waveforms for the phase/amplitude spectra models with different division number

The above results show a difference in the sensitivity of the averaging modeling method between the phase spectrum model and the amplitude spectrum model. Therefore, it is considered that the phase spectrum model dominated waveform shape relatively more than the amplitude spectrum model.

3.3.4 Waveform reconstruction using the phase/amplitude spectra models

In Sections 3.3.2 and 3.3.3, we model only one kind of spectra, either phase or amplitude spectra, to create a reconstructed waveform. In this section, we construct the average models for both kinds of spectra, set the division number of $f^{(j,k)}$ using the above methods, reconstruct the corresponding waveforms, and confirm the reproducibility for the observed waveform.

Figure 10 shows the reconstructed waveforms for the phase and amplitude spectra models. A solid pink line here shows the one-division average, a solid green line shows the two-division average, and a solid red line shows the four-division average. We also superimpose the figure's originally observed waveform on a solid blue line. Investigations reveal that the reproducibility of the reconstructed waveform for the observed waveform improve as the division number increase. This result is almost consistent with the reconstructed waveforms shown in Figs. 6(b) to 7(b) when the phase spectrum model was examined. As a result, the phase spectrum model is more dominant than the amplitude spectrum model when assess from the viewpoint of influence on reproducibility.

3.4 Quantitative dissimilarity evaluation between the observed and reconstructed waveforms

Section 3.3 confirm the effects of the phase and amplitude spectrum models on reconstructed waveforms through the superimposition of observed and reconstructed waveforms. Afterward, their dissimilarities are qualitatively discussed. However, in Section 3.4, we attempt to quantitatively evaluate the effects of phase and amplitude spectra models in more detail by applying the Jensen–Shannon (JS) divergence¹¹⁾ (which has been used in machine learning), focusing on the dissimilarity between the phase/amplitude spectra of the observed waveform and the phase/amplitude spectra model, to create a reconstructed waveform.

3.4.1 Method of quantitative dissimilarity evaluation between two waveforms using JS divergence

JS divergence¹¹⁾ is originally an index that measures the dissimilarity between two probability-based distributions whose value is greater than or equal to 0. Notably, while its value is 0 if two probability distributions are exactly consistent, its value is greater if these distributions differ. Assuming two probability distributions $p(x)$ and $q(x)$, the JS divergence (D_{JS}) can be calculated using Eqs. (8)–(10).

$$D_{JS}(p, q) = \frac{1}{2}D_{KL}(p, M) + \frac{1}{2}D_{KL}(q, M) \quad (8)$$

$$M(x) = \frac{1}{2}p(x) + \frac{1}{2}q(x) \quad (9)$$

$$D_{KL}(p, q) = \sum p(x) \log \frac{p(x)}{q(x)}, \quad (10)$$

where $D_{KL}(p, q)$ denotes the Kullback–Leibler (KL) divergence between the probability distributions $p(x)$ and $q(x)$.

In this paper, assuming that the above $p(x)$ and $q(x)$ are the phase spectrum (p_P) of the observed waveform and the phase spectrum model (q_P) of the reconstructed waveform or the amplitude spectrum (p_A) of the observed waveform and the amplitude spectrum model (q_A) of the reconstructed waveform, the dissimilarity of the two waveforms (the observed and reconstructed waveforms) can be calculated using the index D_{JS_W} , as defined by Eq. (11).

$$D_{JS_W} = \sqrt{D_{JS_P}^2 + D_{JS_A}^2} \quad (11)$$

$$D_{JS_P} = D_{JS}(p_P, q_P), D_{JS_A} = D_{JS}(p_A, q_A), \quad (12)$$

where D_{JS_P} represents the JS divergence related to the phase spectrum and D_{JS_A} represents the JS divergence related to the amplitude spectrum, respectively.

Afterward, while p_P and q_P can be defined by Eq. (13), p_A and q_A can be defined by Eq. (14).

$$p_P(k) = \frac{\phi_k}{2^{j_{max}-1}}, q_P(k) = \frac{\phi'_k}{2^{j_{max}-1}}, (k = 1, \dots, 2^{j_{max}-1}) \quad (13)$$

$$p_A(k) = \frac{X_k}{|x|_{max}} \cdot \frac{1}{2^{j_{max}-1}}, q_A(k) = \frac{X'_k}{|x|_{max}} \cdot \frac{1}{2^{j_{max}-1}}, (k = 1, \dots, 2^{j_{max}-1}), \quad (14)$$

where ϕ_k and X_k represent the phase/amplitude spectra obtained by Fourier transforming the observed waveform, ϕ'_k and X'_k represent the phase/amplitude spectra models obtained by Fourier transforming the reconstructed waveform, and $|x|_{max}$ represents the absolute value of the maximum

amplitude of the observed waveform, where X_k and X'_k are divided by $|x|_{max}$ to normalize the amplitudes of the observed and reconstructed waveforms, respectively.

Since ϕ_k and ϕ'_k have $[-\pi \sim \pi]$ as their principal values, they result in an uncertainty of 2π . Therefore, both p_P and q_P must be positive, just as the probability distributions p and q are positive, to apply JS divergence. Hence, we add 2π to ϕ_k and ϕ'_k , which are less than or equal to 0, to make them fall within the range of $0 < (\phi_k, \phi'_k) \leq 2\pi$.

Finally, we separately examine the systematic changes in phase shift or amplitude for a set of sine waves having similar periods to confirm the basic properties and applicability of D_{JS_W} in Eq. (11). From our investigations, D_{JS_W} quantitatively show dissimilarities in the two different waveforms while preserving the basic properties of the JS divergence.

3.4.2 Results of quantitative dissimilarity evaluation

Next, we attempt to quantitatively evaluate the dissimilarity between the observed waveform and various reconstructed waveforms, using the dissimilarity index D_{JS_W} of Eq. (11), to confirm the effects of modeling the phase/amplitude spectra discussed in Section 3.3. For each reconstructed waveforms in Figs. 4(b) to 9(b) and 10, D_{JS_P} , D_{JS_A} , and D_{JS_W} are calculated for the observed waveform. The maximum number of j used for these calculations is 12, Table 1 shows a list of the calculation results, and these results are shown in Fig. 11.

From the top (Nos. 1–5) of Table 1, we show a case where the group delay time is almost the same as the observed waveform in Fig. 4, another where the random number generation considers the standard deviation of the group delay time (Fig. 5), one where only the average value of the group delay time is reflected (Fig. 6), and the case of two-division or four-division averages (shown in Fig. 7) for the group delay time in which the division number in the support interval is increased. Each phase spectrum is called “quasi-observation,” “random number model,” “one-division model,” “two-division model” and “four-division model,” respectively. In the case of these reconstructed waveforms, since the amplitude spectrum of the observed waveform is used, D_{JS_A} is zero, and $D_{JS_W} = D_{JS_P}$. The D_{JS_W} of these reconstructed waveforms are, in descending order, 0.30 for the one-division model, 0.28 (difference 0.02) for the random number model, 0.23 (0.05) for the two-division model, 0.15 (0.08) for the four-division model, and 0.07 (0.08) for quasi-observation. These rankings agree with the ones in which the reconstructed waveforms are improved in Section 3.3.2. Additionally, the value of the differences between the models described above is consistent with the qualitative improvement in reproducibility.

Following the above, the middle row (Nos. 6–8) of Table 1 shows the results of the amplitude spectra models, reflecting a case in Fig. 8 (one-division model) indicating the average value of the observed waveform’s amplitude spectrum and the case of a two-division average or a four-division average (Fig. 9) with an increased division number in the support interval (two- or four-division model). Investigations reveal that the D_{JS_W} is 0.09 for the one-division model, 0.08 (difference 0.01) for the two-division model, and 0.07 (0.01) for the four-division model, agreeing with the qualitative reproducibility in Section 3.3.3. Regarding the modeling method based on the divided average of the phase/amplitude spectra models, we observe that the D_{JS_P} is 0.30–0.15 in the case of the phase spectrum model, whereas the D_{JS_A} is 0.07–0.03 in the case of the amplitude spectrum model. These values represent smaller influence of the modeling method in the amplitude spectrum model. Overall, the results quantitatively show that the phase spectrum model relatively dominates the waveform shape more than the amplitude spectrum model in Section 3.3.3.

