Assessing Storm Surge Multi-Scenarios based on Ensemble Tropical Cyclone Forecasting

Md. Rezuanul Islam[1], Le Duc[1], and Yohei Sawada[1]

[1]The University of Tokyo

March 26, 2023

Abstract

Ensemble forecasting is a promising tool to aid in making informed decisions against risks of coastal storm surges. Although tropical cyclone (TC) ensemble forecasts are commonly used in operational numerical weather prediction systems, their potential for disaster prediction has not been maximized. Here we present a novel, efficient, and practical method to utilize a large ensemble forecast of 1000 members to analyze storm surge scenarios toward effective decision making such as evacuation planning and issuing surge warnings. We perform the simulation of TC Hagibis (2019) using the Japan Meteorological Agency's (JMA) non-hydrostatic model. The simulated atmospheric predictions were utilized as inputs for a statistical surge model named the Storm Surge Hazard Potential Index (SSHPI) to estimate peak surge heights along the central coast of Japan. We show that Pareto optimized solutions from an ensemble storm surge forecast can describe potential worst (maximum) and optimum (minimum) storm surge scenarios while exemplifying a diversity of trade-off surge outcomes among different coastal places. For example, some of the Pareto optimized solutions that illustrate worst surge scenarios for inner bay locations are not necessarily accountable for bringing severe surge cases in open coasts. We further emphasize that an in-depth evaluation of Pareto optimal solutions can shed light on how meteorological variables such as track, intensity, and size of TCs influence the worst and optimum surge scenarios, which is not clearly quantified in current multi-scenario assessment methods such as those used by JMA/National Hurricane Center in the United States.

Assessing Storm Surge Multi-Scenarios based on Ensemble Tropical Cyclone Forecasting

Md. Rezuanul Islam[1*], Le Duc[1,2], and Yohei Sawada[1,2]

[1]Institute of Engineering Innovation, The University of Tokyo, Tokyo, 113-0032, Japan

[2]Meteorological Research Institute, Japan Meteorological Agency

Corresponding author: Md. Rezuanul Islam (fahiemislam@gmail.com)

Key Points:

- The potential of ensemble tropical cyclone forecasting for assessing storm surge multi-scenarios is shown.
- Pareto optimized solutions from an ensemble storm surge forecast can efficiently illustrate potential worst and minimum storm surge scenarios.
- Analyses of meteorological variables of ensemble members in Pareto frontiers help understand the impact of a tropical cyclone on predicted storm surge multi-scenarios.

Abstract

Ensemble forecasting is a promising tool to aid in making informed decisions against risks of coastal storm surges. Although tropical cyclone (TC) ensemble forecasts are commonly used in operational numerical weather prediction systems, their potential for disaster prediction has not been maximized. Here we present a novel, efficient, and practical method to utilize a large ensemble forecast of 1000 members to analyze storm surge scenarios toward effective decision making such as evacuation planning and issuing surge warnings. We perform the simulation of TC Hagibis (2019) using the Japan Meteorological Agency's (JMA) non-hydrostatic model. The simulated atmospheric predictions were utilized as inputs for a statistical surge model named the Storm Surge Hazard Potential Index (SSHPI) to estimate peak surge heights along the central coast of Japan. We show that Pareto optimized solutions from an ensemble storm surge forecast can describe potential worst (maximum) and optimum (minimum) storm surge scenarios while exemplifying a diversity of trade-off surge outcomes among different coastal places. For example, some of the Pareto optimized solutions that illustrate worst surge scenarios for inner bay locations are not necessarily accountable for bringing severe surge cases in open coasts. We further emphasize that an in-depth evaluation of Pareto optimal solutions can shed light on how meteorological variables such as track, intensity, and size of TCs influence the worst and optimum surge scenarios, which is not clearly quantified in current multi-scenario assessment methods such as those used by JMA/National Hurricane Center in the United States.

Plain Language Summary

Ensemble forecasting generates multiple predictions of a weather event with various possible outcomes based on varying initial conditions, model parameters, and physics. The potential of ensemble tropical cyclone (TC) forecasting for assessing storm surge multi-scenarios has largely been overlooked previously. Enhanced analysis can unlock and maximize the benefit of ensemble forecasting. This study simulated an extremely large ensemble (=1000 members) to reforecast past TC Hagibis which hit the central coast of Japan in 2019 and utilized the results to predict storm surges. We propose that Pareto optimality can identify good ensemble members that reasonably represent potential worst/minimum storm surge scenarios, meaning no other ensemble members can represent better than those. Comprehensive analyses of Pareto members can give forecasters and decision makers a better understanding of how the predicted track, wind intensity, and size of a TC can impact the worst and best storm surge scenarios. This type of analysis is expected to improve the planning of evacuations and the issuing of storm surge warnings.

1 Introduction

Since 1737, 29 coastal storm surge events have claimed at least 5,000 people globally. Two of these events happened in the 21st century and ranked as two of the five worst coastal disasters in the running millennium (Needham et al., 2015; Takagi et al., 2022). Rappaport (2014) has shown that 49% of tropical cyclone (TC)-induced deaths are directly attributed to storm surges. Hence, it is crucially important to improve the understanding of storm surge and their associated risk as it is among the deadliest and most destructive natural disasters.

In recent years, forecast services have likely reduced TC-induced deaths relative to historical standards. For example, several countries have already adopted a dynamical TC ensemble prediction system (EPS) to capture forecast uncertainties and reduce sampling errors in the three-dimensional meteorological simulation (Sharma et al., 2022). Numerical weather prediction centers such as Japan Meteorological Agency (JMA), National Centers for Environmental Prediction in the United States (US), European Centre for Medium-Range Weather Forecasts generate TC track forecasts from their ensemble forecast models and utilize them in their operational settings (Swinbank et al., 2016). Yamaguchi et al. (2015) have shown that EPS can provide skillful guidance of TC genesis forecasts with a forecast lead time extending to two weeks in seven TC basins. Nevertheless, there is a great potential to maximize the use of this EPS not only in TC activity (e.g., track, intensity) forecast but also in forecasting hazards (e.g., storm surge), aiding end users to be prepared better before the dangerous situation (Kobayashi et al., 2020; Duc et al., 2021).

