



Robust uncertainty quantification of the volume of tsunami ionospheric holes for the 2011 Tohoku-Oki earthquake: towards low-cost satellite-based tsunami warning systems

Ryuichi Kanai^{1,2}, Masashi Kamogawa⁴, Toshiyasu Nagao⁵, Alan Smith³, and Serge Guillas^{1,2}

¹Department of Statistical Science, University College London, London, UK

²The Alan Turing Institute, London, UK

³Department of Space and Climate Physics, University College London, London, UK

⁴Global Center for Asian and Regional Research, University of Shizuoka, Shizuoka, Japan

⁵Institute of Oceanic Research and Development, Tokai University, Shizuoka, Japan

Correspondence: Ryuichi Kanai (ryuichi.kanai.16@ucl.ac.uk)

Received: 21 April 2021 – Discussion started: 12 July 2021

Revised: 8 February 2022 – Accepted: 8 February 2022 – Published: 15 March 2022

Abstract. We develop a new method to analyze the total electron content (TEC) depression in the ionosphere after a tsunami occurrence. We employ Gaussian process regression to accurately estimate the TEC disturbance every 30 s using satellite observations from the global navigation satellite system (GNSS) network, even over regions without measurements. We face multiple challenges. First, the impact of the acoustic wave generated by a tsunami onto TEC levels is nonlinear and anisotropic. Second, observation points are moving. Third, the measured data are not uniformly distributed in the targeting range. Nevertheless, our method always computes the electron density depression volumes, along with estimated uncertainties, when applied to the 2011 Tohoku-Oki earthquake, even with random selections of only 5 % of the 1000 GPS Earth Observation Network System receivers considered here over Japan. Also, the statistically estimated TEC depression area mostly overlaps the range of the initial tsunami, which indicates that our method can potentially be used to estimate the initial tsunami. The method can warn of a tsunami event within 15 min of the earthquake, at high levels of confidence, even with a sparse receiver network. Hence, it is potentially applicable worldwide using the existing GNSS network.

1 Introduction

The damage caused by tsunamis can be devastating. For example, almost 20 000 people died in the tsunami following the 2011 Tohoku-Oki earthquake in Japan. One reason for such levels of casualties is that current tsunami height predictions are relatively unreliable, even following an identified earthquake event, and so early warning systems are not as effective as required. Initial sea surface deformations are typically indirectly determined from seismological inversions of the earthquake source. However, some of these early estimates are sometimes much lower than expected: for instance the initially estimated value for the 2011 Tohoku-Oki earthquake of M_w 7.9 was used for warnings, but the actual magnitude was M_w 9.1.

Research on tsunami warnings has been conducted for a long time and has undergone remarkable technical evolution with the development of various technologies (Bernard and Titov, 2015; Wächter et al., 2012). For example, in recent years, tsunami warning systems have been developed for tsunamis of seismic origin in and around the Mediterranean Sea (Amato et al., 2021).

Furthermore, the initial-tsunami wave cannot be precisely inferred from seismic information alone due to the complexity of the relationship between the earthquake source and the initial wave. For example so-called tsunami earthquakes generate much larger tsunamis than expected from the seismic source, e.g., the 2010 Mentawai tsunami (Lay et al., 2011;

Satake et al., 2013), whereas some powerful earthquakes sometimes produce tsunamis much smaller than expected, e.g., for the M_w 8.6 2005 Nias earthquake. These deficiencies in the seismic approach become even greater when considering additional contributions to the tsunami wave such as splay faults and submarine landslides not well picked up by seismic monitoring. One could account for the uncertainties in the earthquake source estimates and propagate these to the initial-tsunami height in real time (Giles et al., 2021), but these approaches cannot realistically model in 3D and in real time the seabed deformation arising from the earthquake source due to epistemic, computational, and observational inadequacies. In addition, in some cases, seismic gaps and co-seismic slips do not overlap, which makes tsunami prediction based on seismic data highly difficult (Lorito et al., 2011). Hence, Bernard and Titov (2015) illustrate a more advanced warning system that uses detecting devices such as pressure sensors and predicts tsunami heights using tsunami information. Specifically, observations closely related to the actually generated tsunami wave are more likely to provide more precise warnings. One example is the successful data assimilation of tsunami waves from buoys, with either dense or possibly sparse networks (Tanioka and Gusman, 2018; Wang et al., 2019). Still, some problems remain, such as the high maintenance cost of those devices and avoiding high ocean currents for installation locations (Bernard and Titov, 2015). We explore here the use of real-time satellite data due to their global coverage, low expense, low maintenance, and rapid access.

A path towards accurate warnings is to estimate the tsunami ionospheric holes (TIHs) generated in the ionosphere after the initial-tsunami occurrence (Kamogawa et al., 2016). The formation of a TIH, which is a decrease in total electron content (TEC) in the ionosphere, can be explained by the following physical mechanisms (Kamogawa et al., 2016; Shinagawa et al., 2013). First, a displacement of the sea surface caused by a tsunami generates acoustic waves that propagate vertically upward and reach the ionosphere. Then, the plasma is moved along the magnetic field by the sound waves, and the downward flow is larger than the upward flow partly because the gravity force causes downward motion. The downward plasma causes recombination, and ion production is suppressed, resulting in a decrease in TEC, and the depression in TEC is called a TIH. The TIH observed in the ionosphere at the time of the 2011 Tohoku-Oki earthquake has been reproduced by performing numerical simulations of this physical phenomenon (Shinagawa et al., 2013; Zettergren et al., 2017; Zettergren and Snively, 2019).

In Japan, the GPS Earth Observation Network System (GEONET), which is a network of more than 1200 receivers, enables us to observe the behavior of TEC in the ionosphere with a large number of data points. The most prominent case of the TEC changes in the ionosphere observed by GEONET is the tsunami following the 2011 Tohoku-Oki earthquake. By focusing on the changes in the ionosphere after the earth-

quake and observing the high-frequency component of the TEC fluctuations, Tsugawa et al. (2011) observed a rapid decrease in TEC near the epicenter approximately 7 min after the earthquake: the rapid fluctuation of the high-frequency component of TEC was detected as concentric waves that radiated outward, and these concentric waves were confirmed to have had a central point source. A. Saito et al. (2011) analyzed the unfiltered TEC fluctuations in which a significant decrease in TEC was observed, with an amplitude of up to 5 TECU (TEC units), which is 1.0×10^{16} electrons m^{-2} , and an area of 500 km. Similarly, Kakinami et al. (2012) showed that the amplitude of the decrease in TEC exceeds 5 TECU, analyzing TEC without frequency filtering.

Furthermore, Kamogawa et al. (2016) examined the behavior of the TEC depression in the ionosphere after the tsunami, examining the low-frequency component of TEC in a variety of tsunami cases including the 2011 Tohoku-Oki earthquake. They discovered a positive correlation between the initial-tsunami height and the rate of TEC depression. It is thus likely possible to detect an initial tsunami by evaluating the magnitude of a TIH, which is the reduction of TEC in the ionosphere. However, it is still challenging to define the scale of a TIH, because even if a dense network of global navigation satellite system (GNSS) receivers is maintained, such as in Japan, there are areas where TEC cannot be measured by the network. Moreover, the TEC measurement locations move in the same way as the satellite moves, and those locations are not uniformly distributed within the target range. The shape of the TIH cannot be completely captured from the measurement points alone. In addition, in regions where GNSS observation networks are less dense, the number of available data is even smaller, making it very difficult to detect the TIH confidently.

To overcome these problems, we implement below a statistical method for the analysis of TEC using satellite data, which allows us to estimate TEC values even over areas with no measurements and to evaluate the whole TIH even without a dense measurement network such as GEONET in Japan. Our approach does not make any assumptions about the nature of the source of the tsunami. This method enables us to calculate the volume (with uncertainty) of the hole as an assessment of the scale of the TIH, and we propose to use its volume as a measure of the TIH. We believe that estimating the TIH provides a new and important tool for tsunami early warning systems that is independent of seismology.

In Sect. 2, the pre-processing and characteristics of the data are described in detail to ensure that this study is reproducible. In addition, we describe our surface-fitting method. In Sect. 3, we present the results of fitting surfaces computed by our new method and the time series analysis of the TIH volume. In Sect. 4, we conclude and mention future possibilities offered by this method.

2 Data and method

2.1 Data

In this study, TEC is calculated using GEONET data operated by the Geospatial Information Authority of Japan, and the following assumptions are made in processing the data (Heki, 2022; Hatanaka, 2022). First, we approximate the F region, which contains many more electrons than other regions in the ionosphere, as a thin layer at an altitude of 300 km, because the two effects of the chemical reactions and diffusion are balanced, and the electron density is maximized at an altitude of 300 km. According to the results of the analysis of Maruyama et al. (2011) of ionogram information on the day of the 2011 Tohoku-Oki earthquake, the electron density peak of the ionosphere was at an altitude of 306 km in the data from Kokubunji, which is the closest to the epicenter at 440 km. This result suggests that the assumption of a hypothetical thin ionosphere at an altitude of 300 km in our analysis is reasonable. Note that the analysis of Maruyama et al. (2011) was conducted for the 11 March 2011 earthquake, and a similar analysis is needed to determine whether setting the ionosphere to 300 km is the most appropriate assumption for other earthquake cases. The point where the line connecting a GNSS satellite and a receiver intersects with this approximated thin layer is called the ionospheric point (IP). The footprint of the IP to the surface is called the sub-ionospheric point (SIP).

Two radio signals from the GNSS satellites, 1575.42 and 1222.60 MHz, are transmitted to the GNSS receivers, and the propagation time of the radio signals depends on the electron density in the atmosphere. Therefore, TEC between the GNSS satellites and the GNSS receivers can be estimated from the phase delay of these two types of radio signals. This TEC, which is in the pathway between a satellite and a receiver, is called the slant TEC, noting that in general the line of sight to the satellite is not vertical. The slant TEC at the time of the earthquake is used as the reference value for the time series slant TEC data. The time series slant TEC is defined as the difference between the slant TEC and the reference TEC value for each satellite receiver pair.

For each time series slant TEC data, a quadratic fitting is performed by the ordinary least-squares method for data points from 30 min before to 7 min after the time of the earthquake to be consistent with the previous study (Kamogawa et al., 2016). These fitting curves are assumed to represent the time series slant TEC data, as it would have been in the absence of the effect of TEC depression caused by acoustic waves induced by the tsunami because it takes almost 7 min for acoustic waves to reach the ionosphere.

Then, we calculate the difference between the fitting curves and the time series slant TEC for each case. By multiplying the time series differences by the cosine of the angle θ between the vertical upward direction and the straight line between the satellite and the receiver, we obtain $\Delta v\text{TEC}$,

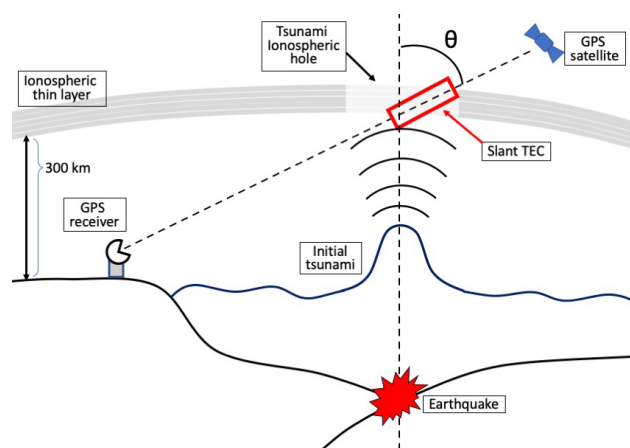


Figure 1. The schematic image of TEC depression detected by a satellite and a receiver.

which is the variation of the vertical component of the slant TEC time series data. The conceptual diagram of the description of the data processing so far is drawn in Fig. 1.

To apply a low-pass filter to this $\Delta v\text{TEC}$, we take a 300 s backward-moving average to obtain the low-frequency components of $\Delta v\text{TEC}$. Since a TIH is a hole formed by the decrease of TEC, we want to focus on the decrease of TEC in our analysis. For this reason, in the following sections, we use the data of which the low-frequency $\Delta v\text{TEC}$ is less than 1. The low-frequency $\Delta v\text{TEC}$ is hereinafter referred as TEC for the sake of simplicity.

Since the time resolution of the available data is the 30 s interval, we set the time of the 2011 Tohoku-Oki earthquake occurrence to 05:46:30 (UTC) even though the exact occurrence time is 05:46:18 (UTC) according to the Japan Meteorological Agency.