Finally, the lower part (Nos. 9–11) of Table 1 shows the results of the reconstructed waveform by both the phase/amplitude spectrum models. Investigations reveal that the D_{JS_W} is 0.30 for the one-division model, 0.24 for the two-division model, and 0.15 for the four-division model, which is almost similar to each D_{JS_P} . From these results, it can be confirmed that the influence of the phase spectrum model is greater than that of the amplitude spectrum model on the dissimilarity of the reconstructed waveform. It is also quantitatively shown that the shapes of the reconstructed waveforms shown in Fig. 10 are almost the same as in Figs. 6(b)–7(b).

Table 1 List of the calculation results of dissimilarity between observed waveform (G802) and reconstructed waveforms

No.	Combinations of phase/amplitude spectra in waveform reconstruction		$D_{JS,P}$ (Y axis*)	$D_{JS,A}$ (X axis*)	$D_{JS,W}$ (Polar coordinate*)	Reference	
	Phase	Amplitude				Figures	Sections
1	Quasi-Obs.	Obs.	0.07	0.00	0.07	Fig. 4(b)	Section 3.3.2
2	Random Model		0.28		0.28	Fig. 5(b)	
3	1-division Model		0.30		0.30	Fig. 6(b)	
4	2-division Model		0.23		-		
5	4-division Model		0.15		0.15	Fig. 7(b)	
6	Quasi-Obs.	1-division Model	0.07	0.07	0.09	Fig. 8(b)	Section 3.3.3
7		2-division Model		0.04	0.08	-	
8		4-division Model		0.03	0.07	Fig. 9(b)	
9	1-division Model	1-division Model	0.30	0.07	0.30	Fig. 10	Section 3.3.4
10	2-division Model	2-division Model	0.23	0.04	0.24		
11	4-division Model	4-division Model	0.15	0.03	0.15		

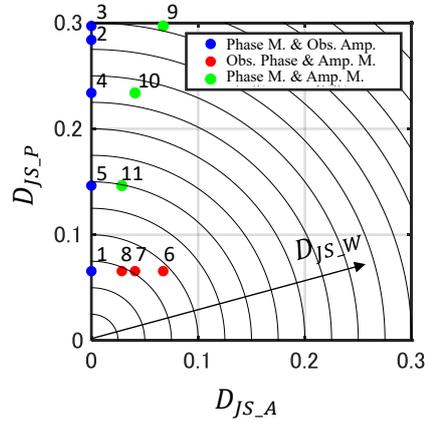


Fig. 11 Dissimilarity index $D_{JS,W}$ between observed and reconstructed waveforms

Subsequently, the calculation results shown in Table 1 are plotted on a single graph in Fig. 11, where the Cartesian and polar coordinate systems are displayed superimposed, with $D_{JS,A}$ on the X axis, $D_{JS,P}$ on the Y axis, and $D_{JS,W}$ plotted as the dynamic radius (distance from the origin) of the polar coordinate system based on Eq. (11). In the figure, blue circles represent the phase spectrum model, red circles represent the amplitude spectrum model, and green circles represent the results using both the phase/amplitude spectra models, respectively. The number in each circle corresponds to the No. in Table 1. This figure helps to visually understand the dissimilarities between each of the above waveforms. Considering the qualitative results of the waveform comparison in Figs. 4(b) to 9(b) and 10, including the quantitative results using the dissimilarity index $D_{JS,W}$ in Fig. 11, one can find that the two waveforms almost coincide when $D_{JS,W}$ is less than 0.1. However, such thresholds should be set individually according to the purpose of use. Since this paper aims to create phase/amplitude spectra models, it is necessary to show how much simplification is allowed. Consequently, by identifying $D_{JS,W}$ as a threshold value of 0.15 greater than 0.1, we adopt the “four-division model” for the phase/amplitude spectra models.

3.5 Uncertainty of the initial integral value via the phase spectrum model

Although the reconstructed waveform in Section 3.3 use the phase spectra models obtained by integrating various group delay time models in the angular frequency domain, the phase spectrum values of the observed waveform are adopted as the initial integral values. Since these values represent the interval boundary determined for each support interval $f^{(j)}$, two values of $\pm 0.5\pi$ are observed when the observed waveform is examined. However, it is uncertain which value will be obtained. Hence, in the generation of an artificial tsunami waveform described later, reconstructed waveforms with many initial integral values that combine two values for each $f^{(j)}$ should be created to investigate their effects since it is impossible to specify the initial integral value in advance. This also indicates the need for a method to extract appropriate ones from many combinations. This section investigates the influence of initial integral value uncertainty and the extraction method of combining initial integral values to create a phase spectrum model from a group delay time model.

3.5.1 Methodology

When creating a phase spectrum model from the group delay time model in Section 3.3, one value is given as the initial integral value for each support interval $f^{(j)}$. However, since the needed scale factor j ranges from 2 to 14, 13 values are needed. Consequently, during investigations of the observed waveform, results are either 0 or π when j is 2, and they are either 0.5π or -0.5π when j is 3 to 14. Therefore, the total number of initial integral value combinations is 2^{13} .

Subsequently, we create reconstructed waveforms from these combinations, estimate the dissimilarity index D_{JS_W} for the observed waveform and a water level H_{1500} for the reconstructed waveform at 1500 s, which is the maximum water level generation time on the observed waveform, followed by an organization of these relationships. Then, to create reconstructed waveforms for the observed waveform G802, both the phase/amplitude spectra models are comprised of “four-division models” based on the results of Sections 3.3 and 3.4, and the maximum number of j used to calculate D_{JS_W} is 12 as in Section 3.4.

3.5.2 Results

Figure 12 shows a D_{JS_W} and H_{1500} scatterplot for the reconstructed waveforms of 2^{13} initial integral value combinations. With the horizontal axis representing D_{JS_W} and the vertical axis representing H_{1500} , each reconstructed waveform is represented by one blue dot mark. In the figure, any case in which the phase spectrum value of the observed waveform used as the initial integral value is indicated by a red circle, a red square shows any case in which the value deviated π from the integral initial value of the observed waveform, and H_{1500} of the observed waveform is indicated by a red dashed line. Investigations reveal that while the red circle indicating the phase spectrum value of the observed waveform as the initial integral value show the smallest D_{JS_W} at 0.15, H_{1500} is 6.44 m, which is close to the water level in the observed waveform (6.66 m). Furthermore, although the D_{JS_W} differ, we observe cases where H_{1500} is close to the water level of the observed waveform. These reconstructed waveforms are numbered 1–4 in a descending D_{JS_W} order (Fig. 13).

Investigations also reveal that although the reconstructed waveform 1 is the case in which the initial integral value is an observed waveform, and the reconstructed waveform 2 is almost the same as No. 1, only a slight difference is observed around 2000 s. However, the reconstructed waveforms 3 and 4 almost coincide with the reconstructed waveforms 1 and 2 because the value of the amplitude spectrum model of the support interval (j is 12) shown in Fig. 9(a) is very small. These results suggest that even if there is a difference in the phase spectrum model of this support interval, it will be difficult for this difference to appear in the waveform’s amplitude. Incidentally, the case indicated by a red square in Fig. 12 shows the exact opposite of the reconstructed waveform 1 and corresponds to the case of an upside-down inverted waveform, using the initial integral value of the π deviation.

From these results, we focus on H_{1500} of the reconstructed waveform as the method to extract the reconstructed waveform close to the observed waveform from many candidates due to the uncertainty

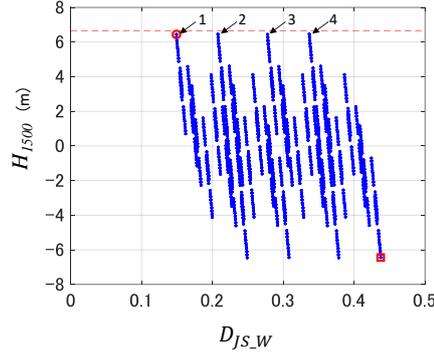


Fig. 12 D_{JS_W} and H_{1500} scatterplot for the reconstructed waveforms of 2^{13} initial integral value combinations

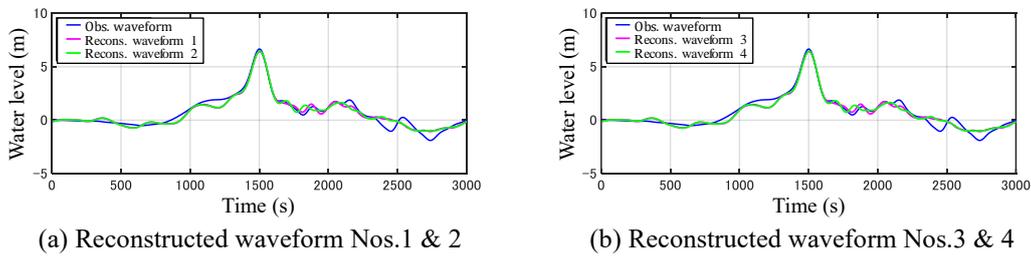


Fig. 13 The reconstructed waveforms considering the uncertainty of initial integral values of the phase spectra models in which H_{1500} are close to that of the observed waveform

of the initial integral value and obtain the prospect of extracting targets using H_{1500} of the observed waveform as the target water level. Hence, the applicability of this extraction method is largely due to the fact that the maximum water level generation time t_a is fixed in advance.