Titley et al. (2019) have recently conducted a questionnaire survey at operational TC forecast centers worldwide to understand the current and potential use of EPS in operational TC forecasting. They reported that over 90% of respondents used an ensemble forecast for TC track forecast, followed by genesis and intensity forecasts. In contrast, less than 10% of surveyed forecasters use ensemble products for hazard (e.g., storm surge) forecasting. Deterministic forecasts are often used for hazard forecasting as it is produced using the best available TC data and unperturbed models. In some cases, ensemble mean (e.g., track and intensity of TC) is used as inputs for hazard forecast to compare the result with the deterministic forecasts, although the full use of EPS in hazard forecasting remains challenging (Titley et al., 2019). A lack of detailed analysis of ensemble members (beyond ensemble mean/median analysis) and less technical expertise on ensemble-based hazard forecasts hinder its' application among hazard forecasters. Wilson et al. (2019) reported that a deterministic mindset resulted in tendencies to modify understanding of probabilistic concepts when presented with different meteorological variables. Furthermore, local authorities responsible for hazard forecasting avoid EPS information as citizens and emergency managers habitually trust a single forecast only, and they are not sufficiently educated to deal with the probabilistic prediction (Lombardi et al., 2018). These findings highlight that ensemble-based hazard (e.g., storm surge) forecast is unfamiliar in disaster risk management communities.

Notwithstanding the challenges mentioned above, ensemble surge prediction system (ESPS) has recently received considerable attention from both the research and operational communities. For instances, Flowerdew et al. (2013), Greenslade et al. (2017), and Kristensen et al. (2022) have successfully developed and evaluated the performance of an operational ESPS for United Kingdom, Australia, and Norway, respectively. Along the coastline of Canada, it was found that 20-member ESPS could reasonably estimate both the uncertainty in peak surge height and timing of surge events resulting from imperfectly forecast atmospheric conditions six days before (Bernier & Thompson, 2015). A 50-member ensemble simulation of 10 surge events during 2010 in Venice by Mel & Lionello (2014) has shown that the distribution of maximum sea level is acceptably realistic with respect to the deterministic forecast. They also found that the uncertainty became its maximum during storm surge peaks and increased linearly with the forecasting lead time. Although these ensemble simulation studies paved the way for a robust surge hazard assessment over a single forecast-based assessment, they considered ensemble TC forecast information only for developing and evaluating the performance (skill and accuracy) of an ESPS. In addition to quantifying the uncertainty of surge height, ensemble-based storm surge multi-scenario (e.g., worst/optimum case) analysis is equally important, aiding disaster risk managers in evacuation planning (Kohno et al., 2018).

To the best of our knowledge, the potential of ensemble TC forecasting for assessing storm surge multi-scenarios has largely been overlooked previously. However, recent developments have seen the introduction of multi-scenario storm surge predictions, such as the worst-case scenario from six typical TC tracks by the JMA (H. Hasegawa et al., 2017) and the maximum storm tide height by the National Hurricane Center in the US (NHC, n.d.). These worst-case scenarios are composite products, representing the maxima among all scenarios. Therefore, it is possible that the worst-case values for two adjacent locations may have come from two different ensemble TC track run. Therefore, the users (e.g., emergency managers) cannot understand which forecasted TC track or which combination of forecasted TC meteorological variables (track, intensity, size, translation speed) may trigger the worst surge scenario for a particular location based on a composite product. This can make it difficult for decision-makers to determine the appropriate level of storm surge warning and evacuation orders. In addition, storm surge is spatially heterogeneous because of its' dependency on a TC characteristic and coastal geometry. It is entirely plausible that the worst case scenario may not occur everywhere within a forecasted TC threat zone (Islam & Takagi, 2020a, 2020b). If the decision makers in cities/tourist districts with highly valuable economies issue a higher warning level without any concrete understating over a worst event, they will inevitably suffer significant economic losses because of false alarming (in case the area has not affected by worst storm surge) and eventually can lower citizens trust over official warning (Sawada et al., 2022; Takagi et al., 2018).

Here we present Pareto optimality - a novel way of assessing storm surge multi-scenarios based on ensemble TC forecasts. Our approach is more advanced than existing assessments. We employed a multi-objective function to determine possible worst/optimum cases to quantify the hazards in a large region. Our approach involved a comprehensive analysis of Pareto optimal solutions in understanding the combination of forecasted TC meteorological variables - such as track, intensity, size, and translation speed of TC - that could result in the worst/optimum surge scenario. We utilized an extremely large ensemble (=1000 member) forecasts of TC Hagibis that made landfall in central Japan in 2019. Our Pareto-based optimal solutions provide an instantaneous overall assessment of storm surge multi-scenarios without any computational burdens. The proposed method will allow forecasters to predict storm surge multi-scenarios harnessing ensemble TC forecasts efficiently and help emergency responders as means of quantifying surge hazards effectively.

2 Data and Methods

2.1 TC Hagibis and ensemble forecast

TC Hagibis in 2019, one of the most destructive and deadliest TC that hit Japan in decades (Shimozono et al., 2020; Ma et al., 2021), has been chosen to demonstrate our multi-scenario storm surge assessment. Hagibis was formed in the western North Pacific Ocean on 2 October 2019 and made landfall in central Japan on 12 October 2019 (around 0900 UTC), as depicted in Figure 1. At the landfall time, its maximum wind speeds sustained at 80 kt. This combined with heavy rainfall, resulted in high storm surges and severe flooding in the area (Shimozono et al., 2020; Ma et al., 2021; JMA, 2021).