In addition, the data include outliers due to broken receivers. We detected them using the k -nearest neighbor algorithm (Cover and Hart, 1967) and removed these outliers from the data to be analyzed. The general principle of the k -nearest neighbor algorithm is that for a given data point, k -nearest neighbor data sets are identified, and labels are assigned to these k -nearest neighbor data sets and the given data point. Generally, this method requires the definition of the metric for measuring the distance (similarity) between data points, and the number k needs to be chosen as well. Here, simple but effective choices are made: the Euclidean distance is used to measure the similarity, and k is defined as the square root of the number of data points in the targeting range.

2.2 Robust fitting method with Gaussian process regression

We analyze data over the area of 10° of latitude and 10° of longitude centered at 38.297° N, 142.373° E, the location of the epicenter of the 2011 Tohoku-Oki earthquake as reported

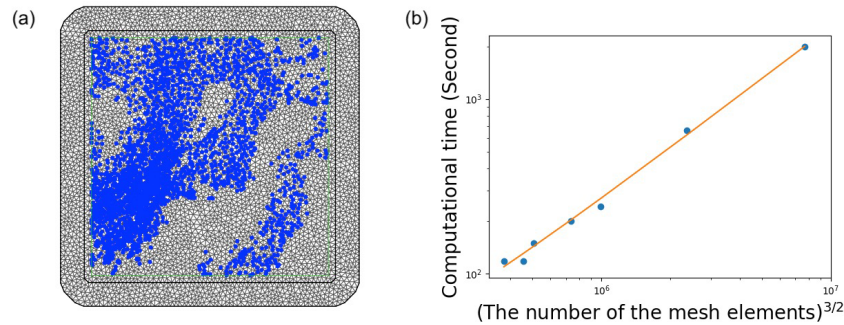


Figure 2. Panel (a) is the mesh elements and observed data, which is plotted with a blue marker, at 06:00:00 UTC. The number of mesh elements is about 5200. Panel (b) is the relationship between computational time and the number of the mesh elements to the 3/2nd power. The computation time is the average value of the time at 05:46:30, 06:00:00, 06:08:00, and 06:16:00. The orange solid line is the regression line. The coefficient is 0.000259, and the intercept is 12.044.

by the USGS. Gaussian process (GP) regression (Rasmussen and Williams, 2006) is a method of regressing a function Y (here TEC is a function of horizontal coordinates) using a flexible nonlinear model based on a set of observed data. A GP is in fact a generalization of the multivariate normal distribution to infinite dimensions: any marginal distribution projected to finite dimensions is multivariate normal. The fitted GP here probabilistically represents all possible TEC surfaces that interpolate (up to a so-called nugget noise level) the observations. We employ here the Matérn kernel with an additional nugget that accounts for some noise about the observations:

$$k_v(x_p, x_q) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{l} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}r}{l} \right), \quad (1)$$

$$\text{cov}(y_p, y_q) = k_v(x_p, x_q) + \sigma^2 \delta_{p,q}. \quad (2)$$

Here, $\text{cov}()$ means a covariance function, $r = |x_p - x_q|$, K_ν is a modified Bessel function of the second kind, Γ is the Gamma function, ν and l are positive parameters, σ^2 is the variance of the noise (i.e., the nugget), and $\delta_{p,q} = 1$ if $p = q$ and 0 otherwise. The Matérn kernel's smoothness ν generates a GP whose smoothness relates to ν and should thus be carefully chosen to match the smoothness of the function Y . By setting $\nu = 5/2$, we use a kernel function that is twice differentiable, which, in our analysis, conforms very well to the physical phenomena of TEC reduction.

After fitting our GP, the joint distribution of the estimates at any new locations are estimated (with uncertainty) even in areas where there is no measurement data. Here we predict the TEC surface over the area in increments of 0.01° in both latitude and longitude. However, using 1200 receivers, it takes more than 10 min to fit the full data due to costs of $\mathcal{O}(n^3)$; that is, the computational cost is proportional to the cubic of the number of data points n .

A stochastic partial differential equation (SPDE) approach using the integrated nested Laplace approximation (INLA) (Lindgren et al., 2011; Rue et al., 2009) can reduce the cost

Table 1. Computational time for the TEC surface fitting and the TEC value estimation.

	GP regression	INLA-SPDE
Mean time (s)	763.9	43.1
Standard deviation time (s)	151.2	4.5

of fitting the GP. Such an approach is not only faster but also has demonstrated that spatial predictions are more accurate, less uncertain, and more robust than the standard covariance-based fitting of a GP, e.g., when mapping stratospheric ozone (Chang et al., 2015). In other words, the estimated values provided by the method are closer to the true values and have less dispersion in the spatial interpolation, and the method is robust against the absence of measurements. Exploiting Gauss–Markov random fields (GMRFs), the INLA-SPDE reduces costs to $\mathcal{O}(n^{\frac{3}{2}})$ for two dimensions. The crucial point is that a Gaussian spatial process with a Matérn covariance function is the stationary solution to a certain SPDE that can be solved using finite-element approaches and approximated using the INLA in the GMRF setting. Nevertheless, some effort must be put into creating a reasonable mesh that solves the SPDE using finite elements, shown in Fig. 2.

The number of elements in the mesh cannot be too large, as the computational burden would become too high, and not too small, as the fitting surface would not be a good approximation of the actual surface.

Using this INLA-SPDE method with about 5200 mesh elements, the average computational time to fit the full data and predict the surface in 30 s increments from 05:30:00 to 06:30:00 becomes less than 1 min, with a standard deviation of less than 5 s, whereas the average computational time based on the standard GP regression method is more than 10 min, with a standard deviation of more than 2 min. The exact values are described in Table 1.

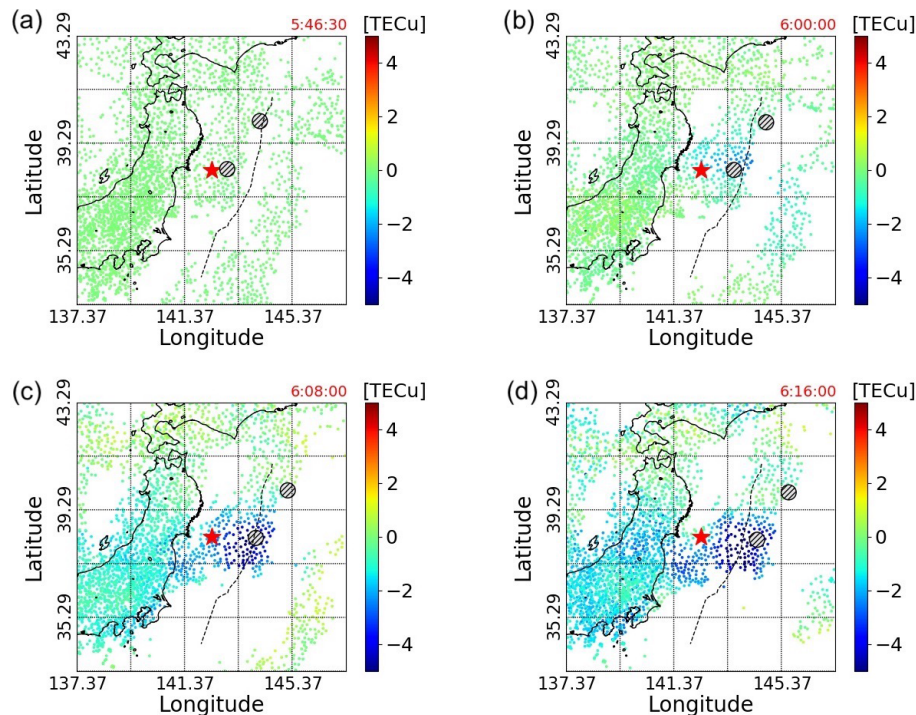


Figure 3. The red star is the epicenter, and the two large black circles are the outliers. Panels (a–d) are the TEC data measured by the GNSS network at 05:46:30, 06:00:00, 06:08:00, and 06:16:00 (UTC), respectively; 05:46:30 was the time of the earthquake occurrence. The black dashed line indicates the location of the axis of the Japan Trench.

3 Results

Figure 3 displays the measured TEC data at different times. It can be seen that the measurement points are moving and are not uniformly distributed in the targeting range, which is 33.297–43.297° N, 137.373–147.373° E. Moreover, although it can be confirmed from Fig. 3c and d that the TIH is formed, we argue that a single data point alone such as the minimum TEC value is not enough to evaluate the scale of the TIH. Indeed, doing so accounts for neither the width nor the anisotropy of the TIH. Figure 3 highlights the fact that in order to properly evaluate the TIH, it is necessary to analyze the data using statistical methods, rather than using the observed data as is.

3.1 Outlier detection

For the two data points that are determined to be outliers by our method, we validate them as outliers as follows. Figure 4a is the 3D plot of the captured data by satellites at 06:00:00, and the two red dots are the points identified as outliers. It is clear that these two outliers have very different values from their neighboring measurement points, but such spikes do not correspond to any genuine physical variations. These points will move further away from the surrounding points over time, and eventually, the absolute values of the

two outliers reach over 50 TECU, which can distort the fitting surface considerably and prevent proper TIH analysis.

In addition, to validate further, we create a semi-variogram cloud. In the semi-variogram cloud, half the value of the squared difference between feature values of two data points is plotted against the difference in geographical space between the two data points. Figure 4b shows the semi-variogram cloud of the data at 06:00:00 (UTC), where the red points are associated with points considered outliers by our method. We can see that these two points present extremely different values when compared to the other measurement points. This corresponds to a clear lack of correlation between these two points and the rest of the data set.

Our method identifies the receivers that correspond to the outliers. In this case, these outliers are observed by the receiver 960588 and 950175, respectively. According to the Geospatial Information Authority of Japan, either the antennas of the receiver or the receiver itself was replaced by a new one in the year following the 2011 Tohoku-Oki earthquake.

In the analysis, it is inappropriate to include observations measured by receivers that would have been broken. Therefore, the exclusion of these outliers is essential to the TIH analysis, and hence all the analyses in this study are implemented after removing the outliers.

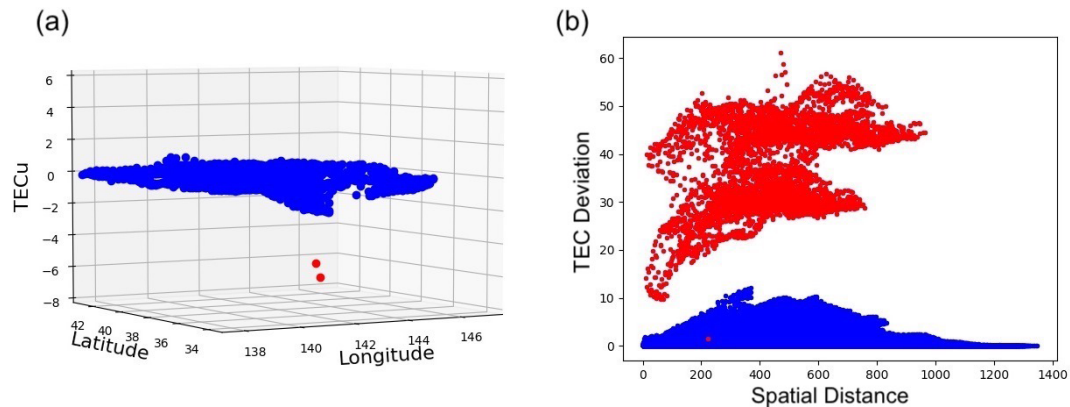


Figure 4. Panel (a) is the 3D plot of the captured data. The two red dots are outliers identified by our method. Panel (b) is the semi-variogram cloud of the data. In both panels (a) and (b), the data depicted and analyzed are captured by satellites at 06:00:00 (UTC).

3.2 Surface fitting with full data

Figure 5 shows that measured TEC data, 3D plot of its fitting surface, and 2D mapping of the fitting surface and confidence interval (CI) for both full data and sparse data at 06:08:30 (UTC). Although Fig. 5a shows that the observed TEC decreases near the epicenter, the position where it decreases the most and the range of the depression cannot be described in detail due to the limited number of data points. In addition, the degree of TEC decrease in Fig. 5a and the 3D plot of the TEC depression represented by blue dots in Fig. 5c show that the TEC values gradually change from 0 in the area where the TEC values continue to be stable to the minimum value in the region where they are decreasing the most. These facts indicate that, in the evaluation of a TIH, it is necessary to estimate the TEC values where no captured data exist, in order to use a measure that can take the overall variation into account rather than using a single data point.

Here, we present our method's success in surface fitting with uncertainty as expressed in Fig. 5c, e and g. In Fig. 5c, the fitting surface is depicted with red colors. The fitted surface in Fig. 5c is an almost perfect fit to the TEC data observed by the satellites. We can thus confirm that the TEC values on the surface do not change linearly from the stable area to its minimum. This fact indicates that the TEC oscillation, which is thought to be the effect of high-frequency components (Tsugawa et al., 2011), remains even after implementing the data pre-processing, which is a low-pass filter. Figure 5e, which is the 2D projection of the fitting surface, shows that our method enables us to estimate the TEC values with a fine granularity even over the region where the data are not detected by the GNSS satellites and receivers. The estimated values are computed and displayed in increments of 0.01° in latitude and longitude. The result shows that our method can capture the TIH anisotropic structure (Zettergren and Snively, 2019). In addition, the 2D projection result shows that the epicenter of the earthquake and the

place where the minimum TEC value is obtained are explicitly different. It is also found that the shape and region of the TIH region are almost the same as those of the initial tsunami shown by the simulation T. Saito et al. (2011).