4. GENERATING ARTIFICIAL TSUNAMI WAVEFORMS USING STATISTICAL METHODS

Based on the artificial tsunami waveform requirements in Chapter 2 and the results from basic studies on modeling phase/amplitude spectra using the observed waveform in Chapter 3, this chapter proposes a method for generating artificial tsunami waveforms that can closely/organically link PTHA and tsunami fragility analysis. Then, we present an example of the artificial tsunami waveforms generated using this method, thereby showing its usefulness.

4.1 Procedure for generating the phase/amplitude spectra models and artificial tsunami waveforms using a group of analytical waveforms

By defining many scenario tsunamis to consider the uncertainty of tsunami occurrence, many analytical waveforms have been obtained at the PTHA target point. As listed in Requirement 3 of Chapter 2, the analytical waveforms include local/regional characteristics, such as the seafloor topography, the earthquake type around the target area, and the diversity of tsunami sources. Therefore, based on the basic study results in Chapter 3, we extract the information needed for modeling the phase/amplitude spectra from the analytical waveforms to create models using these statistics. Then, we propose a method for generating an artificial tsunami waveform using statistical methods.

Figure 14 shows a flowchart of the procedure for generating phase/amplitude spectra models and artificial tsunami waveforms using analytical waveforms. Since (1) Preprocessing of analytical waveforms and (4) Waveform reconstruction using phase/amplitude spectra models are similar in Chapter 3, we omit their explanations here. Thus, only (2) Generation of phase spectra models and (3) Generation of amplitude spectra models are described in detail below.

First, we describe (2): The method for generating phase spectra models using analytical waveforms. In the basic study of Chapter 3, while the observed waveform is only targeted, the phase spectra models are generated here using the analytical waveforms from multiple scenario tsunamis. Then, as shown in the first half of Fig. 14, we conduct wavelet transformation on each analytical waveform, followed by the mean value $\mu_{tgr}^{(j,k)}$ calculation of the group delay time for each division range $f^{(j,k)}$. Since the procedure up to this point is similar to Chapter 3, it is conducted for all analytical waveforms. In the

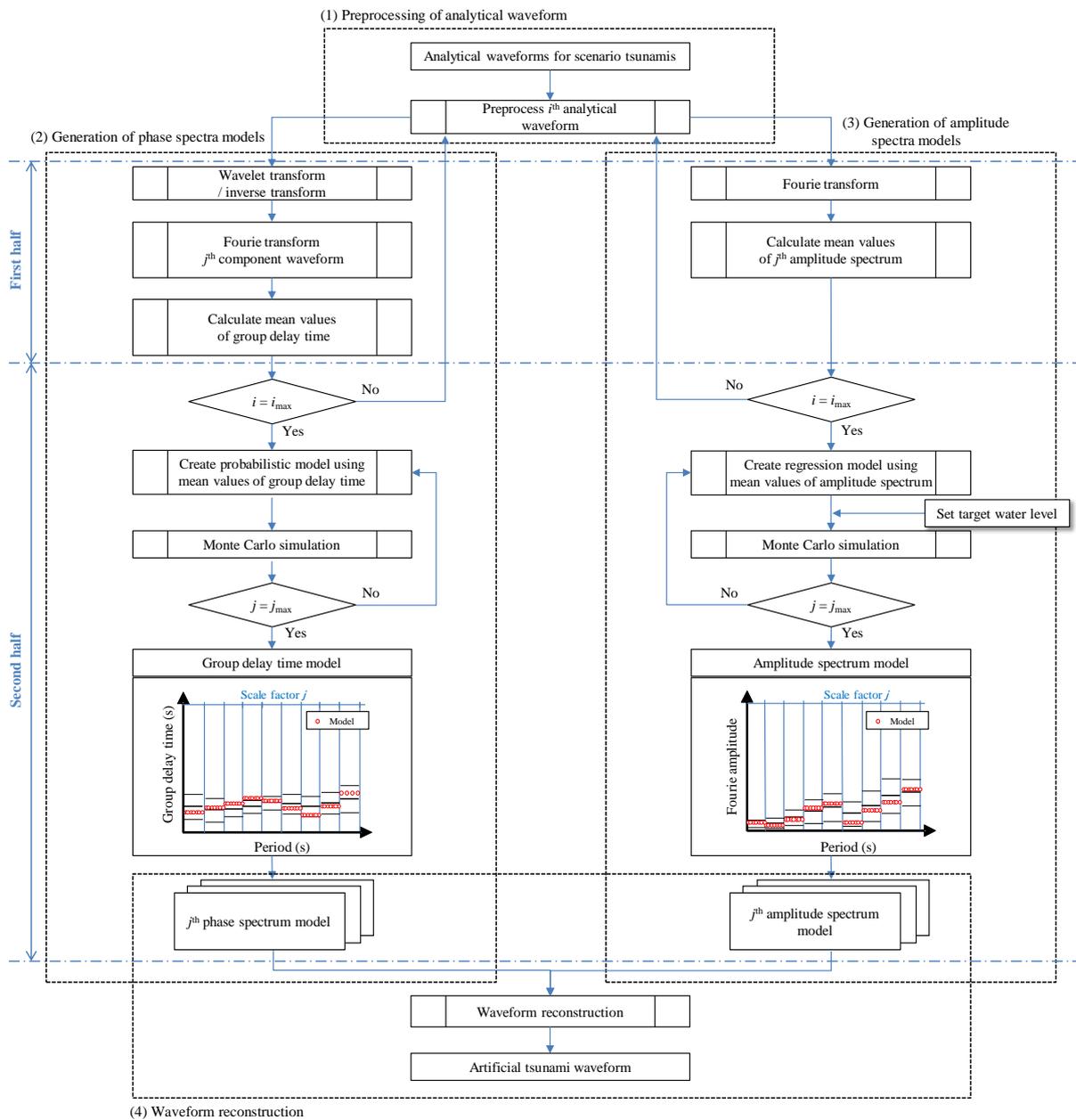


Fig. 14 The procedure for generating phase/amplitude spectra models and artificial tsunami waveforms

second half, however, we calculate the mean $\tilde{\mu}^{(j,k)}_{tgr}$ and standard deviation $\sigma^{(j,k)}_{tgr}$ based on the data group of the mean values for these group delay times, followed by a creation of the group delay time model $t'_{gr}{}^{(j,k)}$ using the MC method, assuming normal distributions based on these statistics. Finally, the group delay time model $t'_{gr}{}^{(j,k)}$ is integrated by ω into the phase spectrum model $\phi^{(j)}(\omega)$, thereby generating phase spectra models for all initial integral value combinations based on the uncertainty results of the initial integral value in Section 3.5.

Subsequently, (3): The method for generating amplitude spectra models using analytical waveforms is described. The procedure for generating them is also divided into first and second halves, as shown in Fig. 14, where the first half is the same procedure as the basic study in Chapter 3, and the second half is the newly added procedure. In the first half, we Fourier transform each analytical waveform to obtain the amplitude spectrum $A(\omega)$. Then, we calculate the mean value $\mu^{(j,k)}_{amp}$ of $A^{(j,k)}$ for each $f^{(j,k)}$ as in the phase spectrum model. The procedure up to this step is conducted for all analytical waveforms. Conversely, in the second half, an average amplitude spectrum model $\mu'^{(j,k)}_{amp}$ is created using the data group of $\mu^{(j,k)}_{amp}$ for each analytical waveform. Based on the correlation between $\mu'^{(j,k)}_{amp}$ and H_{1500} in the analytical waveform, we subsequently model it using the regression formula in Eq. (15), employing H_{1500} as an explanatory parameter to perform regression analysis by the least squares method and determine the coefficients $\alpha^{(j,k)}$ and $\beta^{(j,k)}$. Details are described in the next section with examples. The symbol “ $\ln(\cdot)$ ” in Eq. (15) represents the natural logarithm.

$$\ln\left(\mu'^{(j,k)}_{amp}\right) = \alpha^{(j,k)} + \beta^{(j,k)} \times \ln(H_{1500}) \quad (15)$$

Equation (15) expresses the average trend of a data group with variability. There are residuals between this equation and the individual source data. As these residuals are a modeling uncertainty that must be considered at the generating stage of the amplitude spectrum model described later, we calculate their standard deviation $\sigma'^{(j,k)}_{amp}$.