The atmospheric ensemble forecasts of TC Hagibis were obtained by running JMA's former operational limited-area model called NHM (non-hydrostatic model; Saito et al., 2006). The integration domain (see Figure S1) had a grid spacing of 5 km consisting of 817 × 661 horizontal grid points and 50 vertical levels. Boundary conditions were interpolated to the NHM domain from JMA's global model forecasts and the forecast perturbations of JMA's operational one-week EPS.

Since we used NHM for all forecast members, the only source of uncertainty stemmed from initial conditions. This uncertainty is encapsulated in error covariances of current atmospheric states (analysis error covariances), estimated using a data assimilation system. An ensemble Kalman filter (EnKF) was employed to sample from these error covariances and generate an analysis ensemble. While operational forecast centers generally use around 100 ensemble members, a state-of-the-art data assimilation system with 1000 ensemble members, called the four-dimensional variational-ensemble assimilation technique (4DEnVAR), was utilized in this study (Kobayashi et al., 2020). Our 4DEnVAR system only applied horizontal localization, with the horizontal localization length scales derived from the JMA's operational four-dimensional variational assimilation system's climatological horizontal correlation length scales. This helped to remove sampling noise in estimating forecast error covariances and maintain the coherent vertical structure between atmospheric fields, which is critical in predicting tropical cyclones. As the ensemble member count was large (=1000), localization was relaxed by retaining vertical correlations and removing horizontal correlations at distant locations (Duc et al., 2021).

Unlike EnKF, EnVAR solely estimates the means of analysis ensembles and not the analysis ensembles themselves, even though this method heavily relies on forecast ensembles to estimate these means. To solve this issue, a common approach is to run a separate EnKF in parallel to generate analysis ensembles. However, our 4DEnVAR system is unique in that an EnKF is not necessary. Instead, the same EnVAR program was used to generate analysis perturbations, as suggested in the context of inflation functions (Duc et al., 2020), where we demonstrated that using quadratic inflation functions implies using the Kalman gain to generate analysis perturbations. Using the same program for analysis means and analysis perturbations is essential because it ensures consistency between the two when the same background error covariance, localization, and observations are utilized in both cases. The assimilation system commenced at 00UTC on 7 October 2019, with a 3-hour assimilation cycle and continued until 18:00UTC 10 October 2019. The final analysis ensemble was then used as initial conditions for 39h forecasts with NHM. The assimilation domain was chosen the same as the forecast domain in Figure S1 and we assimilated all routine observations obtained from JMA's database. Here, we opted for a 39h forecast horizon because JMA's operational Meso-scale Ensemble Prediction System (MEPS) also generates 39h forecasts at 6-hour intervals (JMA, 2023).

2.2 Ensemble storm surge forecast

We used storm surge hazard potential index (SSHPI; eq. 1), a statistical model to compute peak storm surge height. While the coastal engineers and ocean modelers are interested in the forecast of storm surge hydrograph, most of the decision makers responsible for issuing surge warning and relief measures have a primary interest in the value of predicted peak surge height. The SSHPI uses meteorological variables sensitive to storm surge, including TC intensity (V_{max}), size (radius of 50-kt wind; R_{50}), and translation speed (S). In addition, the SSHPI considers coastal geometry (a = 1 = open coasts and a = −1 = bays), landfall location sensitivity (D_L), and regional scale bathymetry (L_{30}). The SSHPI does not incorporate factors associated with wave set-up and astronomic tide to keep the configuration simple. TC Hagibis ensemble forecasts (=1000 member; see Section 2.1), particularly during landfall, was used as meteorological forcing of the SSHPI. We produced corresponding 1000 perturbed surge forecasts with a lead time of 39h. The bathymetry of the target region was obtained from the Japan Oceanographic Data Center (Japan Oceanographic Data Center, 2020). The effectiveness of the SSHPI for predicting peak surge hazard potential was discussed in Islam et al. (2021, 2022). The formulation of the SSHPI is the following:

$$\mathrm{SSHPI} = \left(\frac{V_{max}}{V_{ref}}\right)^2 \left(\frac{R_{50}}{R_{ref}}\right) \left(\frac{S}{S_{ref}}\right)^a \left(\frac{L_{30}}{L_*}\right) (D_L) \qquad (1)$$

$$\frac{R_{50}}{R_{ref}} = \begin{cases} 1.5 & if\ \frac{R_{50}}{R_{ref}} \geq 1.5 \\ \frac{R_{50}}{R_{ref}} & if\ 0.5 < \frac{R_{50}}{R_{ref}} < 1.5 \\ 0.5 & if\ \frac{R_{50}}{R_{ref}} \leq 0.5 \end{cases} ; \quad \left(\frac{S}{S_{ref}}\right)^a = \begin{cases} 1.5 & if\ \left(\frac{S}{S_{ref}}\right)^a \geq 1.5 \\ \left(\frac{S}{S_{ref}}\right)^a & if\ 0.5 < \left(\frac{S}{S_{ref}}\right)^a < 1.5 \\ 0.5 & if\ \left(\frac{S}{S_{ref}}\right)^a \leq 0.5 \end{cases} ; \quad \frac{L_{30}}{L_*} = \begin{cases} \frac{L_{30}}{L_*},\ if\ \frac{L_{30}}{L_*} \geq 1 \\ 1,\ if\ \frac{L_{30}}{L_*} \leq 1 \end{cases}$$

$$D_L = \begin{cases} 1 & if\ the\ surge\ estimated\ point\ falls\ right\ side\ of\ TC\ track\ and\ x \leq 20 \\ & OR \\ & if\ the\ surge\ estimated\ point\ falls\ left\ side\ of\ TC\ track\ and\ x \leq 10 \\ 1 - \dfrac{0.03(x-20)}{20} & if\ the\ surge\ estimated\ point\ falls\ right\ side\ of\ TC\ track\ and\ x > 20 \\ 1 - \dfrac{0.05(x-10)}{10} & if\ the\ surge\ estimated\ point\ falls\ left\ side\ of\ TC\ track\ and\ x > 10 \end{cases}$$