3.3 Surface fitting with sparse data

Unlike the case of full data displayed in Fig. 5a, it can be clearly seen from Fig. 5b that the number of data points in the sparse-data case is not sufficient to entirely analyze the TIH, where 95 % of the GNSS receivers are randomly removed from the observed data. This situation is highly possible in areas where a dense network of GNSS satellites and receivers has not been installed, and hence the appropriate analysis for the TIH is almost impossible from the captured data. Nevertheless, our new analysis method can overcome this sparse-data problem effectively.

In Fig. 5d, the fitting surface and almost 5 % of captured data are depicted by a red surface and blue dots, respectively. Despite the extremely small number of data points, our method is successful in surface fitting and uncertainty evaluation.

A 2D projection of the fitting surface with sparse data and its 99 % CI, in Fig. 5f and h, shows that the location and range of the TIH are adequately estimated and consistent with those of the full-data case shown in Fig. 5e and g. To be more specific, our method is able to estimate the anisotropic TIH, which is consistent with the location and shape of the initial tsunami. The uncertainty naturally increases by comparing Fig. 5g with h, since the number of data points is smaller than that of the original data. These results demonstrate that our new method based on GP regression overcomes the sparse-data problem by implementing surface fitting that adequately estimated TEC variation with uncertainty and captured TIH shape.

In this study, surface fitting for the sparse data is performed using the data received by the remaining 5 % of receivers after randomly excluding 95 % of the GNSS receivers. With

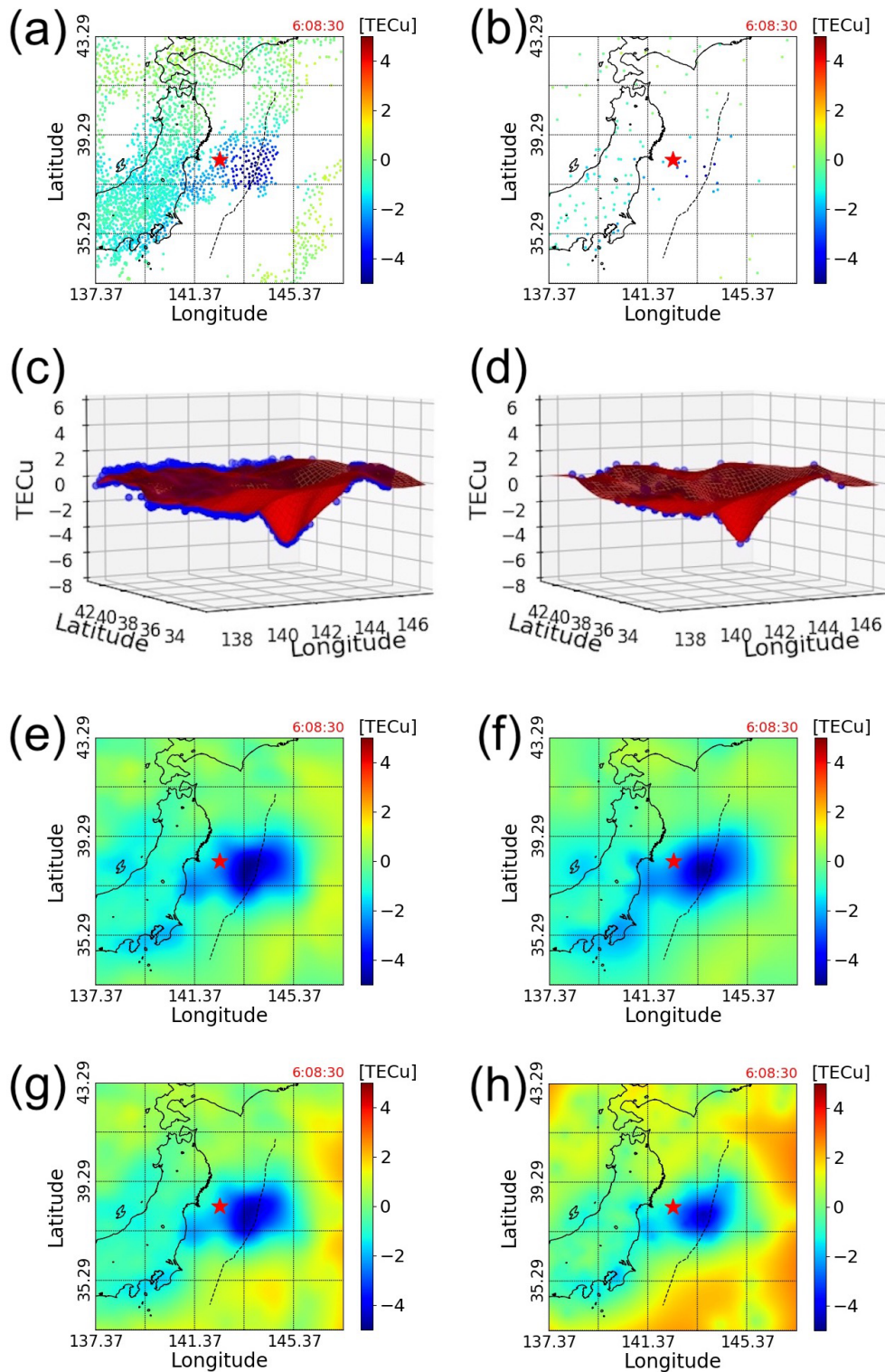


Figure 5. Left-hand side is for the full data; right-hand side is for the sparse data using only 5 % of the GEONET receivers. Panels (a) and (b) are measured TEC data. Panels (c) and (d) are measured TEC data (blue dots) and the fitting surface (red surface). Panels (e) and (f) are 2D projections of the fitting surface. Panels (g) and (h) are 2D projections of the 99 % one-sided CI of the fitting surface. The fitting surface is computed using the INLA-SPDE method. The black dashed line indicates the location of the axis of the Japan Trench.

Table 2. Table of the minimum TEC values and the receiver numbers that observed them at three different times.

	At 06:08:00		At 06:12:00		At 06:16:00	
	Minimum observed	Receiver number at which the minimum is observed	Minimum observed	Receiver number at which the minimum is observed	Minimum observed	Receiver number at which the minimum is observed
Full	−4.95	0043	−5.57	3011	−5.19	0589
Random 1	−4.95	0043	−5.41	0043	−4.89	0043
Random 2	−4.78	3007	−5.26	3007	−4.70	3007
Random 3	−4.73	0951	−5.17	0951	−4.94	0592
Random 4	−4.41	3016	−5.13	3016	−4.63	3016
Random 5	−4.63	3005	−5.49	3005	−5.04	3005
Random 6	−4.88	0950	−5.48	0950	−4.98	0950
Random 7	−4.87	3001	−5.43	3001	−4.94	0587
Random 8	−4.71	0587	−5.36	0587	−4.94	0587
Random 9	−4.22	0215	−4.74	0212	−4.14	0212
Random 10	−4.35	0582	−5.04	3023	−4.47	3032

such a small number of receivers, in theory it could happen that none of the 5 % of randomly chosen receivers would observe the data points that exist in the specific TIH region of great importance for the detection of the tsunami.

In our analysis, the minimum number of GNSS receivers within the target area detecting data from the satellite between 05:46:30 and 06:16:30, which is the period of analysis, is 832 (at 05:46:30), and 40 receivers were selected at random as a conservative estimate of 5 % of that number. Consider the aforementioned case where only receivers that do not measure the TIH are randomly selected to make up these 40 units. For example, at 06:12:00, there are 66 receivers receiving data with a TEC value of -4 or less and 25 receivers receiving data with a TEC value of -5 or less. At 06:12:00, there are 917 receivers detecting data from the satellite in the target area, and the probability that at least 1 of the 40 receivers randomly selected from these receivers contains 1 of the 66 receivers that receive -4 or less TEC value is around 95 %. Also, the probability that at least 1 of these 40 receivers contains 1 of the 25 receivers that receive -5 or less is approximately 70 %. Therefore, a random selection of receivers that cannot measure the TIH can happen, though the probability of that is very low. In this experiment, we analyze the usefulness of surface fitting in the sparse case, where there are receivers that measure the TIH.

The specific details of the situation of this experiment are described below: 10 experiments were conducted to randomly select 5 % of receivers, and the minimum observed values measured in each case are shown in Table 2. With a rare exception, the receiver number by which the minimum value is measured in each case is different. For example, at 06:12:00, in the case of random 9, the minimum value is about 0.8 TECU larger than the case with all data. Also, at 06:16:00, the minimum value of the full data is -5.19 , but there is only one case, random 5, where the minimum value is less than -5 .

As described above, we can see that the minimum value and distribution of TEC are different when we randomly choose receivers. And the following Fig. 6 shows that our fitting method works for such different sparse distributions of TEC. In the left half of Fig. 6, three sets of randomly chosen sparse data are drawn, indicating that the data points are distributed differently. On the right-hand side of Fig. 6, the 2D projections of the fitting surface for these sparse data are displayed, and it can be seen that the shape of the TIH is almost the same, though not exactly the same. Specifically, Fig. 6b2 shows a slightly wider TIH form than Fig. 6a2, and Fig. 6c2 has a TIH that is evenly spread in the east–west direction compared to Fig. 6a2 and b2. These results show that if we can measure a few observation points with reduced TEC, we can apply our fitting method to capture a TIH even if there are far fewer data points relative to the original data.

How the calculation of the TIH volume, i.e., the volume of the area where the value of TEC is less than 0 in the target area, is affected in this situation is analyzed in Fig. 7. The volume of the TIH is used as a measure to determine the effect. It can be seen from Fig. 7 that in each case, there are very few measurement points with a value of -4 or lower, and even the number of measurement points with a value of -3 or less is fewer than 10. However, when compared to the volumes calculated using all the data represented by the horizontally lines, it can be seen that across all cases the values are very close to one another, at each time of the three selected times. Furthermore, although the number of data points below -2 varies, the computed volumes are almost similar. Obviously, if there are no data points in the TIH area, volume calculation itself is impossible but is highly unlikely as we explained above. However, if data points shown in the figure are measured in the TIH, volume calculation is possible. Since a single receiver can receive data from multiple satellites, the number of receivers needed to receive data in the TIH is extremely small.

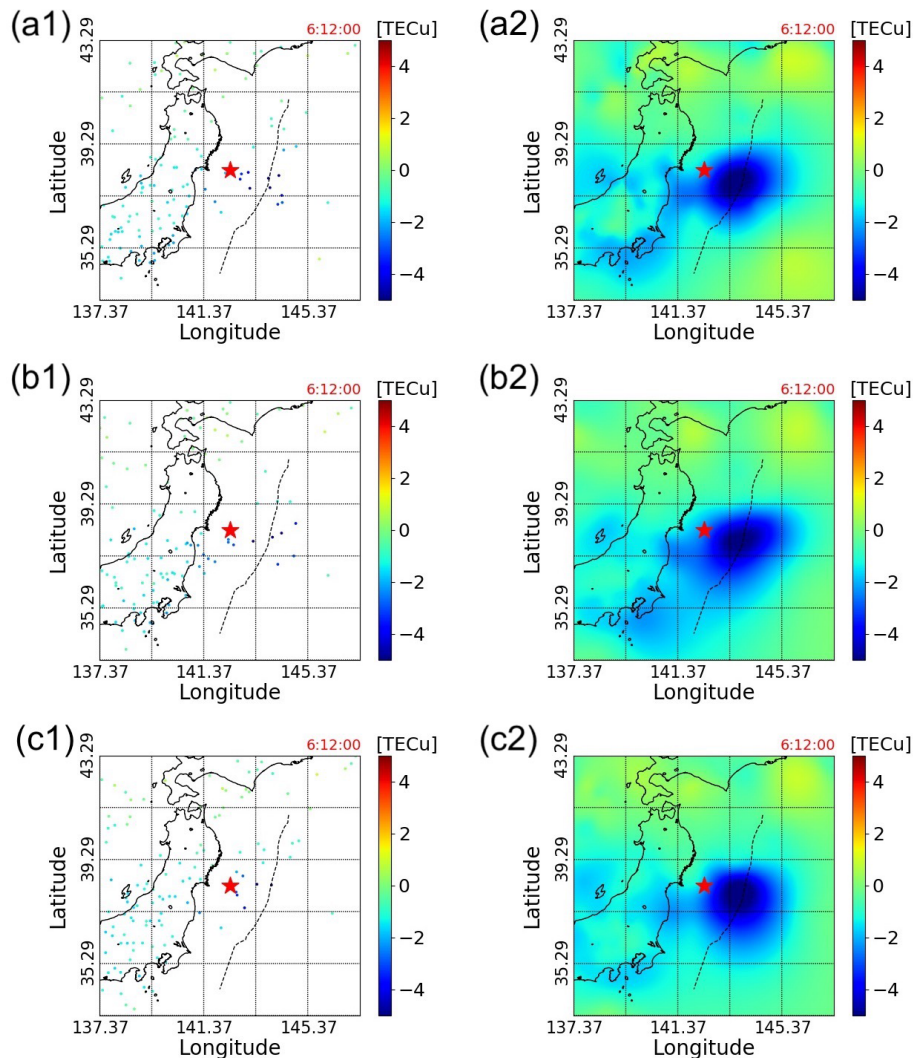


Figure 6. Three different sparse-data distributions, random 0, random 3, and random 8, and these fitting surfaces mapped in two dimensions at 06:12:00 (UTC). The black dashed line indicates the location of the axis of the Japan Trench.