Next, we set the target water level H_t of the artificial tsunami waveform. Then, sampling is performed employing the MC method to generate the amplitude spectrum model $A'^{(j,k)}(\omega)$, assuming a lognormal distribution using $\mu'^{(j,k)}_{amp}$ and $\sigma'^{(j,k)}_{amp}$ obtained by substituting H_t into H_{1500} in Eq. (15).

After generating the phase/amplitude spectra models in the previous procedure, we perform waveform reconstruction, using the same procedure as Chapter 3 to obtain the artificial tsunami waveform. Due to the uncertainty of the initial integral value, the number of phase spectra models is $2^{j_{max}-2}$ at this stage, resulting in similar number of candidates for the artificial tsunami waveforms. Therefore, based on the results in Section 3.5, we extract one from many candidates for the artificial tsunami waveforms whose water level at the preset maximum water level occurrence time t_a is close to the target water level.

Since the above procedure adopt analytical waveforms for the scenario tsunamis with regional characteristics, these regional characteristics are reflected in the phase/amplitude spectra models. Additionally, by adopting a probabilistic model and the MC method for the phase/amplitude spectra models, statistically setting reasonable and explanatory variabilities becomes possible. As a result, multiple artificial tsunami waveforms with different shapes can be generated even if the target water level is the same. By adopting a regression equation, using H_{1500} as an explanatory parameter in Eq. (15) for the generation of the amplitude spectrum model, and extrapolating the range of the highest water level of the analytical waveforms for the scenario tsunamis, setting the target water level beyond this range while ensuring rational explanatory also becomes possible.

Subsequently, since the only information obtained from the hazard curve is the maximum water level, which is the output of PTHA, we further conduct regression analyses using only H_{1500} as an explanatory parameter in Eq. (15). Then, we make the most effective use of the horizontal axis index of the hazard curve to ensure continuity between PTHA and tsunami fragility analysis, thereby enhancing

the practicality of the proposed method for generating an artificial tsunami waveform.

Tentatively, as in the method of Sato et al.⁶⁾, it is possible to construct a regression equation, using the magnitude and epicenter distance of the earthquake as explanatory parameters. However, since these settings are required as input values when generating an artificial tsunami waveform in such cases, difficulties in setting these values with statistically rational explanations are expected, indicating the need for further examinations.

4.2 Estimation of artificial tsunami waveforms

4.2.1 Group of analytical waveforms for scenario tsunamis

This paper has proposed a method to generate artificial tsunami waveforms for close/organic linkages between PTHA and tsunami fragility analysis. Therefore, this section applies the proposed method and estimates artificial tsunami waveforms using the results of previous PTHA studies.

An example of PTHA is our previous study¹⁸⁾, where a location of 150 m depth off the coast of Fukushima Daiichi Nuclear Power Station was defined as the hazard evaluation point. Since the assumed tsunami sources were earthquake-induced tsunamis, the characterizing model¹⁹⁾ for earthquake-induced tsunamis was set considering various earthquake magnitudes and earthquake types along the Japan Trench, followed by many analytical waveform calculations. Subsequently, “Relative maximum water level rise” was used as a horizontal axis tsunami hazard curve index to consider the effects of crustal uplift and subsidence according to fault movement at the power plant's location. Then, we calculated the relative maximum water level rise by subtracting crustal deformation (with a positive uplift side) at the power plant's location from the maximum water level rise at the hazard assessment point in each scenario tsunami. This approach is an important viewpoint for linking PTHA and tsunami fragility analysis via artificial tsunami waveforms. For details on the analysis conditions of PTHA, please refer to the original paper.

A similar index called “Relative maximum water level rise” is also applied to the analytical waveforms to generate artificial tsunami waveforms in this paper. Here, the analytical waveforms of many scenario tsunamis at hazard assessment points are treated as relative values (relative water levels expressed as the waveform) by subtracting the crustal deformation (with a positive uplift side) associated with each fault movement from the water level of the analytical waveform. However, since this approach is related to the fact that the artificial tsunami waveform format proposed in this paper is the input waveform for the tsunami propagation and run-up analysis, these analyses can not directly consider crustal deformation according to fault movement. Nevertheless, we propose that the influence of crustal deformation can still be indirectly considered by generating artificial tsunami waveforms based on analytical waveforms expressed by relative values.

Figure 15 shows the analytical waveforms comprising 118 scenario tsunamis selected from the interplate earthquakes assumed in the PTHA¹⁸⁾. The waveforms in the figure are displayed by classifying the relative maximum water levels in 6 m increments. As in the preprocess of the observed waveform in 3.3.1, the relative maximum water level occurrence time t_a is set as 1500 s, the duration time T as 3000 s, the taper processing time T_t as 600 s, j_{max} of the total number of data $2^{j_{max}}$, which is applied zero data, as 15, and the data interval in each waveform as one second. Then, each waveform has the characteristic with a two-step rise¹²⁾ (water level rise), as seen in the waveform observed by the GPS buoy during the Tohoku Earthquake Tsunami, and we can see them the two-step rise variabilities in water level and waveform.

Conversely, Table 2 shows a list of selected scenario tsunami sources. In the table, the No. of scenario tsunami sources, the moment magnitude M_w , and the placement pattern symbol for slip distribution are shown. All these items are referred to as the values and symbols from the previous study¹⁸⁾. The symbols for the placement slip distribution pattern are shown in the schematic of the tsunami source in Fig. 16. The symbols “_A, _B, and _C” represent the placement patterns of the large slip area in the tsunami source, indicating the northern, central, and southern, respectively, whereas the symbols “1,..., 5” represent the placement patterns in which the placement, shape, and number of super-

Table 2 Selected scenario tsunami sources

Source No.	Mw	Placement pattern of slip distribution				
404	8.3	404				
409	8.7	409				
410	8.7	410				
413	8.9	413 C1				
414	8.9	414 A1				
		414 B1				
		414 C1				
415	8.8	415				
416	9.1			416 C3	416 C4	416 C5
417	9.0	417 B1				
		417 C1				
418	9.0	418 A1				
		418 B1				
		418 C1				
419	9.2	419 B1	419 B2	419 B3	419 B4	419 B5
		419 C1	419 C2	419 C3	419 C4	419 C5
420	9.1	420 A1	420 A2	420 A3	420 A4	420 A5
		420 B1	420 B2	420 B3	420 B4	420 B5
		420 C1	420 C2	420 C3	420 C4	420 C5
421	9.2			421 A3	421 A4	421 A5
		421 B1	421 B2	421 B3	421 B4	421 B5
		421 C1	421 C2	421 C3	421 C4	421 C5
615	9.3				615 B4	615 B5
		615 C1	615 C2	615 C3	615 C4	615 C5
619	9.3	619 C1	619 C2	619 C3	619 C4	619 C5
620	9.3	620 B1	620 B2	620 B3	620 B4	620 B5
		620 C1	620 C2	620 C3	620 C4	620 C5
623	9.4	623 C1	623 C2	623 C3	623 C4	623 C5
624	9.4	624 B1	624 B2	624 B3	624 B4	624 B5
		624 C1	624 C2	624 C3	624 C4	624 C5
626	9.5	626 C1	626 C2	626 C3	626 C4	626 C5
						627 B5
627	9.4	627 C1	627 C2	627 C3	627 C4	627 C5
						628 C5
628	9.5	628 C1	628 C2	628 C3	628 C4	628 C5
						629 C5
629	9.5	629 C1	629 C2	629 C3	629 C4	629 C5
						630 C5
630	9.6	630 C1	630 C2	630 C3	630 C4	630 C5