V_{ref}, R_{ref}, and S_{ref}, are reference constants as follows: 50-kt equivalents of the tropical storm category, 95 NM (historical mean R_{50} at the time of landfall in Japan mainland), and 35 km/h (historical mean S at the time of landfall in Japan), respectively (Islam et al., 2021). L_{30} is the horizontal distance (km) between the shoreline and the 30-m depth contour. L^* was chosen to be 10 km. D_L is defined by different expressions depending on the surge estimated points' (e.g., tidal station) position (right/left) respective to the TC track and horizontal distance (x in km) between the TC landfall location and a surge estimated point. Compared to V_{max}, the upper and lower bounds of R_{50}, S, and L_{30} in eq. 1 restrict their contribution in generating surge hazards and, thus, prevents discrete jumps in the SSHPI.

Figure 1 shows a storm surge modeling domain and the position of tide gauges used for validating surge model and predicting surge hazards in this study. There are two domains, covering Tokyo Bay and Suruga Bay individually. Each domain has tide gauges located both in inner bays and open coasts. It should be noted that the tide gauges chosen for this study are the only stations that possess recorded (historical) storm surge data, which is kept by JMA (JMA, 2022) and Japan Coast Guard (Japan Oceanography Data Center, 2021). The empirical relationship for expected storm surge in each tide gauge was determined in Islam et al. (2021, 2022) by drawing a line of best fit through the historical surge data and the SSHPI and thus, used for the surge forecasts in this study.

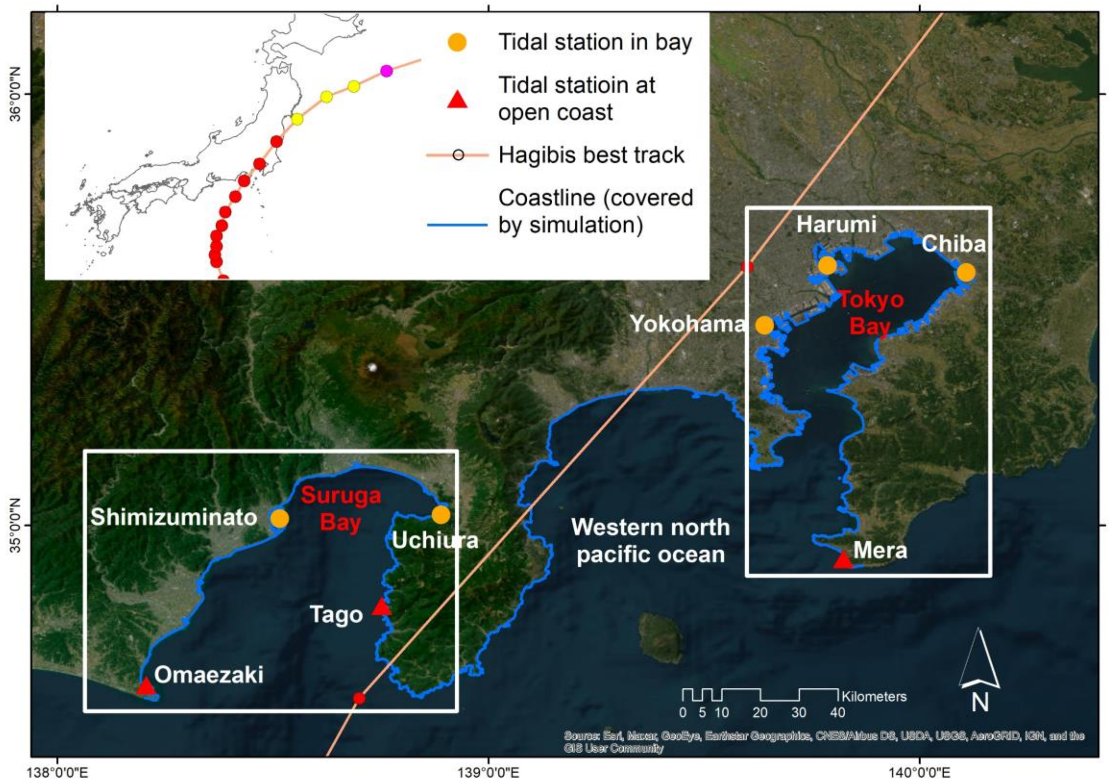

Figure 1. Domain of the storm surge forecasts model and the locations of the tide gauges used for model validation and surge forecasts.

2.3 Pareto optimality and assessing storm surge multi-scenarios

It is unrealistic to anticipate a "nice" forecast scenario that accurately predicts the exact intensity of a hazard at all locations within a given domain for a particular condition (e.g., worst/optimum). An improved forecast at one location is usually accompanied by a deterioration of forecast at another location and vice versa. The best we can do is to quantify the trade-off between different objectives. Here, we conducted multi-objective optimization to select ensemble forecast members (among 1000 ensemble forecasts; see Section 2.1 and 2.2) that reasonably characterize the potential worst/optimum storm surge case for a particular location (e.g., Tokyo Bay) by computing the Pareto frontier. The Pareto frontier captures the trade-offs between objectives. It is the set of all Pareto-optimum solutions where a single Pareto optimal solution denotes a solution that is not dominated by any other solution (Kochenderfer & Wheeler, 2019).