3.4 TIH expansion

Since we are able to estimate all the TEC variations in the target area, we analyze the shape of the TIH in detail using the observed data. Figure 8 displays the distribution of TEC below different levels at different times. Specifically, in Fig. 8a1 to a3, the shapes of the TIH limited to TEC values of -2 or less are drawn at 06:08:00, 06:12:00, and 06:16:00 (UTC). Similarly, in Fig. 8b1 to b3 and c1 to c3, the shapes of TIHs with TEC values of -3 or less and -4 or less, respectively, are drawn. The tsunami source was set to 38° N, 143.4° E, which was calculated in a previous study (Kamogawa et al., 2016) as the average of tsunami sources obtained from previous research papers. According to Shinagawa et al. (2013), the initial tsunami reproduces the TIH, and the TIH is spread out over the initial tsunami. Then, Kamogawa et al. (2016) use the TEC values around the tsunami source area to de-

rive the relationship between the TIH and the initial tsunami. Based on these previous studies, using the observed data, the expansion of the TIH around the tsunami source was analyzed as follows.

As shown in Fig. 8a1 to a3, the region with TEC less than -2 can be seen to expand less to the northern direction than the other directions. In addition, a TEC decrease isolated from the TIH directly above the tsunami can be seen in the southwestern direction. As an overall trend, it can be said that the expansion of the TIH in the eastern and southwestern directions is more pronounced than in the other directions.

However, when we observed the behavior of Fig. 8 b1 to b3, where the value of TEC is less than -3 , we found that the expansion of the TIH is different from that of the TIH with a TEC value less than -2 described in Fig. 8a1 to a3. First of all, the area of the TIH has not expanded as much as that of the TIH with TEC less than -2 . In addition, the

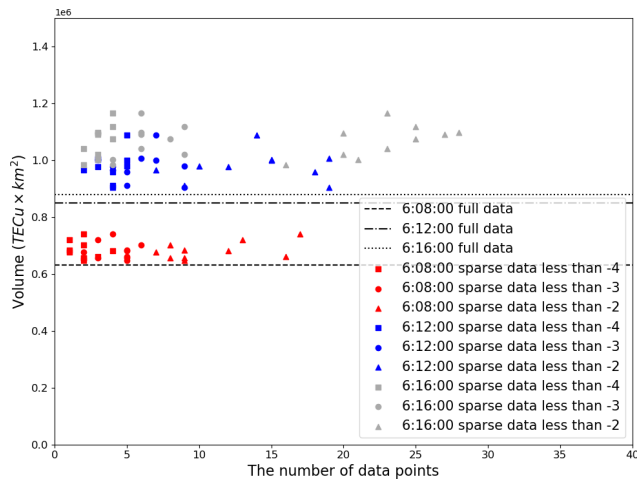


Figure 7. Comparison of the volume of the TIH calculated by sparse data (40 receivers) and the volume calculated using all data. Random choices are independently implemented 10 times. Points with square marks indicate the number of data points with a TEC value of -4 or less and the computed volume of the TIH; round marks and triangular marks indicate those of -3 or less and -2 or less, respectively. The red color shows the data at 06:08:00. Also, blue and dark grey are at 06:12:00 and 06:16:00, respectively. The horizontal lines show the volumes calculated using all the data at 06:08:00, 06:12:00, and 06:16:00.

southwestward expansion is not as large as that of the TIH in Fig. 8a1 to a3. For example, the eastward expansion appears to be the most pronounced, and also the northward expansion is less pronounced than in other directions, which are similar characteristics.

In the case of the TIH with TEC less than -4 , shown in Fig. 8c1 to c3, the westward expansion is smaller than the expansion in the other directions. In addition, the TIH center does not remain directly above the vicinity of the tsunami source but appears to be moving to the southeast. It can also be seen that the region where TEC is less than or equal to -4 does not expand significantly during the period of 20 to 30 min after the earthquake, but it does not shrink significantly either. In other words, when we focus on the TEC values below -4 , the shape of TEC is almost stable.

3.5 TIH expansion distance in each direction

From Fig. 8, it can be inferred that the time evolution of the shape of the region with small TEC variation and that with large TEC variation do not necessarily coincide. To investigate this point in more detail, we analyze the graphs with the distance in kilometers plotted by time for the most expanded points in the eight directions from the tsunami source. Here, the distance is computed by Hubeny's distance formula (Sato et al., 2017).

Here, the location of the tsunami source is the same value used in Kamogawa et al. (2016), and its coordinates are

calculated by referring to Maeda et al. (2011), Grilli et al. (2013), Ohta et al. (2012), and T. Saito et al. (2011). The eight directions are evenly divided into north, northwest, west, southwest, south, southeast, east, and northeast from its tsunami center position.

In Fig. 9, the distance between the most expanded point in the eight directions and the tsunami source when the TEC value is -2 or less (Fig. 9a), -3 or less (Fig. 9b), and -4 or less (Fig. 9c) are drawn, respectively. The eight directions and the color coding for each are the same as in Fig. 8.

In Fig. 9a, it can be seen that the expansion in the north-eastern direction progresses earlier than other directions. A little later, the eastward expansion continues to progress, and finally the eastward expansion progresses more than the northeastward expansion. The expansion to the south, after increasing for some reason, begins to decrease and then continues to expand again. Eventually, it will be as large as the eastward expansion. The westward expansion is a discontinuous movement due to the definition of distance and direction in this analysis. Specifically, as depicted in Fig. 8a2 and a3, the region with TEC less than -2 has a special shape in the western direction, and the distance in this direction tends to be more discontinuous than the distances in other directions. What is noteworthy is the expansion of the distance in the northern and northwestern directions. The progress of the distances in these directions is clearly smaller than those in other directions. For example, compared to the expansion in the eastern direction, the distance in the northern and northwestern directions is less than half. Also, it is worth mentioning the TIH withdrawal (05:59:30–06:01:00) in the southward direction. Due to the influence of the geomagnetic field, the plasma tends to move more to the south. In addition, the backward-moving average frequency filter applied in this study does not completely exclude the high-frequency component. For these reasons, we detected a decrease in the electron density as a result of the recombination caused by the oscillation of the plasma due to the high-frequency component on the southern side. This temporary decrease in electron density is caused by a different mechanism than the TIH formation, which is caused by the low-frequency component.

In Fig. 9b, the time evolution of the distances in all directions appears to be continuous. In this case, too, the expansion toward the northeast progresses in the initial stage, but then the expansion toward the east progresses significantly. The expansion in the southern direction can be said to be slower than that of the two directions mentioned above, but it eventually expands by the same distance as that in the eastern direction. The southeastern direction shows a trend similar to the expansion in the southern direction, and although the initial expansion is not so large, the final value is large. It can be seen that the northern and southwestern directions do not expand as much even at the beginning and then remain in a stable state with almost the same value over time. The northwestern direction, as with Fig. 9a, is one of the directions that does not expand the most. The most significant difference be-

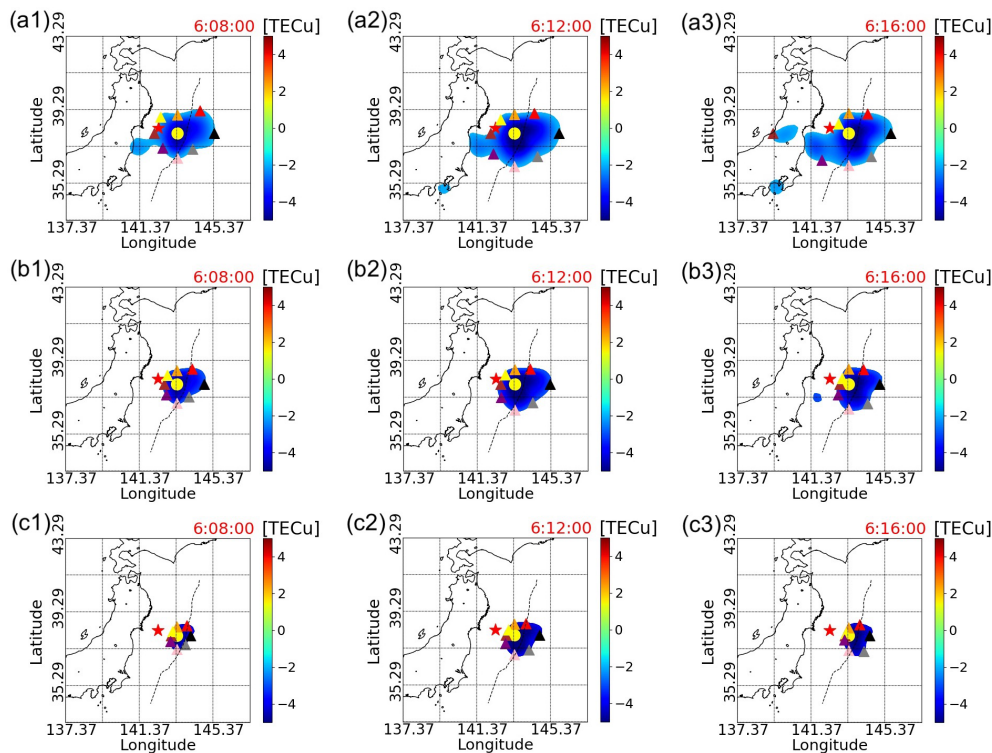


Figure 8. Expansion of the tsunami ionospheric hole. Panels (a1–a3) are plots of the TIH with TEC less than or equal to -2 . Panels (b1–b3) and (c1–c3) are less than or equal to -3 and -4 , respectively. These plots are at 06:08:00, 06:12:00, and 06:16:00 (UTC). The eight triangles are the south, north, west, east, northeast, southeast, northwest, and southwest directions from the tsunami source, respectively. The colors are pink, orange, brown, black, red, grey, yellow, and purple, respectively. The red-star mark is the location of the epicenter. The yellow-circle mark is the location of the estimated tsunami source. The black dashed line indicates the location of the axis of the Japan Trench.

tween Fig. 9b and a is the expansion in the western direction. In Fig. 9a, the distance in the western direction expands discontinuously and finally reaches the highest value. On the other hand, in Fig. 9b, the distance in the western direction is the smallest in almost all time periods.

In Fig. 9c, the time at which the image begins to expand is greatly delayed because the threshold is set even smaller than in Fig. 9a and b. Initially, the expansion in the northwestern and southern directions is significant, and as time passes, the expansion in the eastern direction becomes the largest. Eventually, the distance in the southeastern direction becomes larger than the distances in the other directions at the end of the period, but its maximum value is smaller than the value recorded by the eastern direction around 06:10:30 (UTC). The expansions in the northern and southwestern directions show almost the same development. For the northwestern and western directions, as in Fig. 9b, only the smallest expansion is shown during the period.

3.6 TIH overlapping with the initial tsunami

Unlike the high-frequency component of the TEC variation (Tsunami et al., 2011), the low-frequency component of the

TEC variation displays large drops in its values and remains fixed within a region, as shown in Fig. 8. Since this TEC reduction is caused by the initial tsunami, staying in the same location is theoretically correct. As shown in Fig. 10, the initial tsunami estimated by T. Saito et al. (2011) using inversion analysis mostly overlaps with the TIH where the TEC reduction is large. Although there have been studies to estimate the tsunami source using TEC fluctuation data, these studies estimate the tsunami source as a point (Liu et al., 2019). We show here for the first time that our method of estimating the entire TIH by GP regression can be used to estimate the initial-tsunami region as shown in Fig. 10. In the case of TEC values less than -3 shown in Fig. 10a1–a3, the TIH almost overlaps the initial-tsunami areas; in other words, the TIH with TEC values less than -3 is located in the region which is almost the same as the initial-tsunami region, while Fig. 10b1–b3 with TEC values less than -4 shows that the TIH stays within the initial-tsunami region.