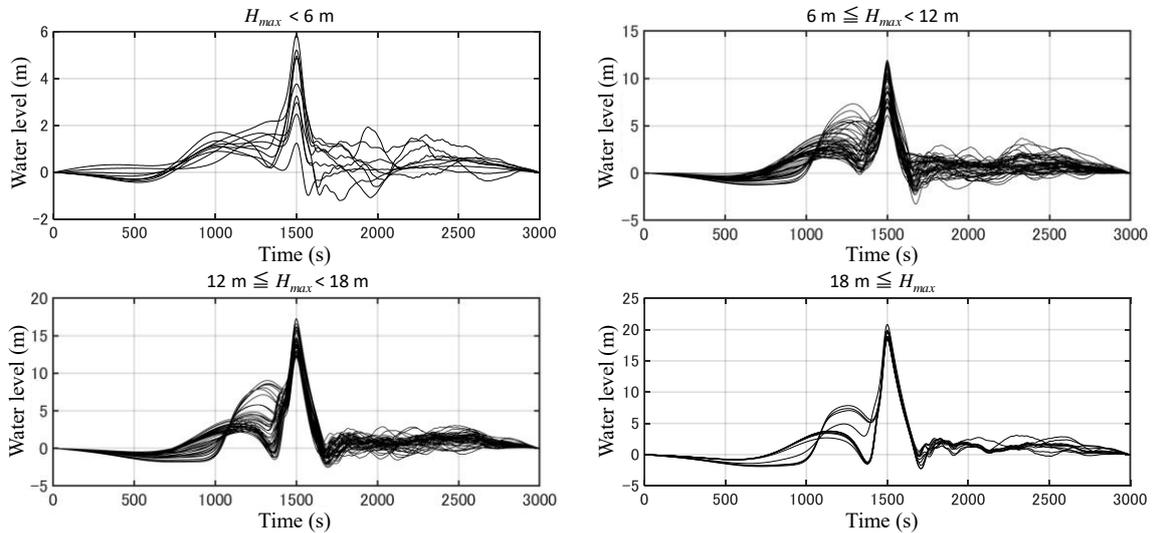


Fig. 15 The analytical waveforms comprising 118 scenario tsunamis selected from the interplate earthquakes assumed in the PTHA¹⁸⁾

large slip areas are changed. Here, while 1–3 are cases of a super-large slip area changing, 4 and 5 are the changing shape in two super-large slip areas.

In the analytical waveforms for the scenario tsunamis before selection, a mixture of cases is observed: where the maximum water level at the hazard assessment point occurs at a time zone that

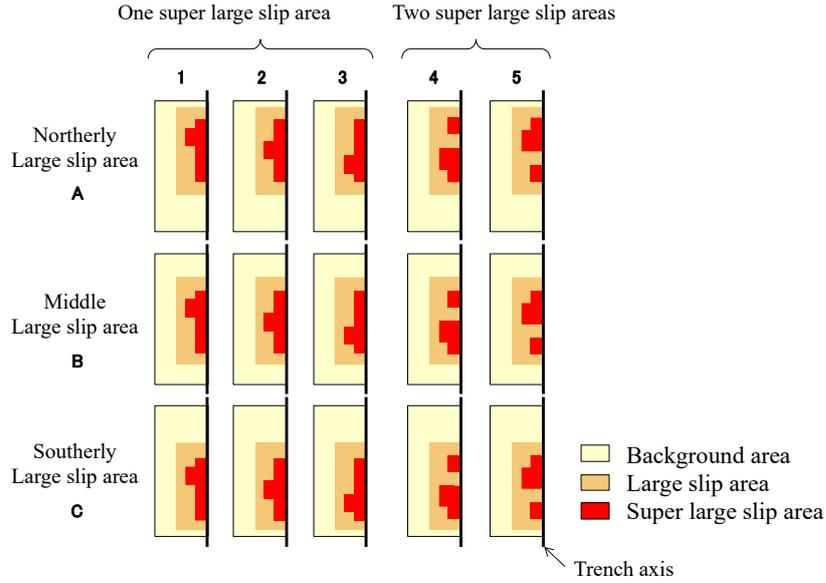


Fig. 16 The placement slip distribution patterns on scenario tsunami sources

arrives directly from a source area (Direct Wave) and where the waves reflected from the coastal area return and overlap (Superimposed Wave). Considering that the artificial tsunami waveform is an input condition of the tsunami propagation and run-up analysis on the tsunami fragility analysis for facilities located on the coast, adding the latter superimposed waves is not necessary. Therefore, when estimating the artificial tsunami waveform, we decide to exclude the latter. However, to distinguish between direct and superimposed waves, it remains necessary to know when the effects of reflected waves begin to appear during the analytical waveform.

Based on the above, we determine the hypothetical tsunami source model as a preparatory analysis, then perform two tsunami propagation analyses using a standard topographic model and another topographic model for passing waves, followed by a comparison of the analytical waveforms at hazard assessment points. In the latter topographic model, shallow waters and land area's elevation at an altitude of -100 m or more are uniformly changed to -100 m to suppress reflection from land. As a result, the reflection effect is confirmed about 45 min after the earthquake occurrence time. Therefore, we select scenario tsunamis induced by interplate earthquakes treated in the hazard analysis based on this result, in which the maximum water level occur before 45 min, and its level is 1 m or more. The selected scenario tsunamis shown in Table 2 are all sources at which the shallow and deep areas between the plates are destroyed simultaneously.

The abovementioned selection method and results for the analytical waveforms are trial calculation examples showing artificial tsunami waveforms. However, when using the proposed method in practice, the selection of analytical waveforms should be considered based on individual situations.

4.2.2 Example of creating phase/amplitude spectra models using a group of analytical waveforms

Following the procedure in Fig. 14, this study successfully generate phase/amplitude spectra models using the analytical waveforms described above.

First, the mean value $\tilde{\mu}^{(j,k)}_{tgr}$ and the standard deviation $\sigma^{(j,k)}_{tgr}$ of the group delay time are calculated for each $f^{(j,k)}$ after the wavelet transform of 118 analytical waveforms. The calculation results are shown in Table 3. Since the values in the table ought to be expressed in radians, we convert from seconds to radians by multiplying with $2\pi/T_d$ (T_d : Total data time ($= 1 \times 2^{15}$ seconds)). Incidentally, the time 1500 s, at which the relative maximum water level occur, is 0.29 in radian, which is helpful when considering the distribution in Fig. 17. Adopting examples of $j = 5$ and 7, Fig. 17 shows a histogram of $\mu^{(j,k)}_{tgr}$ from the source data of the calculated values in Table 3. Investigations reveal

Table 3 Mean and standard deviation of group delay time model [unit: rad]

j	12				11				10				9				8				7			
k	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1
$\tilde{\mu}^{(j,k)}_{tgr}$	0.46	0.51	0.53	0.54	0.53	0.52	0.50	0.40	0.44	0.37	0.34	0.30	0.30	0.29	0.28	0.29	0.29	0.29	0.29	0.31	0.33	0.29	0.29	0.35
$\sigma^{(j,k)}_{tgr}$	0.14	0.13	0.13	0.10	0.14	0.13	0.13	0.04	0.02	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.05

j	6				5				4				3				2			
k	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1
$\tilde{\mu}^{(j,k)}_{tgr}$	0.33	0.31	0.28	0.41	0.48	0.31	0.25	0.10	-0.28	0.13	0.24	0.70	0.77	0.47	0.30	0.02	-0.96	0.31	0.32	0.32
$\sigma^{(j,k)}_{tgr}$	0.05	0.05	0.06	0.13	0.15	0.04	0.03	0.08	0.12	0.06	0.03	0.07	0.09	0.01	0.02	0.07	0.05	0.02	0.02	0.02

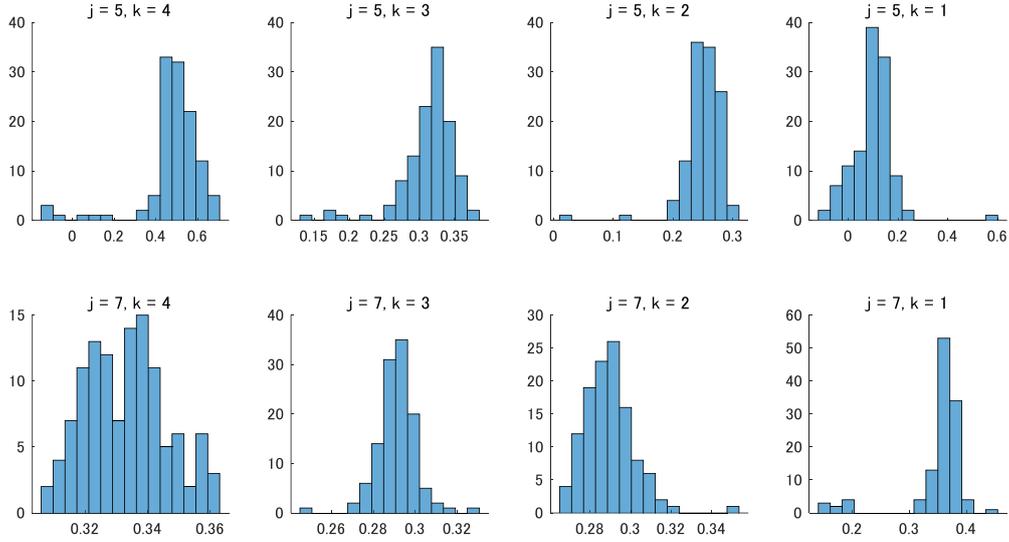


Fig. 17 Histograms of mean group delay times $\mu^{(j,k)}_{tgr}$ of the analytical waveforms

that while some of these distributions are clustered around the relative maximum water level time, 0.29 rad, some are not, causing the phase characteristics of each j and k in the analytical waveforms to be shown here.