In this study, we analyzed a subset of 1000 forecasted surge scenarios for each tide gauge in a specific domain, referred to as solution z. Each scenario is evaluated based on d objectives, represented by the values $y^1(z), y^2(z), \ldots, y^d(z)$. As an example, we considered a scenario in Tokyo Bay where the objectives are to maximize the forecasted peak surge at four tide gauges: Harumi, Chiba, Yokohama, and Mera (as shown in Figure 1). This objective function considers the potential worst-case scenario in Tokyo Bay for a set of 1000 ensemble surge forecasts. We compared the given two solutions z and z', if for every objective i, $y^i(z) \geq y^i(z')$ and the strict inequality holds for at least for one objective, we considered that solution z dominates z'. In other words, if one solution (e.g., ensemble member no. 10) provides 160 cm, 155 cm, 140 cm, 110 cm of forecasted peak surge in Harumi, Chiba, Yokohama, and Mera, respectively, but another solution (e.g., ensemble member no. 20) yields 160 cm, 155 cm, 140 cm, 90 cm of forecasted peak surge for the same tide gauges, then the solution given by ensemble member no. 10 dominates the solution provided by ensemble member no. 20. This is because a smaller storm surge (90 cm) in Mera is undesirable as stated in the objective function. A similar objective function but minimizing the forecasted peak surge at four tide gauges: Harumi, Chiba, Yokohama, and Mera (as shown in Figure 1) was considered to determine ensemble forecasts member that characterizes a potential optimum surge case. In order to select the most appropriate ensemble TC member that may cause the worst surge case in the inner bay, but also results in the optimum surge at the open coast, we further assume that objectives are to be maximized for the inner bay tide gauges (Harumi, Chiba, and Yokohama), but minimized for open coast tide gauges (Mera) at the same time.

The details of the implementation of the algorithm used here are described in Tommy (2021). This algorithm can compute the Pareto frontier for four objectives within a minute, meaning the runtime should be acceptable to any operational hazard forecast settings.

3 Results

3.1 Model evaluation

3.1.1 TC ensemble forecasts validation

From the forecasts of atmospheric fields given by NHM, TC tracks and intensities were detected. Here, TC centers are defined as the average of the mean sea level pressure minima, geopotentials at 850 hPa and 700 hPa. Figure 2(b) shows 1000 track forecasts generated from 1000 initial conditions obtained from the 4DEnVAR data assimilation system, along with the ensemble mean forecast, control forecast, and best track. The ellipses in the figure illustrate uncertainties of the TC centers, which are determined by the forecast error covariances of the TC centers. The arrows show the distance errors between the observations (i.e., the best track) and the ensemble mean. For comparison, the operational 20-member JMA ensemble forecast (MEPS) is included in Figure 2a. As shown in Figure 2, 4DEnVAR outperforms MEPS in terms of track forecasts and its ensemble mean is almost identical to the best track with only minor distance errors.

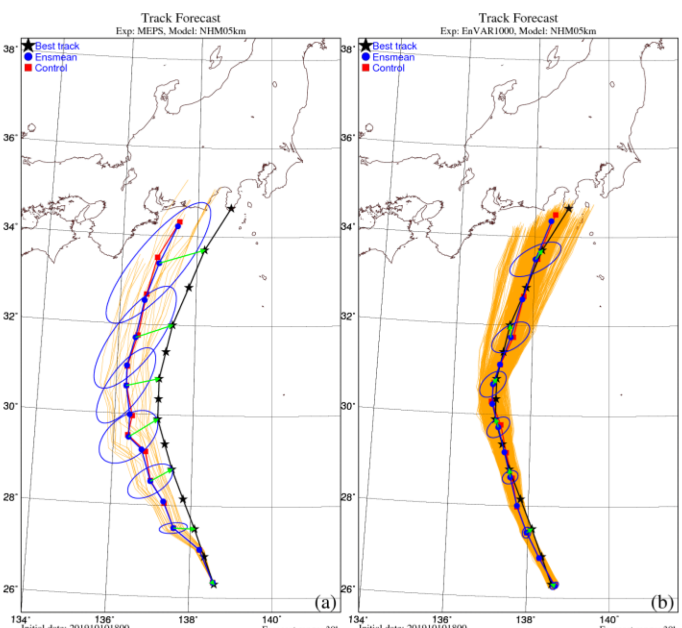

Figure 2. A comparison of 39h (at 1800 UTC on 10 October 2019) ensemble track forecasts for TC Hagibis issued by (a) the operational ensemble prediction system MEPS of JMA and (b) the 4DEnVAR data assimilation system. The ellipses represent forecast error covariances of the TC centers. The arrows denote the distance errors between the ensemble mean and best track.

Figure 3 presents the forecasted intensity of TC Hagibis as indicated by its central pressure. It is evident from the figure that the 4DEnVAR (Figure 3b) surpasses JMA's operational MEPS (Figure 3a) in predicting the intensity. Even though both ensemble forecasts show overestimation of intensity, the tendency of overestimation becomes more apparent with increasing forecast ranges in MEPS (Figure 3a). Despite having a smaller number of ensemble members, MEPS exhibits greater uncertainty in intensity forecast as compared to 4DEnVAR. This can be attributed to the fact that JMA's operational MEPS employs singular vectors to generate initial conditions for ensemble members, which maximizes their spread (JMA, 2023).