Figure 11 shows the uncertainty in estimating the values of TEC using the full data. In this figure, the uncertainty is defined as 3 times the standard deviation. In general, interpolation between data points can be performed with a small uncertainty, while extrapolation has a larger uncertainty. Even

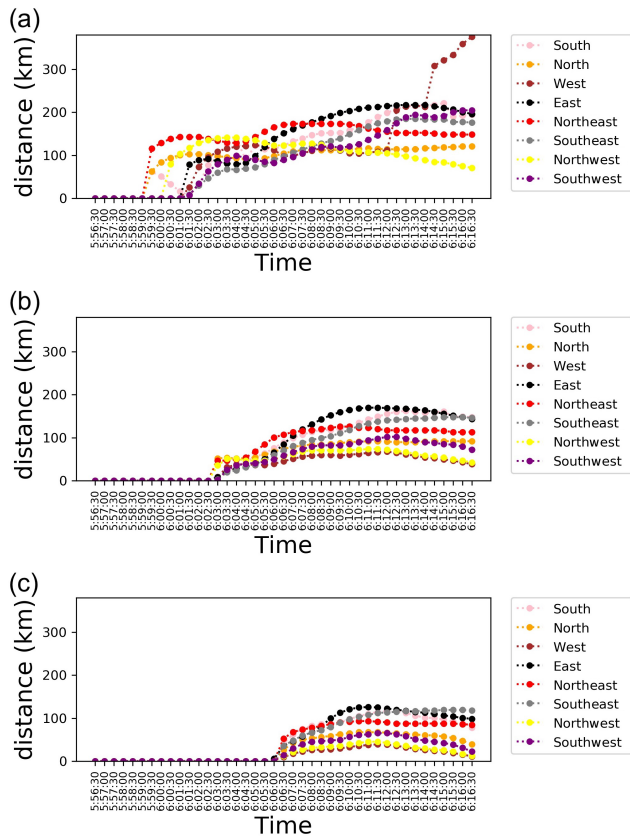


Figure 9. Expansion of the tsunami ionospheric hole. Panel (a) is the time series of the distance from the tsunami source in eight directions for the TIH with TEC less than or equal to -2 . Panels (b) and (c) are the time series of distance with TEC less than or equal to -3 and -4 , respectively.

in the case of interpolation, if the data points are sparse, the uncertainty will be large. Therefore, it can be seen from Fig. 11 that the uncertainty is larger in areas where the measurement points are sparse or where extrapolation is performed, as can be seen by comparing with Fig. 3.

In Fig. 10, the initial-tsunami and TIH regions overlap, but at 06:12:00 and 06:16:00, the TIH region extends more toward the eastern direction of the Pacific Ocean than the initial-tsunami region. Specifically, this tendency is observed in the area east of the initial-tsunami area and west of the dotted line at 145.37° E. In addition, the TIH region at 06:12:00 appears to be wider eastward of the Japan Trench than the TIH region at 06:16:00. In this regard, for example, looking at the uncertainty values for each region listed in Fig. 11, at 06:12:00, the uncertainty for the region east of the initial-tsunami region and west of the dotted line at 145.37° E is slightly larger. However, it is unlikely that this was the reason for capturing the non-overlapping phenomenon of the initial-tsunami and TIH regions.

Heki and Ping (2005) show that acoustic waves propagate upward, which are gradually refracted, and their effects

propagate horizontally in the ionosphere. According to this principle, the TIH is expected to spread evenly in the east–west direction of the tsunami generation area. Kakinami et al. (2012), who analyzed the measured data, showed that slant TEC decreased in the area east of the tsunami generation area after the 2011 Tohoku–Oki earthquake. The reason why the TIH calculated by our method appears to be extended to the east of the Japan Trench when compared to the initial-tsunami area is that the initial-tsunami height is highest on the Japan Trench. The acoustic waves from the highest region on the trench propagate into the atmosphere and affect the neutral atmosphere evenly in the east–west direction, resulting in the recombination of ions and electrons, causing the TIH to spread in the east–west direction around the trench. Therefore, we believe that the estimation by our method correctly captures the variation of the electron density. In future reverse calculations of the initial-tsunami area and height based on TIH information, it is necessary to take into account that the area with the highest initial tsunami has a large impact on TIH formation.

Also, the initial tsunamis estimated by other researchers (Ohta et al., 2012; Takagawa and Tomita, 2012) show that the initial-tsunami regions overlap the TIH region as in Fig. 12, where these initial tsunamis are roughly estimated by looking at them. In Ohta et al. (2012), they used an algorithm developed by Okada (1985) to compute the initial sea surface displacement based on their fault-determination procedure. On the other hand, in Takagawa and Tomita (2012), they investigated the effect of the rupture process on a tsunami source inversion. Estimated sea surface elevation of a tsunami source by their inversion method based on the assumption of finite rupture velocity of 2 km s^{-1} is shown. These results indicate that our TIH estimates overlap not only with the initial tsunami estimated by T. Saito et al. (2011) but also with the initial tsunami determined by other researchers.

3.7 Detailed comparison between TIH depth and initial-tsunami height

In this section, the relationship between the TIH and the initial tsunami is analyzed in more detail with the help of data provided by Tatsuhiko Saito, the first author of T. Saito et al. (2011). Based on the assumption that the TIH is triggered by tsunamis, it is expected that the locations of high tsunamis and high TEC reduction in the TIH are in close proximity to each other.

Figure 13a1 to a3 show the contour lines of the initial-tsunami height and the overlaid TIH that is plotted at three different time points. From these figures, it can be seen that the region with the largest decrease in TEC and the region with the highest initial-tsunami wave height are very close to each other. Furthermore, the decrease in TEC appears to correspond with the increase in the initial-tsunami height.

To make a more precise comparison in this regard, graphs comparing the initial tsunami and TIH at the latitude and lon-

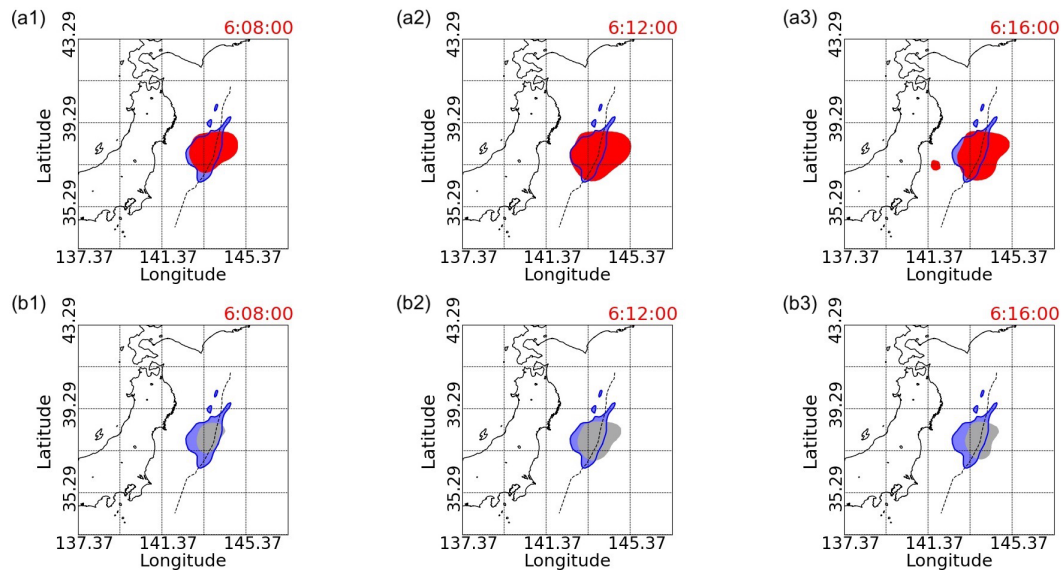


Figure 10. The initial tsunami and TIHs with TEC values less than -3 or -4 . The blue area is the simulated initial tsunami by inversion analysis with 130 small basis functions implemented by T. Saito et al. (2011). In panels (a1–a3), the red areas are TIHs with TEC less than -3 , while the dark grey areas in panels (b1–b3) illustrate TIHs with TEC less than -4 . Panels (a1) and (b1) are snapshots at 06:08:00 (UTC); panels (a2) and (b2) are at 06:12:00 (UTC); and panels (a3) and (b3) are at 06:16:00 (UTC). The black dashed line indicates the location of the axis of the Japan Trench.

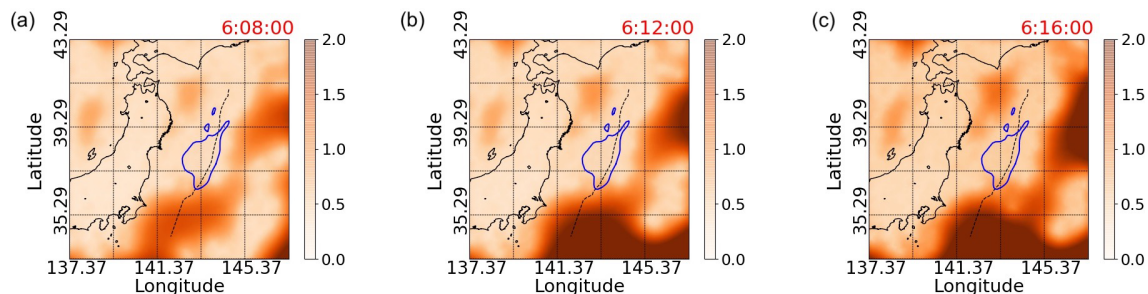


Figure 11. Uncertainty of the estimated TEC values at 06:08:00, 06:12:00, and 06:16:00. The uncertainty in this case is defined as 3 times the standard deviation. The area surrounded by the blue line is the simulated initial tsunami by inversion analysis with 130 small basis functions implemented by T. Saito et al. (2011). The black dashed line indicates the location of the axis of the Japan Trench.

gitude where the initial-tsunami height is the highest are illustrated in the second and third rows of Fig. 13, respectively. From the figures shown in the second line, it can be seen that the longitude with the largest decrease in TEC and the longitude with the highest initial-tsunami height roughly match. On the western side of the location where the initial tsunami is highest, the area where the tsunami height starts to increase and the area where TEC starts to decrease appear to be almost the same, but this is not necessarily the case on the eastern side.

From the figures shown in the third row, the latitudes with the largest decrease in TEC and the highest initial-tsunami height roughly matched, and unlike the east–west asymmetry when the latitude is fixed, the TEC decrease is generally sym-

metrical in the north–south direction. This symmetry seems to correspond to the shape of the initial tsunami.

Furthermore, the time series variation of TEC from 05:46:30 to 06:16:30 (UTC) is shown in Fig. 14, when fixed at the latitude and longitude with the highest initial-tsunami height, respectively. Figure 14a shows that the TEC decrease is the largest in the region where the initial-tsunami height is the highest when we observe the time series variation of TEC for 30 min after the earthquake. However, in the process of TEC decrease, it is also observed that there is a time period when TEC decreases the most in a slightly western region. In more detail, we can see that the area where TEC is decreasing the most is slightly moving from the west to the east and expanding with a larger decrease. In addition, there is an asymmetry of TEC decrease in the east–west direction.

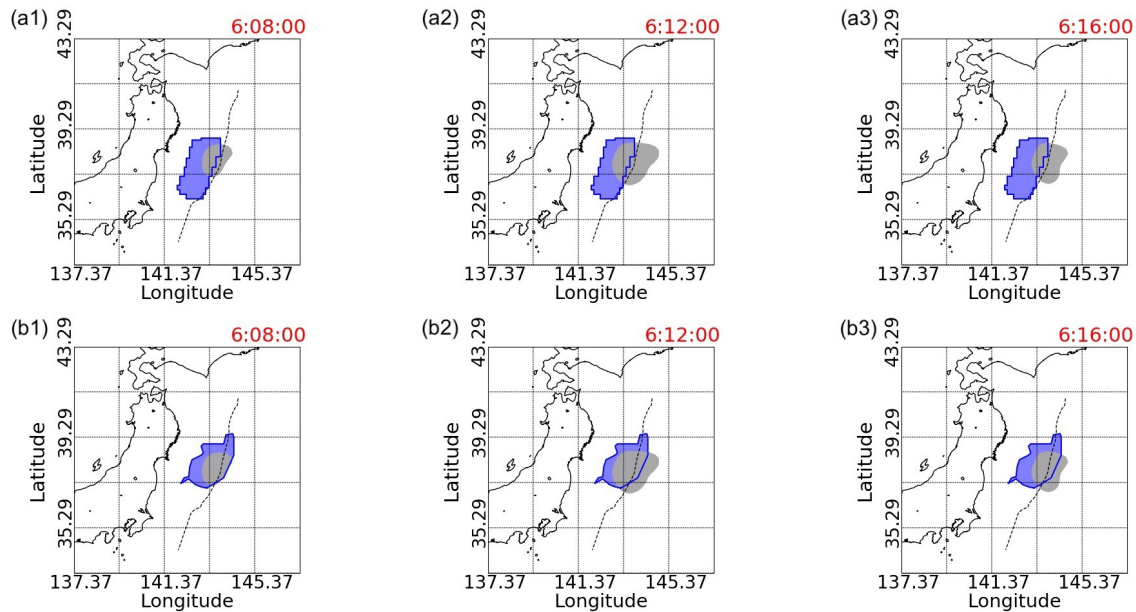


Figure 12. TIH and the initial-tsunami comparison. In the first row, the initial tsunami estimated by Ohta et al. (2012) is shown. In the second row, the initial tsunami estimated by Takagawa and Tomita (2012) is shown. In both cases, the TIHs with TEC values less than -4 are described.