Next, the Fourier transform obtain the amplitude spectrum for the analytical waveforms, followed by average amplitude spectrum calculations at each j and k . Then, regression analysis is performed using Eq. (15) to calculate coefficients $\alpha^{(j,k)}$, $\beta^{(j,k)}$ and $\sigma'^{(j,k)}_{amp}$. While the calculation results are shown in Table 4, scatterplots of the source data $\mu^{(j,k)}_{amp}$ and the regression models are shown in Fig. 18 as an example of $j = 5$ and 7. In this figure, the horizontal axis represents the relative maximum water level H_{1500} , the vertical axis represents $\mu^{(j,k)}_{amp}$, as displayed in a double logarithmic graph, the solid line represents the regression model of Eq. (15), and the dashed line represents the standard deviation $\pm 1\sigma$ from the regression model, respectively. Since we observe correlation between $\mu^{(j,k)}_{amp}$ and H_{1500} , whose tendencies are similar to another j and k , we decide to set a regression model using Eq. (15), where the intercept $\alpha^{(j,k)}$, slope $\beta^{(j,k)}$, and standard deviation $\sigma'^{(j,k)}_{amp}$ of the regression models show different values for each j and k . The amplitude characteristics of each j and k of the analytical waveforms are shown.

From the analysis, while a normal distribution is assumed as the group delay time model based on the distribution of the group delay time shown in Fig. 17, a lognormal distribution with the regression model as the mean is assumed as the amplitude spectrum model based on the regression analysis results

Table 4 Regression coefficients and standard deviations of amplitude spectrum model

j	12				11				10				9				8				7							
k	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1
$\alpha^{(j,k)}$	-7.23	-6.95	-6.64	-6.24	-5.93	-5.68	-5.25	-1.72	-0.16	-0.61	-0.49	-0.24	0.27	0.60	2.49	2.80	3.46	3.72	3.96	4.00	3.98	3.87	3.93	3.90				
$\beta^{(j,k)}$	0.82	0.80	0.80	0.80	0.81	0.83	0.83	0.48	0.52	0.69	0.98	1.12	1.20	1.25	0.70	0.82	0.62	0.55	0.60	0.85	1.03	1.15	1.21	1.21				
$\sigma^{(j,k)}_{amp}$	0.41	0.44	0.47	0.47	0.47	0.46	0.41	0.21	0.29	0.24	0.31	0.29	0.37	0.41	0.28	0.29	0.34	0.32	0.34	0.20	0.18	0.14	0.14	0.26				

j	6				5				4				3				2			
k	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1	4	3	2	1
$\alpha^{(j,k)}$	4.11	4.08	4.17	4.38	4.88	5.29	5.42	5.40	5.45	5.55	5.71	5.91	6.06	6.14	6.21	6.28	6.32	6.36	6.38	6.38
$\beta^{(j,k)}$	1.07	1.09	1.04	1.02	0.97	0.90	0.88	0.89	0.89	0.88	0.86	0.82	0.78	0.76	0.74	0.73	0.71	0.70	0.70	0.70
$\sigma^{(j,k)}_{amp}$	0.38	0.32	0.38	0.47	0.44	0.37	0.33	0.30	0.23	0.18	0.15	0.15	0.16	0.17	0.17	0.18	0.19	0.19	0.19	0.19

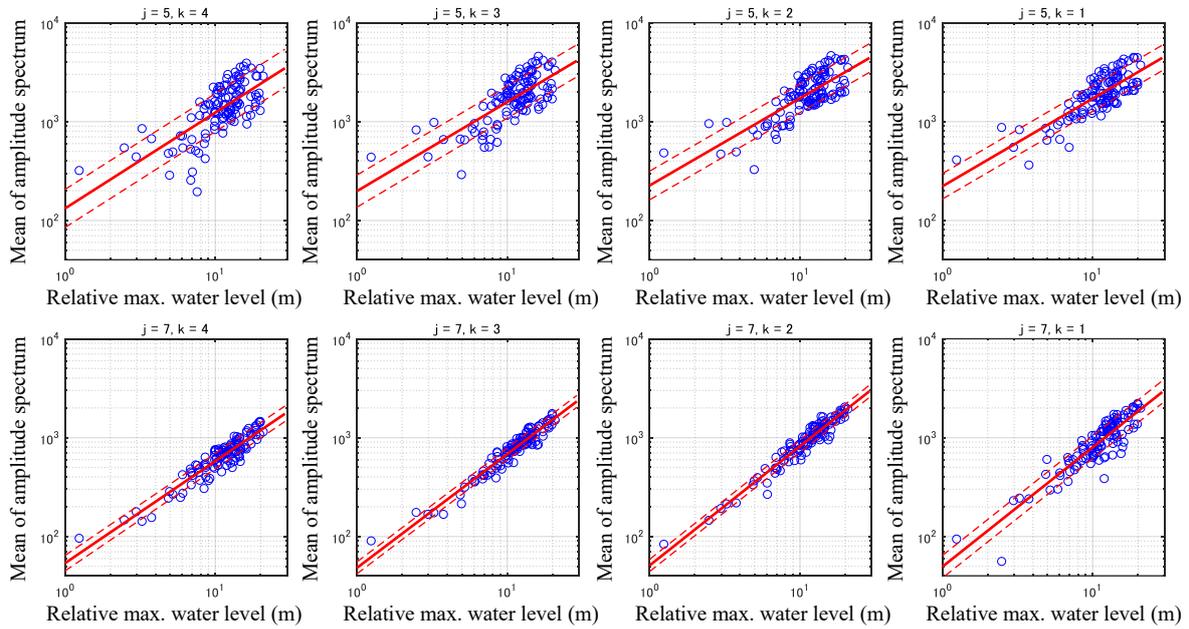


Fig. 18 Relationships between mean amplitude spectra and relative maximum water levels for the analytical waveforms and the regression models

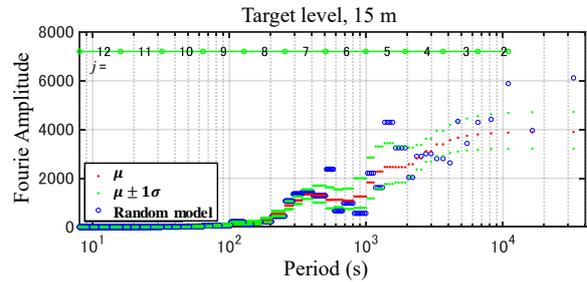
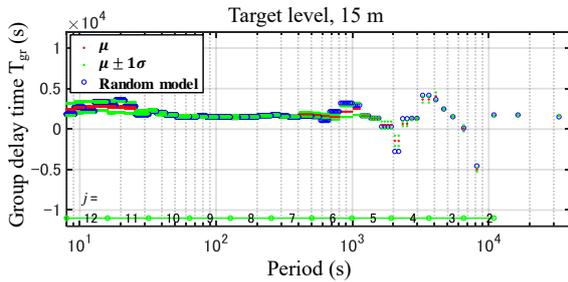


Fig. 19 Group delay time model by MC method Fig. 20 Amplitude spectrum model by MC method

of the mean amplitude spectrum shown in Fig. 18. We also generate normal random numbers by the MC method, including group delay time models and amplitude spectra models. During random number generation, the probability distribution of each j and k is assumed to be independent. Figure 19 shows the group delay time model, and Fig. 20 shows the amplitude spectrum model when the target water

level of the artificial tsunami waveform is 15 m. In both cases, the mean values of the probability distributions are shown as red points, the standard deviations of $\pm 1\sigma$ are shown as green points, and the models by random number generation are shown as blue circles. Overall, these results show examples of random number generation. Primarily, we discover that when different random number combinations are used, different group delay time models and amplitude spectra models are generated, resulting in artificial tsunami waveforms with different shapes, even at similar target water levels.