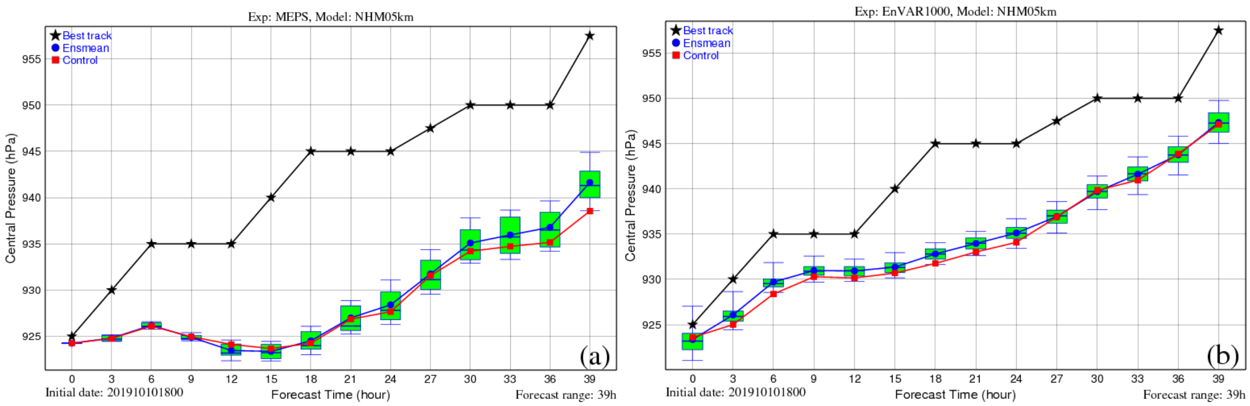

Figure 3. A comparison of 39h (at 1800 UTC on 10 October 2019) ensemble intensity forecast for TC Hagibis issued by (a) the operational ensemble prediction system MEPS of JMA and (b) the 4DEnVAR data assimilation system. The distributions of ensemble intensities are represented as box-and-whisker plots.

3.1.2 Storm surge ensemble forecasts validation

A comparison of the forecasted and measured peak storm surge (Japan Oceanography Data Center, 2021; JMA, 2022) at eight different tide gauges is shown in Figure 4a. It is noted that a significant deviation from the climatological mean surge height is observed at all stations during TC Hagibis. We evaluate JMA best track (JMA, 2021) as an ideal meteorological forcing input as well as our 39h ensemble TC forecasts. Both ensemble median forecasts and best track estimates systematically underestimate the observed peak levels. Nevertheless, the observed peak surge values are enveloped by the full ensemble of the forecasts in most tide gauges, implying that the ensemble spread is large enough to represent the uncertainty in the prediction. The average mean absolute error of the eight stations is 11.3 cm. In the case of Mera and Omaezaki, wave set-up is often the dominant driver for generating storm surges (Islam et al., 2018, 2022), which is not considered in the SSHPI. Therefore, the ensemble surge forecast underestimated observed surges.

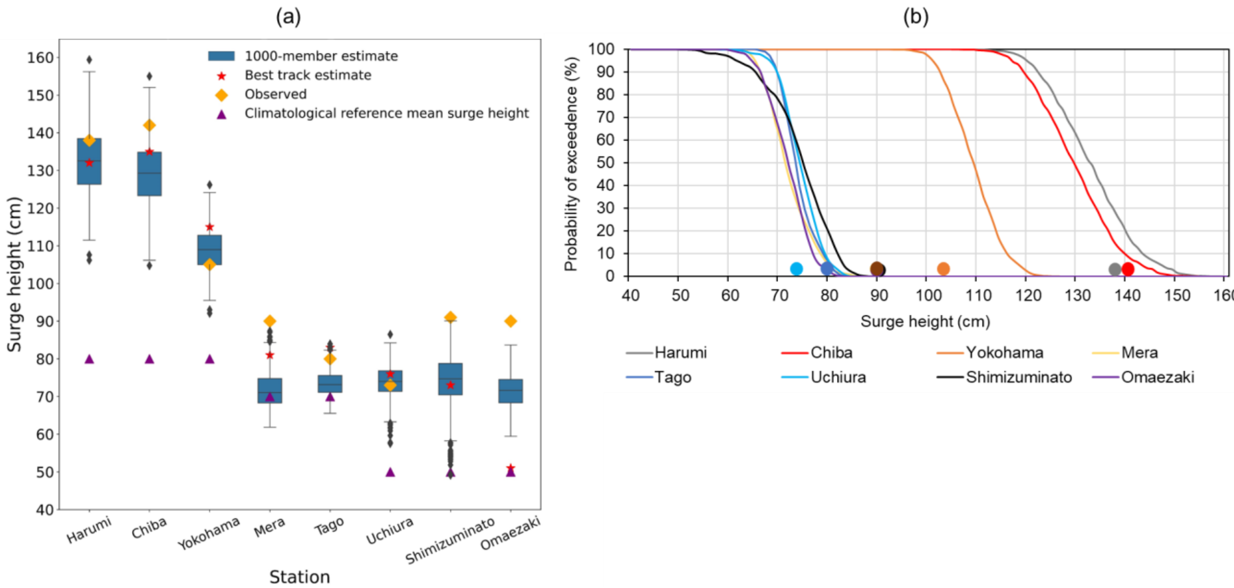

Figure 4. (a) A comparison of the 39h (at 1800 UTC on 10 October 2019) ensemble peak surge forecasts and the observed peak surge height for TC Hagibis at eight tide gauges. (b) Probability of exceeding (Y-axis) for a given storm surge threshold (X-axis) during TC Hagibis landfall time (at 0900 UTC, 12 October 2019). It was estimated by fitting ensemble forecasts empirically. A circle that shares the same color as a line represents the peak surge height recorded at a specific tide gauge.

The probability of surpassing a specific surge threshold during the landfall time of TC Hagibis is illustrated in Figure 4b, as determined by the 39h ESPS. The observed peak surge levels are within the predicted range, except for Mera and Omaezaki. For example, the probability of surpassing the observed peak surge for Harumi (138 cm) is 26.4%. In general, Figure 4 indicates that the SSHPI and its corresponding 39h peak surge forecasts are comparable in quality to those produced by numerical surge models such as Liu et al. (2021). The latter study reported an RMSE of ~10 cm when predicting the maximum total water level in Tokyo Bay during TC Hagibis with a 72h forecast horizon, using atmospheric forcing fields from the Global Forecast System and a hydrodynamic model known as the Semi-implicit Cross-scale Hydroscience Integrated System, which has a nearshore resolution of ~150 m.