On the other hand, from Fig. 14b, the TEC decrease appears to be symmetrical in the north–south direction when the longitude is fixed at the position where the initial tsunami is highest. However, by checking the details of the time series change of TEC decrease, we can see that the position where the TEC decreases the most is moving from north to south, and the decrease is expanding.

3.8 TIH volume computation

Figure 15 shows the time series of the TIH volume, which is computed by the trapezoidal quadrature method for the region where TEC estimated by surface fitting has a negative value. In other words, the volume between the flat surface, where the TEC values are equal to 0, and the fitting surface, which has a negative value, is calculated.

The main effect by acoustic waves induced by the initial tsunami is the reduction of TEC in the ionosphere by moving the plasma along the magnetic field and causing recombination. More specifically, although there are regions where TEC increases due to complex physical mechanisms, the magnitude of the initial tsunami can be assessed by focusing on the decrease in TEC. Therefore, the volume of the region with a negative TEC value is considered to be related to the magnitude of the initial tsunami.

In Fig. 15a, surface fitting is implemented with simple GP regression, while the INLA-SPDE method is applied to the GP regression process in Fig. 15b. The solid lines in Fig. 15a and b display the TIH volumes computed for the full data, and the dashed lines are for the sparse data. In the case of

the sparse data, we repeat 10 times the random selection of 5 % of the receivers and the GP regression to fit the surfaces and then calculate the average value using each computed volume. This iterative process is intended to exclude a possible influence of the random seed used in the sparse-data selection on the results. As shown in Fig. 3, the measurement points are moving and not uniformly distributed in the targeting range. Still, the time series of the computed TIH volumes looks continuous, as shown in Fig. 15a and b.

The volume of the TIH begins to increase almost 10 min after the earthquake occurrence and continues to increase until about 28 min after the earthquake shown in Fig. 15a and b. More specifically, the reason the volume of the TIH continues to increase until 28 min after the earthquake is that the TIH is formed by the extremely low-frequency component of acoustic waves that travel from the sea surface to the ionosphere. The recombination of plasma and electrons caused by this low-frequency component is the direct cause of the TIH, and this recombination takes time, so the volume continues to increase over a long period of time. Simulations of TIH formation using a physical model presented by Shinagawa et al. (2013) also show that the TIH forms over a similar time period. The past study of actual post-earthquake TEC variation data has confirmed that the percentages of TEC depressions for huge simulated tsunamis reach almost 40 times as large as those for smaller simulated tsunamis (Kamogawa et al., 2016). In this context, a huge simulated tsunami means a tsunami with a simulated initial-tsunami height of around 5 m, and a small simulated tsunami means a tsunami with a simulated initial-tsunami height of 1 m or less in their study.

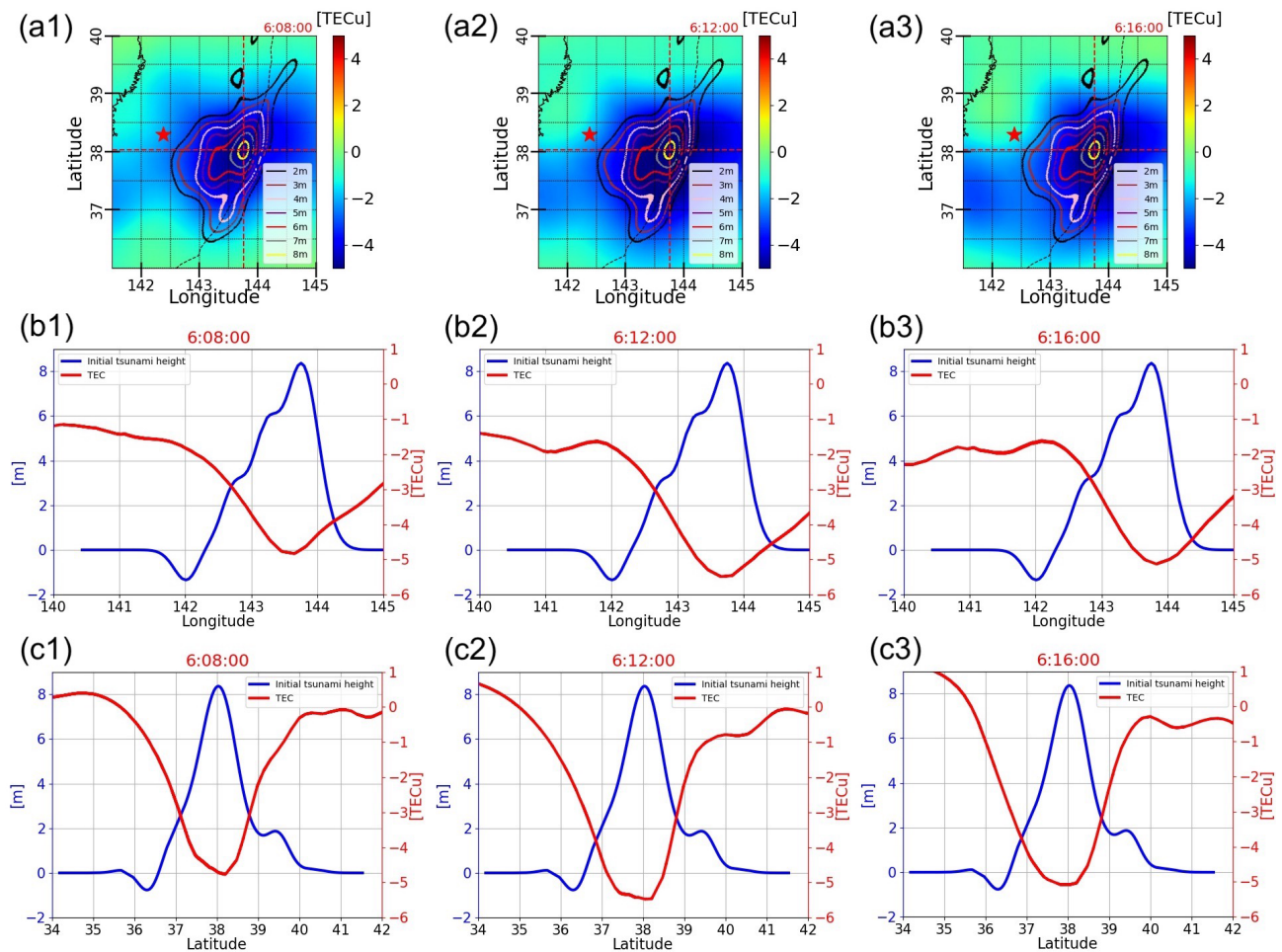


Figure 13. Overlap of initial tsunami and TEC. Panels (a1–a3) are plots of TEC values and contours of the initial-tsunami height by T. Saito et al. (2011). The horizontal and vertical red dashed lines indicate the latitude and longitude where the initial tsunami reaches its highest height. Panels (b1–b3) and (c1–c3) show the initial-tsunami height and TEC values at the latitude and longitude where the initial-tsunami height is the highest, respectively. The initial-tsunami height corresponds to the scale on the left axis (blue), and the TEC value corresponds to the scale on the right (red).

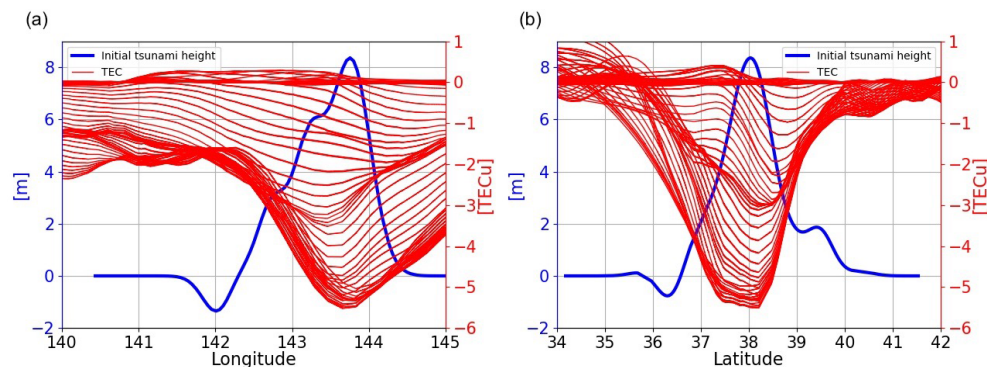


Figure 14. Initial-tsunami height and time series variation of TEC from 05:46:30 to 06:16:30 (UTC). Panel (a) shows the TEC time series when the initial-tsunami height is fixed at the highest latitude, and panel (b) shows the TEC time series when the longitude is fixed. The initial-tsunami height corresponds to the scale on the left axis (blue), and the TEC value corresponds to the scale on the right (red).

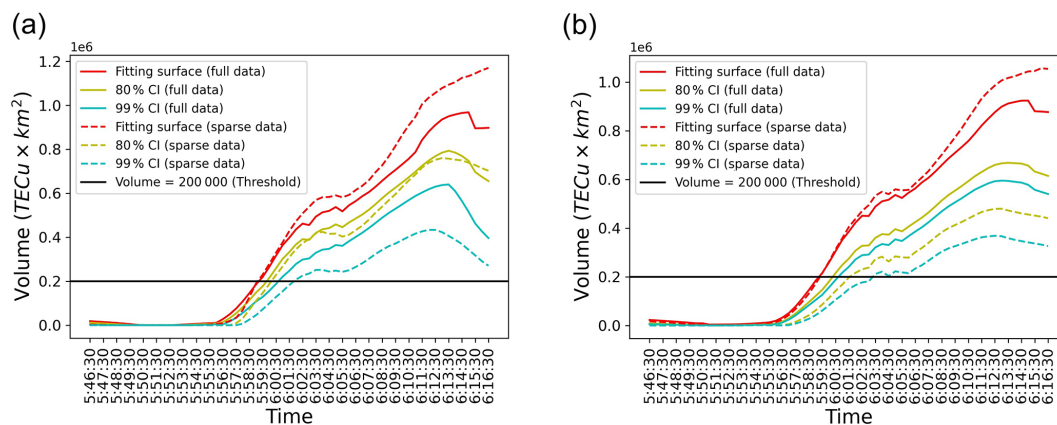


Figure 15. The time series of TIH volume for full data and sparse data with a one-sided CI. The red solid line is the volume calculated with a fitting surface for full data. The solid yellow line and the solid blue line are a one-sided 80 % CI and 99 % CI, respectively. The dashed lines are for sparse data with only 5 % of receivers. The horizontal black line is a provisional threshold. Panel (a) shows the TEC surface computed based on GP regression. Panel (b) shows the TEC surface computed based on GP regression using the INLA-SPDE method.

Although it is impossible to understand the time evolution of TIH volume in advance, since no study has considered and analyzed the entire TEC variation in the targeting range under the extremely complex TIH mechanism and our new method is applied to only one tsunami case, it is a sufficiently conservative estimate that the threshold is set to about 10 %–20 % of the maximum TIH volume according to the aforementioned previous study. Therefore, if the volume exceeds this threshold, it is reasonable to draw the conclusion that a huge tsunami is being generated. Though it is impossible to describe what the definition of a huge tsunami in an explicit and strict manner is, since there is no general definition of a huge tsunami, a huge tsunami in this context means a tsunami with the same magnitude as the 2011 Tohoku-Oki earthquake. By convention, tsunamis triggered by a magnitude of about nine earthquakes are considered to be huge tsunamis. Also, in this analysis, a provisional threshold is set to 200 000 TECU km².

In the case of the full data, both panels show that the volumes calculated from the fitting surface (but not accounting for uncertainties in the approximation) reach the threshold 1 and 2 min earlier, respectively, than the volumes of 80 % and 99 % CI. Similarly, in the case of the sparse data, the time difference is about 2 and 4 min, respectively, to reach the threshold for the volumes computed from the fitting surface and both CIs. It means that thanks to our uncertainty computations, making sure that a warning is at a high level of confidence, based on data, of either 80 % or 99 % results in delays for advisories of only, respectively, 1–2 or 2–4 min.