4.2.3 Examples of artificial tsunami waveforms

Figure 21 shows the artificial tsunami waveforms generated using the group delay time and amplitude spectrum models in Section 4.2.2, where the solid black lines in the figure show the artificial tsunami

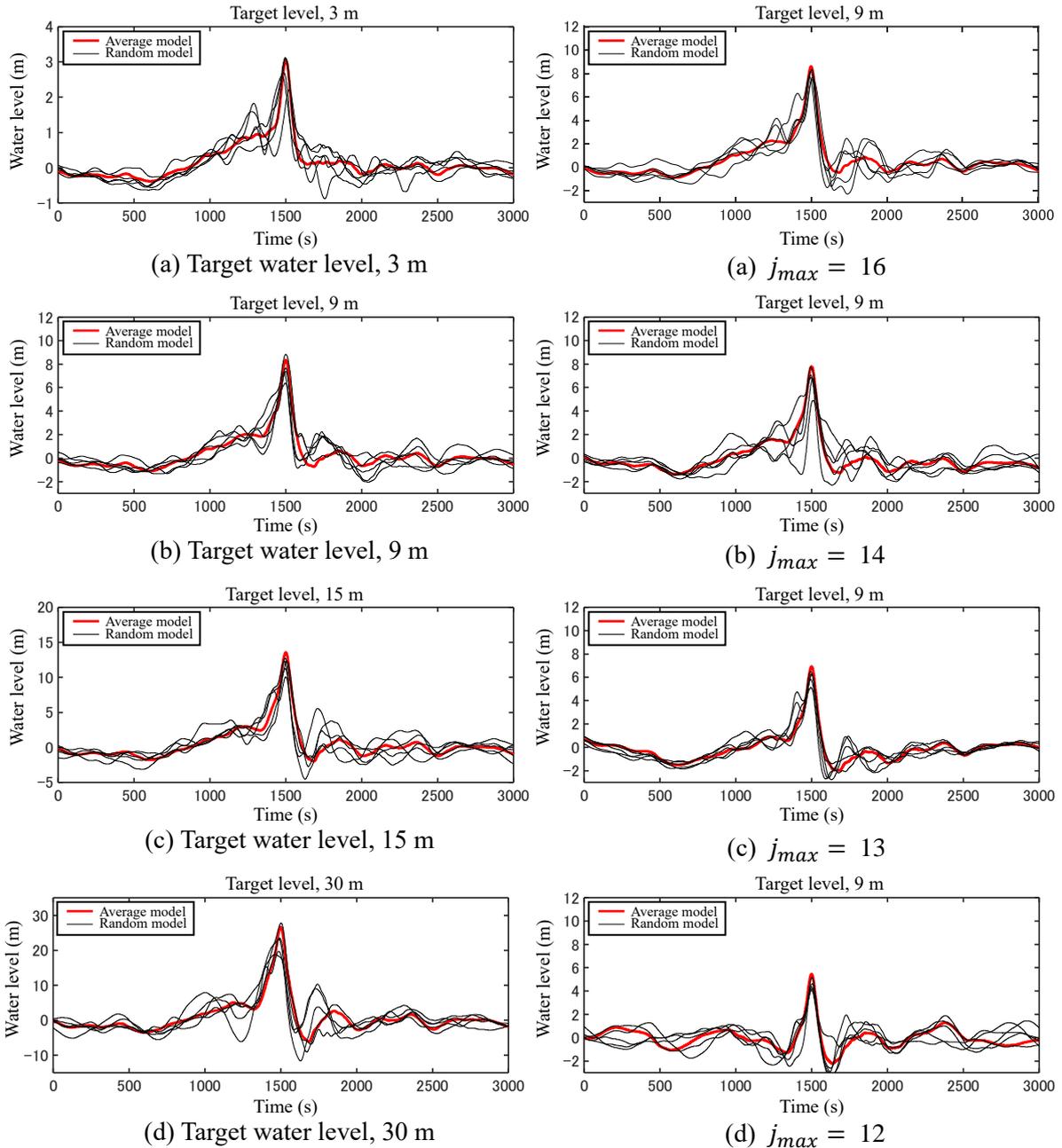


Fig. 21 Artificial tsunami waveforms using statistical method ($j_{max} = 15$)

Fig. 22 Effect of j_{max} on artificial tsunami waveforms (target level, 9 m)

waveforms (called a random model) generated by random numbers due to the MC method from the probability distributions of the group delay time and the amplitude spectrum models, respectively, and the solid red line in the figure shows the artificial tsunami waveform (called an average model) when both the group delay time and the amplitude spectrum models adopt the average value as a reference. Here, while four target water levels are selected: 3, 9, 15 and 30 m, five waves are shown for each target water level. The random model show different waveform shapes for any target water level, with the highest water level being around 1500 s. Investigations also reveal that the values are scattered near the target water level, which is a common finding. Therefore, these random models are distributed so that they are entangled with the average model. We also show six artificial tsunami waveforms as an example, in which generating several artificial tsunami waveforms with different shapes are possible by increasing the random numbers. Even then, variabilities around the average model are expected. Additionally, the method make it possible to generate artificial tsunami waveforms for any target water level by interpolating and extrapolating the regression equations of the amplitude spectrum model.

Based on the above findings, we believe that the successful selection of one appropriate artificial tsunami waveform from the many candidates, considering the uncertainty of the initial integral value in the phase spectrum modeling, is largely because the time of maximum water level occurrence t_a is fixed in advance. Therefore, we consider generating a phase spectrum model without fixing t_a and setting the time of earthquake occurrence to zero. While the time of interest can not be specified, the above method can not be used, leaving the issue of uncertainty in the initial integral value.

The example shown in Fig. 21 is also based on the case where the total amount of data after adding zero to the analytical waveforms is 2^{15} ($j_{max} = 15$). Although the study of Sato et al.⁶⁾ was designed to make the number of data sufficiently long, it remains debatable whether the same results as in Fig. 21 can be obtained by changing the number of zero (the total number of data, j_{max}). Therefore, a sensitivity analysis is performed by varying j_{max} from 16 to 12, then we compare the analysis results. Conversely, Fig. 22 shows the sensitivity analysis results when the target water level is set at 9 m. The legend in the figure is similar to that in Fig. 21. Here, we focus on the average model with a solid red line for comparison to simplify the discussion. When the j_{max} are 16 and 15, the difference is small, decreasing to 14 and 13, and the maximum water level at 1500 s gradually decreases. However, when it reaches 12, the maximum water level at the same time is even lower, significantly changing the overall waveforms. Thus, we are able to confirm that the artificial tsunami waveform generated differ depending on the value of j_{max} , which suggests that j_{max} above a particular value should be used to obtain stable analysis results. Still, the threshold value is expected to depend on the time interval and duration of the original analytical tsunami waveform, which requires further research.

The above results show that the proposed artificial tsunami waveforms satisfy all requirements listed in Chapter 2 for those to be used in the tsunami fragility analysis. However, the shape of the artificial tsunami waveforms represent by the random model, excluding the average model shown in Fig. 20, do not fully capture the shape of the analytical waveforms for the scenario tsunamis group shown in Fig. 15. This limitation may be due to the loss of continuity (correlation between adjacent data) between the group delay time and amplitude spectrum for a group of analytical waveforms due to the assumption that the probability distribution for each $f^{(j,k)}$ is independent when generating random numbers for the group delay time model and amplitude spectrum model. Thus, this point remains to be addressed in the future.

5. CONCLUSION

In this paper, to contribute to advancing the tsunami PRA method for nuclear power plants, we first summarized the requirements for artificial tsunami waveform for an close/organic linkage between PTHA and tsunami fragility analysis. Next, we conducted basic studies to model the phase and amplitude spectra using observed waveforms from past tsunamis, then proposed a method for generating artificial tsunami waveforms applying statistical methods based on these requirements and the results of basic studies. It is noted that this paper assumed the artificial tsunami waveforms to be the input conditions

for the tsunami run-up analysis in the tsunami fragility analysis. We summarize the artificial tsunami waveform requirements and their respective treatments using the proposed method as follows:

Requirement 1 was for the artificial tsunami waveforms to be defined at the PTHA target point. In the proposed method, artificial tsunami waveforms were calculated at the point, and the use of the information regarding phase and amplitude characteristics from the analytical waveforms on PTHA was the key to the close/organic linkage between PTHA and tsunami fragility analysis.