3.2 Multi-scenario analysis

3.2.1 Pareto optimal multi-scenarios

The Pareto-optimal frontier, as shown in Figure 5, illustrates a group of solutions that depict the forecasted potential worst and optimum storm surge scenarios for TC Hagibis in Tokyo Bay. The two-dimensional Pareto frontiers (Figure 5a), allow for a straightforward evaluation of trade-offs among the forecasted peak storm surge levels. The results identify the best one or two members from the 1000 TC forecasts to represent the potential worst (Harumi, Chiba: ~155 cm) or optimum (Harumi, Chiba: ~106 cm) surge scenario in the inner Tokyo Bay. Here, the tide gauges (Harumi, Chiba) possess similar coastal geometry features, including bathymetry, and are situated in close proximity to each other (Figure 1). Therefore, the predicted surge response (from 1000 TC forecasts; Figure 5a) between them is almost linear.

Figure 5b reveals a diversity of trade-off surge outcomes in the Pareto frontier for Tokyo Bay. It includes four Pareto frontiers for worst-surge scenarios and six Pareto frontiers for optimum surge scenarios. This diversity is due to the emergence of trade-offs among tide gauges with distinct coastal geometry characteristics, including bathymetry. For instance, some of the identified most severe surge scenarios for open coastlines in Tokyo Bay do not result in high surge levels in the inner bay. Specifically, a Pareto optimal solution in Figure 5b predicts that Mera would experience the worst surge levels of 90 cm (<1% exceedance probability; Figure 5c), while Harumi and Chiba would experience approximately 135 cm (>25% exceedance probability; Figure 5c) of highest surge levels under the same scenario. This storm surge level (~135 cm) in Harumi and Chiba is substantially less than the worst surge levels (~155 cm) predicted by other optimal solutions. Additionally, the surge intensity may vary across Tokyo Bay, depending on the characteristics of the approaching TC and the impact can be much more severe in some places compared to others. For example, several Pareto optimal solutions shown in Figure 5d predict that the inner Tokyo Bay such as Harumi would witness surge levels higher than 150 cm, while it would keep as minimum as 70 cm along the open coastline (e.g., Mera). Owing to such surge incongruence among the coastal locations, creating multiple scenarios for different coastal places can lead to multiple optimal solutions, as seen in Figures 5b and 5d. This emphasizes the importance of considering multiple scenarios when issuing warnings and assessing the risks posed by extreme weather events like storm surges.