The warning system based on this method is highly feasible because surface fitting and the estimation of the TEC values for the full data can be processed in less than 1 min based on the INLA-SPDE method. However, in the case of the sparse-data fitting, our implementation of the INLA-SPDE method sometimes fails due to the geometric meshing opti-

mized for larger data sets, where the benefit of this method is. Nevertheless, the robustness and feasibility of this method never deteriorate because it is possible to compute the surface and estimated values in less than 10 s for the sparse-data case based on the standard GP regression method. Our method is the first to demonstrate that we can calculate the volume of TIHs accurately in real time and use it as a measure of TIHs even when only a limited number of measurement points are available.

4 Conclusions

In this paper, we compute the volume of the ionospheric depression generated by a tsunami, in real time and with enough confidence to issue warnings. The surface fits the TEC data using a Gaussian process regression after removing outliers. It enables us to estimate the TEC values over the entire target area. Furthermore, uncertainty can be properly evaluated for the estimated values of TEC according to the density of observations.

The TIH captured by our method is located east of the epicenter. This is consistent with the initial tsunami estimated by other research groups being east of the epicenter (T. Saito et al., 2011; Ohta et al., 2012; Takagawa and Tomita, 2012). Also, the estimated TIH almost overlaps with the initial-tsunami area estimated by three different research groups. In the ionosphere, the anisotropic conductance and geomagnetic-field directions theoretically cause ionospheric currents to have complex shapes (Zettergren and Snively, 2019). We concretely show here that the estimated TIH can be anisotropic using observed TEC data and a statistical approach.

As shown in our results, this new method is robust, as it works in situations where measurements are not uniformly distributed and moving and TIHs display anisotropy and even

if the number of observed data points is sparse. Since our estimates of the shape of the anisotropic TIHs reflect the signature of the initial-tsunami wave, we demonstrate that using one specific data point such as the minimum observed value as a scale of a TIH (Kamogawa et al., 2016) is insufficient to characterize the initial wave. Our computation of the volume of TIHs as a measure to assess the scale of TIHs fully takes into account the spatial variations of the TEC depression generated by the tsunami over the domain, including any anisotropy.

In addition, although there have been papers referring to a TIH based on observational data (Kakinami et al., 2012; Kamogawa et al., 2016), it was impossible to give a detailed explanation of the temporal variation of the TIH in those papers due to data limitations. Using our method, however, the detailed TIH expansion analysis based on different thresholds becomes possible. One of the findings is that the way in which the region with small TEC variation expands is dramatically different from that with large TEC variation expands. In our study, it is confirmed that the northward expansion is smaller than the southward expansion, no matter at which threshold level the TIH expansion is checked. This is consistent with the outcomes in previous studies (Heki and Ping, 2005) that in the Northern Hemisphere, the interaction of the geomagnetic field with the movement of charged particles in acoustic waves may have attenuated northward propagation. This result is also consistent with the simulated results of previous studies using 3D simulations (Zettergren et al., 2017; Zettergren and Snively, 2019), where the TEC decrease in the southern direction was larger than that in the northern direction. While the results of the 2D simulations did not show a relatively large decrease in TEC to the south (Shinagawa et al., 2013), the 3D simulations did (Zettergren et al., 2017; Zettergren and Snively, 2019), and we have demonstrated this through statistical analysis based on satellite captured data.

As for the high-frequency component of the TEC variability, past studies (A. Saito et al., 2011) analyzing observational data have confirmed that the eastward propagation of the TEC fluctuation is faster than the westward propagation. Although our analysis is focused on the low-frequency component, we have confirmed for the first time that the westward expansion of the TIH with TEC less than -3 , estimated by this new method, is less rapid than the eastward expansion. However, no similar simulation results have been reported so far for the asymmetry in the east–west direction found in our analysis based on the measured data. Furthermore, the westward expansion observed for TEC values below -2 cannot be satisfactorily explained by the relevant previous studies, and further detailed analysis is needed. Previous papers based on observation data without imposing frequency filters mention that the TIH stops just above the tsunami source, but a more detailed analysis in our study shows that the TIH with a threshold of TEC variation below -2 expands over time. The TIH separated by each of

these thresholds overlaps with the initial-tsunami region calculated by other research groups (T. Saito et al., 2011; Ohta et al., 2012; Takagawa and Tomita, 2012). In the previous study using TEC observational data (Liu et al., 2020), the tsunami source is analyzed, but the result is a point estimation. On the other hand, our method can estimate the tsunami region because it is possible to cover all the targeting area. More significantly, the time series change of the TIH fixed at the latitude and longitude where the initial tsunami is the highest shows for the first time that the TIH reflects the shape of the initial tsunami at that fixed latitude and longitude. This is a fact that could not be found in the previous analysis that used only observation data. From this fact, it is expected that the TIH information can be used to estimate not only the initial-tsunami area but also its shape.

We also believe that our method can estimate the initial tsunami independently from previous methods. Previous methods include, for example, the method of estimating a tsunami from the source and magnitude of an earthquake. By detecting seismic waves at multiple observation points, it is possible to calculate the approximate location of the earthquake and to some extent the exact size and magnitude of the earthquake within 2 min. Based on this information, the initial tsunami can be estimated.

In addition, a method has been developed to calculate the initial shape of a tsunami in reverse by calculating the shift of vessel speed due to the passage of a tsunami from the data of the automatic identification system (AIS), which is required to be installed on ships sailing in the area. This method is expected to be used to predict the initial tsunami around the world. However, some problems exist, such as the fact that the exact magnitude of a huge earthquake with a magnitude greater than 8 is not immediately known, the fault displacement (estimated from seismic waves) does not always match the initial sea level change, and the initial sea level change cannot be known.

In addition to the above methods, Japan has developed a system called the REAL-time GEONET Analysis system for Rapid Deformation monitoring (REGARD), which analyzes GEONET data in real time and extracts crustal deformation during earthquakes to automatically estimate fault models and earthquake scale within 3 min after the earthquake. Moreover, the Seafloor Observation Network for Earthquakes and Tsunamis along the Japan Trench (S-NET) has been established on the seafloor from off of Boso to off of Tokachi. The data are collected in real time 24 h every day. This system is expected to be used to accurately predict the initial tsunami. As a matter of course, it is difficult to accurately estimate the initial-tsunami information because no estimation method is perfect and the area covered may be limited.

Therefore, in addition to the various existing initial-tsunami estimation methods mentioned above, if the initial-tsunami shape and height estimation based on our developed TIH estimation can be realized in the future, it is ex-

pected that the combination of these methods will enable us to realize an even more accurate estimation of initial-tsunami height and range. From this point of view, even if it takes about 20 min to obtain the initial-tsunami information, we believe it is beneficial.

Furthermore, the existence of the second and third waves can be estimated by looking at the time variation of the TIH. Since the presence of large second and third waves affects the shape of the TIH, it can be used to determine whether the tsunami warning should be maintained or canceled after it is issued. In fact, after the 2011 earthquake in Japan, it took 1.5 d for the tsunami warning to be lifted. This is an advantage of our method, even if it takes time to obtain useful information.

As for the method to obtain the initial-tsunami information by calculating an inverse from the TIH information, we expect that the combination of the TIH estimated by our method and the acoustic wave propagation model may be able to estimate the initial-tsunami area. At present, although Shinagawa et al. (2013) and Zettergren and Snively (2019) have been able to reproduce the TIH quite accurately, they have not yet been able to reproduce it completely. Therefore, we expect to be able to back-calculate the region of the initial tsunami at an early stage by supplementing the simulation model with model discrepancies (Brynjarsdóttir and O'Hagan, 2014), which is a statistical method that takes into account the differences between actual measurements and model outputs for estimation and history matching (Vernon et al., 2014), which can limit the likelihood region of model parameters. There have been attempts to construct tsunami early warnings using TEC variations. Liu et al. (2019) demonstrate that the location of the tsunami source can be estimated from 10 IPs by observing TEC variations with a high-pass filter using methods that consider the propagation speed such as the circle method, the ray-tracing technique, and the beam-forming technique. However, these methods only identify the tsunami source with uncertainty and do not take into account the scale and range of the initial tsunami. As larger initial tsunamis cause larger decreases in TEC (Astafyeva et al., 2013; Kamogawa et al., 2016), if a TIH volume reaches a certain threshold, then it indicates that a large-scale initial tsunami has occurred. For example, in Fig. 3 of Kamogawa et al. (2016), we can see the positive correlation between the percentage of TEC depression and simulated initial-tsunami height. To relate our measure, that is, the volume of the TIH, to a corresponding measure of the tsunami size, it is necessary to apply the method to other tsunami cases in future work. As a result, we expect to be able to derive a detailed relationship between the volume of the TIH and the initial tsunami. Therefore, using our method, it is possible to build an early warning system that issues a tsunami warning when the volume of the TIH exceeds a certain threshold, taking uncertainty into consideration. In our analysis, we set a provisional threshold at 200 000 TECU km², and it is clear that the volumes calcu-

lated using both full data and sparse data exceed the threshold within 15 min after the earthquake occurrence or sooner with a lower threshold. Even carrying out the computations in the most exacting case, using 99 % CI and sparse data (5 % of the total observations) only delays the warning by around 4 min. We anticipate that more numerical work, more physical understanding of possible natural levels of TEC variations, and more data analysis will be required to establish more finely the thresholds at which advisories can be issued and thus shorten the advisories to possibly 10 min or so. Although, in some cases, tsunamis reach the coast very fast, to apply our method there must be a minimum window of almost 10 min between generation and arrival. However, this is perfectly valid for tsunami hazard assessment over populous regions with larger arrival times, as for example tsunami hazard assessment in the city of Victoria, British Columbia, from a tsunami generated in the Cascadia subduction zone (Salmanidou et al., 2021). Our implementation on the 2011 Tohoku-Oki earthquake in Japan demonstrates that our method works well there. Hence it is very likely that this method can be applied to tsunamis around the world, caused by any kind of sources. This may enable the construction of a robust worldwide tsunami early warning system using the volume of TIHs as an index.

Code and data availability. GPS data were provided by the Geospatial Information Authority of Japan. Currently, the data can be purchased from the Japan Association of Surveyors at <http://www.jsurvey.jp/eng.htm> (The Japanese government, 2022). Several software packages are used for the data preprocessing. These software can be downloaded from the following URLs.

1. <https://terras.gsi.go.jp/ja/crx2rxn.html> (Hatanaka, 2022)
2. <http://www.ep.sci.hokudai.ac.jp/~heki/software.htm> (Heki, 2022).

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/nhess-22-849-2022-supplement>.

Author contributions. RK conceptualized the project and methodology, curated the data, performed the formal analysis and investigation, created the visualizations, and wrote the original draft of the paper. MK acquired project funding, curated the data, and worked with the software. TN acquired project funding, curated the data, worked with the software, and reviewed and edited the paper. AS supervised the project, conceptualized the methodology, and reviewed and edited the paper. SG conceptualized and supervised the project, acquired project funding, performed the investigation, and reviewed and edited the paper.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The authors acknowledge the use of the UCL Myriad High Performance Computing Facility (Myriad@UCL), as well as associated support services, in the completion of this work. Ryuichi Kanai is supported by the Japan Student Services Organization. This research was partly supported by the Ministry of Education, Culture, Sports, Science and Technology through a Grant-in-Aid for Scientific Research (B, no. 17H02058, 2017–2020, Masashi Kamogawa) and Earthquake Research Institute (University of Tokyo) cooperative research program (Masashi Kamogawa and Ryuichi Kanai). Serge Guillas was supported by the projects “Uncertainty Quantification of complex computer models. Applications to tsunami and climate” (EPSRC grant no. EP/N510129/1) and “Real-time Advanced Data assimilation for Digital Simulation of Numerical Twins on HPC” (EPSRC grant no. EP/T001569/1) of The Alan Turing Institute. Masashi Kamogawa was supported by Adaptable and Seamless Technology transfer Program through Target-driven R & D (A-STEP) from Japan Science and Technology Agency (JST) (grant no. JPMJTM19YG). The authors are grateful to Tatsuhiko Saito, the first author of T. Saito et al. (2011), for sharing his data of the estimated initial tsunami in his paper. In creating Figs. 10, 11, 13, and 14, we use his data to show the contour and the shape of the initial tsunami. Kosuke Heki (Heki, 2022) at Hokkaido University kindly provided us with GNSS data conversion software. We also appreciate the Geospatial Information Authority of Japan, which provided GPS data conversion software.

Financial support. This research has been supported by the Ministry of Education, Culture, Sports, Science and Technology (grant no. 17H02058); the Engineering and Physical Sciences Research Council (grant nos. EP/N510129/1 and EP/T001569/1); Tokai University (grant no. IORD2018-01); and the Adaptable and Seamless Technology Transfer Program through Target-driven R&D (grant no. JPMJTM19YG).