Requirement 2 was for the artificial tsunami waveforms to be defined at any tsunami water level in the tsunami hazard curve. In the proposed method, we estimated the maximum water level and average amplitude spectrum in each periodic band for analytical waveforms which were used to estimate the tsunami hazard curve, created a regression model with the maximum water level as an explanatory parameter based on these data groups and treated this requirement. By extrapolating the maximum water level range of the analytical waveforms with this model, it was possible to set the target water level beyond this range with an explanation.

Requirement 3 was that the variabilities in phase and amplitude owing to regional characteristics were statistically and rationally considered on the artificial tsunami waveforms. In the tsunami fragility analysis, even if similar maximum water levels were reached at an offshore location, different waveforms resulted in different inundation conditions on land, suggesting the necessity of including variabilities in the phase and amplitude characteristics of the artificial tsunami waveforms. To this end, the proposed method considered that the analytical waveforms in PTHA have regional characteristics related to the tsunami occurrences, the requirement to consider the regional characteristics was addressed by modeling the phase and amplitude characteristics at each periodic band using the statistical analysis. The requirement to account for phase and amplitude characteristics variabilities was also addressed by combining the MC method with probability distribution assumptions based on the statistical analysis results.

Requirement 4 was for the artificial tsunami waveforms to consider the effects of crustal uplift and subsidence associated with fault movement. Following this study's approach, this requirement had to be first treated uniformly with PTHA, tsunami fragility analysis and artificial tsunami waveforms for close/organic linkage. Then, the proposed method addressed this requirement by modeling the phase and amplitude characteristics based on the analytical waveforms in PTHA and expressing them as relative water levels that consider the uplift and subsidence of the land area.

Despite the enormous findings from this study, some limitations were encountered. First, from our artificial tsunami waveform calculations, application examples were given only for cases where the shallow and deep parts of the plate boundary ruptured simultaneously. Even for interplate earthquakes, the phase and amplitude spectra naturally differ when only the shallow or deep part of the plate boundary is active or when different categories of earthquakes, such as crustal earthquakes, are targeted. Therefore, in the future, we will confirm the applicability of the proposed method to these analytical waveforms and extend the scope of application. Additionally, we have considered that the artificial tsunami waveforms here were assumedly treated as forced water level fluctuations and boundary inputs to the analysis domain, making them the input conditions for tsunami propagation and run-up analysis. However, the tsunami propagation, which originally is a two-dimensional spread, is simplified to a one-dimensional model. Nevertheless, this approach has not been verified, indicating another research gap for tsunami fragility analysis using the proposed artificial tsunami waveform method. Hence, all these issues should be addressed in future studies. Finally, note that the proposed method can not apply to the waveforms generated in shallow water or inundation areas with strong nonlinearity because the proposed artificial tsunami waveforms are generated at offshore locations.

ACKNOWLEDGMENT

We thank Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) for allowing us to use the observed waveform data during the Tohoku Earthquake Tsunami published by NOWPHAS (National Ocean Wave Information Network for Ports and Harbors), and thank for us to use WaveLab version.850

for the wavelet transform and inverse transform. Furthermore, we thank Dr. MURONO Yoshitaka for his lecture on the Fourier phase spectra unwrap method. We also used some of the project results of the Nuclear Regulation Authority for the regulatory research on seismic and tsunami safety design of nuclear facilities (Analysis of Observed Tsunami Waveforms and Generation of Artificial Tsunami Waveforms) in the fiscal year 2015.

REFERENCES

- 1) Atomic Energy Society of Japan: *Implementation Standard Concerning the Tsunami Probabilistic Risk Assessment of Nuclear Power Plants:2016*, 2019 (in Japanese).
- 2) Sugino, H., Iwabuchi, Y., Nishio, M., Tsutsumi, H., Sakagami, M. and Ebisawa, K.: Development of Probabilistic Methodology for Evaluating Tsunami Risk on Nuclear Power Plants, *The 14th World Conference on Earthquake Engineering (14WCEE)*, Beijing, China, S14-15-0035, 2008.
- 3) Iwabuchi, Y., Sugino, H. and Ebisawa, K.: Guidelines for Developing Design Basis Tsunami Based on Probabilistic Approach, Japan Nuclear Energy Safety Organization, *JNES-RE-2013-2041*, 193 pp., 2014 (in Japanese).
- 4) Kihara, N., Kimura, T., Kaida, H., Iizuka, K. and Fujii, N.: A Methodology of Determination of Scenario Tsunami for Fragility Assessment Based on Hazard Deaggregation—Determination Method of Tsunami Conditions Arriving at a Site with a Specific Scale—, *Journal of Japan Association for Earthquake Engineering*, Vol. 18, No. 1, pp. 35–58, 2018 (in Japanese).
- 5) Ohsaki, Y.: Guideline of Evaluation of Basic Design Earthquake Ground Motions for Nuclear Power Plants, Mainly on the Ohsaki Spectrum, *ORI Research Report*, No. 84-01, 101 pp., 1984 (in Japanese, title translated by the authors).
- 6) Sato, T., Murono, Y. and Nishimura, A.: Empirical Modeling of Phase Spectrum of Earthquake Motion, *Journal of Japan Society of Civil Engineers*, No. 640, I-50, pp. 119–130, 2000 (in Japanese).
- 7) Sasaki, F., Maeda, T. and Yamada, M.: Study of Time History Data Using Wavelet Transform, *Journal of Structural Engineering*, Vol. 38B, pp. 9–20, 1992 (in Japanese).
- 8) Railway Technical Research Institute: *Design Standards for Railway Structures and Commentary (Seismic Structural Design)*, MARUZEN, 444 pp., 2012 (in Japanese).
- 9) Atomic Energy Society of Japan: *A Standard for Procedure of Seismic Probabilistic Risk Assessment for Nuclear Power Plants:2015*, 1012 pp., 2015 (in Japanese).
- 10) Matsumura, A.: *Daijirin 3*, Sanseido, 2976 pp., 2006 (in Japanese).
- 11) Fuglede, B. and Topsoe, F.: Jensen–Shannon Divergence and Hilbert Space Embedding, *IEEE International Symposium on Information Theory*, pp. 31–36, 2004.
- 12) Kawai, H., Satoh, M., Kawaguchi, K. and Seki, K.: The 2011 Off the Pacific Coast of Tohoku Earthquake Tsunami Observed by GPS Buoys, *Journal of Japan Society of Civil Engineers B2 (Coastal Engineering)*, Vol. 67, No. 2, pp. I_1291–I_1295, 2011 (in Japanese).
- 13) Buckheit, J., Chen, S., Donoho, D., Johnstone, I. and Scargle, J.: About WaveLab, version .850, 2005. http://statweb.stanford.edu/~wavelab/Wavelab_850/index_wavelab850.html (last accessed on May 26, 2020)
- 14) Sawada, S., Morikawa, H., Toki, K. and Yokoyama, K.: Identification of Path and Local Site Effects on Phase Spectrum of Seismic Motion, *The 10th Earthquake Engineering Symposium Proceedings*, pp. 915–921, 1998 (in Japanese).
- 15) Meyer, Y.: Orthonormal Wavelets, *Springer*, pp. 21–27, 1989.
- 16) MathWorks: Documentation “MATLAB Function”.
https://jp.mathworks.com/help/matlab/referencelist.html?type=function&s_cid=doc_ftr (last accessed on January 6, 2021) (in Japanese)
- 17) Matsumoto, M. and Nishimura, T.: Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-random Number Generator, *ACM Transactions on Modeling and Computer Simulation*, Vol. 8, No. 1, pp. 3–30, 1998.
- 18) Sugino, H., Iwabuchi, Y., Abe, Y. and Imamura, F.: Effect of the Modeling Method for Scenario

Tsunami on the Probabilistic Tsunami Hazard Assessment, *Journal of Japan Association for Earthquake Engineering*, Vol. 15, No. 4, pp. 40–61, 2015 (in Japanese).

- 19) Sugino, H., Iwabuchi, Y., Hashimoto, N., Matsusue, K., Ebisawa, K., Kameda, H. and Imamura, F.: The Characterizing Model for Tsunami Source Regarding the Interplate Earthquake Tsunami, *Journal of Japan Association for Earthquake Engineering*, Vol. 14, No. 5, pp. 1–18, 2014 (in Japanese).

(Original Japanese Paper Published: February 2021)

(English Version Submitted: February 24, 2023)

(English Version Accepted: April 13, 2023)