Review statement. This paper was edited by Maria Ana Baptista and reviewed by two anonymous referees.

References

- Amato, A., Avallone, A., Basili, R., Bernardi, F., Brizuela, B., Graziani, L., Herrero, A., Lorenzino, M. C., Lorito, S., Mele, F. M., Michelini, A., Piatanesi, A., Pintore, S., Romano, F., Selva, J., Stramondo, S., Tonini, R., and Volpe, M.: From Seismic Monitoring to Tsunami Warning in the Mediterranean Sea, *Seismol. Res. Lett.*, 92, 1796–1816, <https://doi.org/10.1785/0220200437>, 2021.
- Astafyeva, E., Shalimov, S., Olshanskaya, E., and Lognonné, P.: Ionospheric response to earthquakes of different magnitudes: Larger quakes perturb the ionosphere stronger and longer, *Geophys. Res. Lett.*, 40, 1675–1681, <https://doi.org/10.1002/grl.50398>, 2013.
- Bernard, E. and Titov, V.: Evolution of tsunami warning systems and products, *Philos. T. R. Soc. A*, 373, 20140371, <https://doi.org/10.1098/rsta.2014.0371>, 2015.
- Brynjarsdóttir, J. and O'Hagan, A.: Learning about physical parameters: The importance of model discrepancy, *Inverse Probl.*, 30, 114007, <https://doi.org/10.1088/0266-5611/30/11/114007>, 2014.
- Chang, K.-L., Guillas, S., and Fioletov, V. E.: Spatial mapping of ground-based observations of total ozone, *Atmos. Meas. Tech.*, 8, 4487–4505, <https://doi.org/10.5194/amt-8-4487-2015>, 2015.
- Cover, T. and Hart, P.: Nearest neighbor pattern classification, *IEEE T. Inform. Theory*, 13, 21–27, <https://doi.org/10.1109/TIT.1967.1053964>, 1967.
- Giles, D., Gopinathan, D., Guillas, S., and Dias, F.: Faster than real time tsunami warning with associated hazard uncertainties, *Front. Earth Sci.*, 8, 560, <https://doi.org/10.3389/feart.2020.597865>, 2021.
- Grilli, S. T., Harris, J. C., Bakhsh, T. S. T., Masterlark, T. L., Kyriakopoulos, C., Kirby, J. T., and Shi, F.: Numerical simulation of the 2011 Tohoku tsunami based on a new transient FEM co-seismic source: Comparison to far-and near-field observations, *Pure Appl. Geophys.*, 170, 1333–1359, <https://doi.org/10.1007/s00024-012-0528-y>, 2013.
- Hatanaka, Y.: RNXCMP software, Geospatial Information Authority of Japan [code], <https://terras.gsi.go.jp/ja/crx2rn.html>, last access: 8 March 2022.
- Heki, K.: L4 generation from RINEX observation data file and Satellite position calculation from RINEX orbit data file in Earth-Centered Earth-Fixed system [code], <http://www.ep.sci.hokudai.ac.jp/~heki/software.htm>, last access: 8 March 2022.
- Heki, K. and Ping, J.: Directivity and apparent velocity of the coseismic ionospheric disturbances observed with a dense GPS array, *Earth Planet. Sc. Lett.*, 236, 845–855, <https://doi.org/10.1016/j.epsl.2005.06.010>, 2005.
- Kakinami, Y., Kamogawa, M., Tanioka, Y., Watanabe, S., Gusman, A. R., Liu, J.-Y., Watanabe, Y., and Mogi, T.: Tsunami-genic ionospheric hole, *Geophys. Res. Lett.*, 39, L00G27, <https://doi.org/10.1029/2011GL050159>, 2012.
- Kamogawa, M., Orihara, Y., Tsurudome, C., Tomida, Y., Kanaya, T., Ikeda, D., Gusman, A. R., Kakinami, Y., Liu, J.-Y., and Toyoda, A.: A possible space-based tsunami early warning system using observations of the tsunami ionospheric hole, *Sci. Rep.-UK*, 6, 37989, <https://doi.org/10.1038/srep37989>, 2016.
- Lay, T., Ammon, C. J., Kanamori, H., Yamazaki, Y., Cheung, K. F., and Hutko, A. R.: The 25 October 2010 Mentawai tsunami earthquake (M_w 7.8) and the tsunami hazard presented by shallow megathrust ruptures, *Geophys. Res. Lett.*, 38, 6, <https://doi.org/10.1029/2010GL046552>, 2011.
- Lindgren, F., Rue, H., and Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, *J. R. Stat. Soc. B*, 73, 423–498, <https://doi.org/10.1111/j.1467-9868.2011.00777.x>, 2011.
- Liu, J.-Y., Lin, C.-Y., Tsai, Y.-L., Liu, T.-C., Hattori, K., Sun, Y.-Y., and Wu, T.-R.: Ionospheric GNSS total electron content for tsunami warning, *J. Earthq. Tsunami*, 13, 1941007, <https://doi.org/10.1142/S1793431119410070>, 2019.
- Liu, J.-Y., Lin, C.-Y., Chen, Y.-I., Wu, T.-R., Chung, M.-J., Liu, T.-C., Tsai, Y.-L., Chang, L. C., Chao, C.-K., Ouzounov, D., and Hattori, K.: The source detection of 28 Septem-

- ber 2018 Sulawesi tsunami by using ionospheric GNSS total electron content disturbance, *Geoscience Letters*, 7, 1–7, <https://doi.org/10.1186/s40562-020-00160-w>, 2020.
- Lorito, S., Romano, F., Atzori, S., Tong, X., Avallone, A., McCloskey, J., Cocco, M., Boschi, E., and Piatanesi, A.: Limited overlap between the seismic gap and coseismic slip of the great 2010 Chile earthquake, *Nat. Geosci.*, 4, 173–177, <https://doi.org/10.1038/ngeo1073>, 2011.
- Maeda, T., Furumura, T., Sakai, S., and Shinohara, M.: Significant tsunami observed at ocean-bottom pressure gauges during the 2011 off the Pacific coast of Tohoku Earthquake, *Earth Planets Space*, 63, 53, <https://doi.org/10.5047/eps.2011.06.005>, 2011.
- Maruyama, T., Tsugawa, T., Kato, H., Saito, A., Otsuka, Y., and Nishioka, M.: Ionospheric multiple stratifications and irregularities induced by the 2011 off the Pacific coast of Tohoku Earthquake, *Earth Planets Space*, 63, 65, <https://doi.org/10.5047/eps.2011.06.008>, <https://doi.org/10.5047/eps.2011.06.008>, 2011.
- Ohta, Y., Kobayashi, T., Tsushima, H., Miura, S., Hino, R., Takasu, T., Fujimoto, H., Iinuma, T., Tachibana, K., Demachi, T., Sato, T., Ohzono, M., and Umino, N.: Quasi real-time fault model estimation for near-field tsunami forecasting based on RTK-GPS analysis: Application to the 2011 Tohoku-Oki earthquake (M_w 9.0), *J. Geophys. Res.-Sol. Ea.*, 117, B02311, <https://doi.org/10.1029/2011JB008750>, 2012.
- Okada, Y.: Surface deformation due to shear and tensile faults in a half-space, *B. Seismol. Soc. Am.*, 75, 1135–1154, <https://doi.org/10.1785/BSSA0750041135>, 1985.
- Rasmussen, C. E. and Williams, C. K. I.: *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, <https://mitpress.mit.edu/books/gaussian-processes-machine-learning> (last access: 8 March 2021), 2006.
- Rue, H., Martino, S., and Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *J. R. Stat. Soc. B*, 71, 319–392, <https://doi.org/10.1111/j.1467-9868.2008.00700.x>, 2009.
- Saito, A., Tsugawa, T., Otsuka, Y., Nishioka, M., Iyemori, T., Matsumura, M., Saito, S. and Chen, C. H. and Goi, Y., and Choosakul, N.: Acoustic resonance and plasma depletion detected by GPS total electron content observation after the 2011 off the Pacific coast of Tohoku Earthquake, *Earth Planets Space*, 63, 863–867, <https://doi.org/10.5047/eps.2011.06.034>, 2011.
- Saito, T., Ito, Y., Inazu, D., and Hino, R.: Tsunami source of the 2011 Tohoku-Oki earthquake, Japan: Inversion analysis based on dispersive tsunami simulations, *Geophys. Res. Lett.*, 38, L00G19, <https://doi.org/10.1029/2011GL049089>, 2011.
- Salmanidou, D. M., Beck, J., Pazak, P., and Guillas, S.: Probabilistic, high-resolution tsunami predictions in northern Cascadia by exploiting sequential design for efficient emulation, *Nat. Hazards Earth Syst. Sci.*, 21, 3789–3807, <https://doi.org/10.5194/nhess-21-3789-2021>, 2021.
- Satake, K., Nishimura, Y., Putra, P. S., Gusman, A. R., Sunendar, H., Fujii, Y., Tanioka, Y., Latief, H., and Yulianto, E.: Tsunami source of the 2010 Mentawai, Indonesia earthquake inferred from tsunami field survey and waveform modeling, *Pure Appl. Geophys.*, 170, 1567–1582, <https://doi.org/10.1007/s00024-012-0536-y>, 2013.
- Sato, F., Tanabe, T., Murase, H., Tominari, M., and Kawai, M.: Application of a wearable GPS unit for examining interindividual distances in a herd of Thoroughbred dams and their foals, *Journal of Equine Science*, 28, 13–17, <https://doi.org/10.1294/jes.28.13>, 2017.
- Shinagawa, H., Tsugawa, T., Matsumura, M., Iyemori, T., Saito, A., Maruyama, T., Jin, H., Nishioka, M., and Otsuka, Y.: Two-dimensional simulation of ionospheric variations in the vicinity of the epicenter of the Tohoku-oki earthquake on 11 March 2011, *Geophys. Res. Lett.*, 40, 5009–5013, <https://doi.org/10.1002/2013GL057627>, 2013.
- Takagawa, T. and Tomita, T.: Effects of Rupture Processes In an Inverse Analysis On the Tsunami Source of the 2011 Off the Pacific Coast of Tohoku Earthquake, in: *The Twenty-second International Offshore and Polar Engineering Conference*, 17–22 June 2012, Rhodes, Greece, ISBN 978-1-880653-94-4, 2012.
- Tanioka, Y. and Gusman, A. R.: Near-field tsunami inundation forecast method assimilating ocean bottom pressure data: A synthetic test for the 2011 Tohoku-oki tsunami, *Phys. Earth Planet. In.*, 283, 82–91, <https://doi.org/10.1016/j.pepi.2018.08.006>, 2018.
- The Japanese government: Japan Association of Surveyors website, <https://jsurvey.jp/eng.htm>, last access: 8 March 2022.
- Tsugawa, T., Saito, A., Otsuka, Y., Nishioka, M., Maruyama, T., Kato, H., Nagatsuma, T., and Murata, K. T.: Ionospheric disturbances detected by GPS total electron content observation after the 2011 off the Pacific coast of Tohoku Earthquake, *Earth Planets Space*, 63, 875–879, <https://doi.org/10.5047/eps.2011.06.035>, 2011.
- Vernon, I., Goldstein, M., and Bower, R.: Galaxy formation: Bayesian history matching for the observable universe, *Stat. Sci.*, 29, 81–90, <https://doi.org/10.1214/12-STS412>, 2014.
- Wächter, J., Babeyko, A., Fleischer, J., Häner, R., Hammitzsch, M., Kloth, A., and Lendholt, M.: Development of tsunami early warning systems and future challenges, *Nat. Hazards Earth Syst. Sci.*, 12, 1923–1935, <https://doi.org/10.5194/nhess-12-1923-2012>, 2012.
- Wang, Y., Maeda, T., Satake, K., Heidarzadeh, M., Su, H., Sheehan, A., and Gusman, A.: Tsunami data assimilation without a dense observation network, *Geophys. Res. Lett.*, 46, 2045–2053, <https://doi.org/10.1029/2018GL080930>, 2019.
- Zettergren, M. and Snively, J.: Latitude and Longitude Dependence of Ionospheric TEC and Magnetic Perturbations From Infrasonic-Acoustic Waves Generated by Strong Seismic Events, *Geophys. Res. Lett.*, 46, 1132–1140, <https://doi.org/10.1029/2018GL081569>, 2019.
- Zettergren, M., Snively, J., Komjathy, A., and Verkhoglyadova, O.: Nonlinear ionospheric responses to large-amplitude infrasonic-acoustic waves generated by undersea earthquakes, *J. Geophys. Res.-Space*, 122, 2272–2291, <https://doi.org/10.1002/2016JA023159>, 2017.