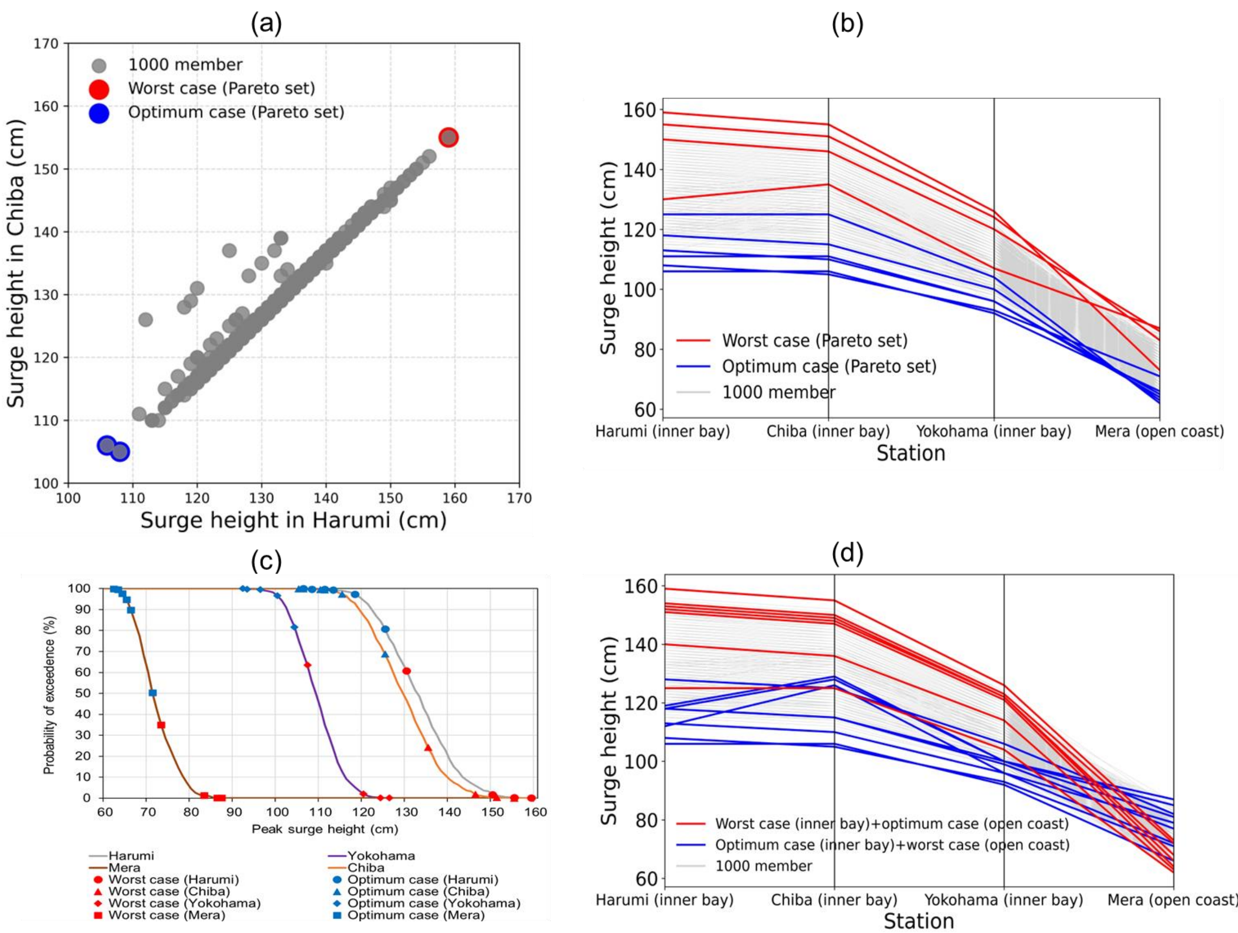

Figure 5. Forecasted Pareto optimal multi-scenarios due to TC Hagibis, apply for (a) Harumi and Chiba in Tokyo Bay [objective function (red dot): max surge height in Harumi and Chiba; objective function (blue dots): min surge height in Harumi and Chiba]; (b) Tokyo Bay (Harumi, Chiba, Yokohama and Mera) using the parallel coordinate plot [objective function (red lines): max surge height in Harumi, Chiba, Yokohama, and Mera; objective function (blue lines): min surge height in Harumi, Chiba, Yokohama, and Mera]; (c) Probability of exceeding (Y-axis) a storm surge threshold (X-axis) determined from each Pareto optimal solution in Figure 5b, at TC Hagibis landfall time (at 0900 UTC, 12 October 2019). It was estimated by fitting ensemble forecasts empirically; (d) Parallel coordinate plot including forecasted Pareto optimal solutions where forecasted peak surge intensity varies across Tokyo Bay [objective function (red lines): max surge height in Harumi, Chiba, and Yokohama + min surge height in Mera; objective function (blue lines): min surge height in Harumi, Chiba, and Yokohama + max surge height in Mera].

The results displayed in Figure 6a are comparable to those in Figure 5b but for Suruga Bay. It includes nine Pareto frontiers for worst-surge scenarios and one Pareto frontier for optimum surge scenario. Figure 6a predicts that all selected tide gauges would experience the worst surge levels of ~85 cm (<1% exceedance probability; Figure 6b), while the minimum surge height would be ~55 cm (>99% exceedance probability; Figure 6b). It is noteworthy that only one worst-surge scenario is found to be shared by both the Pareto frontiers of Tokyo Bay (Figure 5b) and Suruga Bay (Figure 6a), which represents a potential worst-case across the Japanese coastline during TC Hagibis. This indicates that among 1000 TC ensemble forecasts, a particular TC ensemble member has the potential to bring severe surge levels to both Bays. The reason behind this commonality will be discussed in Section 3.2.2.

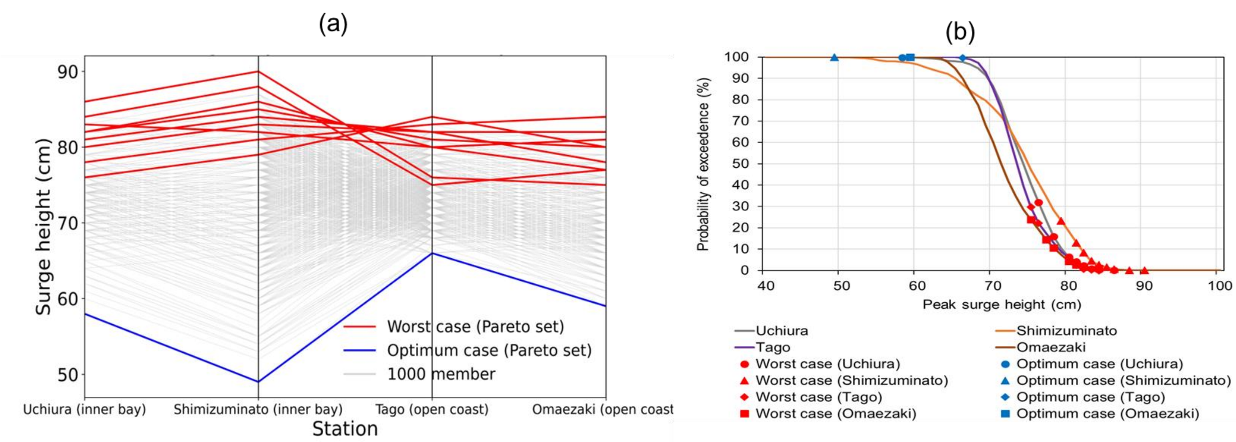

Figure 6. (a) Parallel coordinate plot with Pareto optimal multi-scenarios for Suruga Bay, at TC Hagibis landfall time (at 0900 UTC, 12 October 2019) [objective function (red lines): max surge height in Uchiura, Shimizuminato, Tago, and Omaezaki; objective function (blue lines): min surge height in Uchiura, Shimizuminato, Tago, and Omaezaki]; (b) Probability of exceeding (Y-axis) a storm surge threshold (X-axis) determined from each Pareto optimal solution in Figure 6a. It was estimated by fitting ensemble forecasts empirically.

In addition to Pareto Frontiers illustrated in Figures 5b and 6a, we further noticed many distinct trade-offs when both bays were considered together (Figure 7). For instance, we incorporated the representative tide gauges for each category of coastal geometry, such as inner bay (Harumi in Tokyo Bay and Shimizuminato in Suruga Bay) and open coast (Mera in Tokyo Bay and Tago in

Suruga Bay) in the objective function. This resulted in a large number of Pareto frontiers (worst case: nine; optimum case: twenty-two), with multiple overlapping solutions between worst and optimum cases. Thus, 33% of Pareto optimal solutions (red lines) predict that inner Tokyo Bay will experience the worst surge levels of ~150 cm, while inner Suruga Bay will witness ~60 cm, equivalent to minimum surge cases predicted by 41% of Pareto optimal solutions (blue lines; Figure 7). Although we maximize the potential of the large (i.e., 1000) ensemble forecasts, our proposed method is also found to be useful with the small ensemble size. We repeated the same analysis with 36 ensemble size. Although some worst and minimum storm surge scenarios were missed, a meaningful set of Pareto optimal solutions was still obtained (see Figure S2 in the supplementary section).

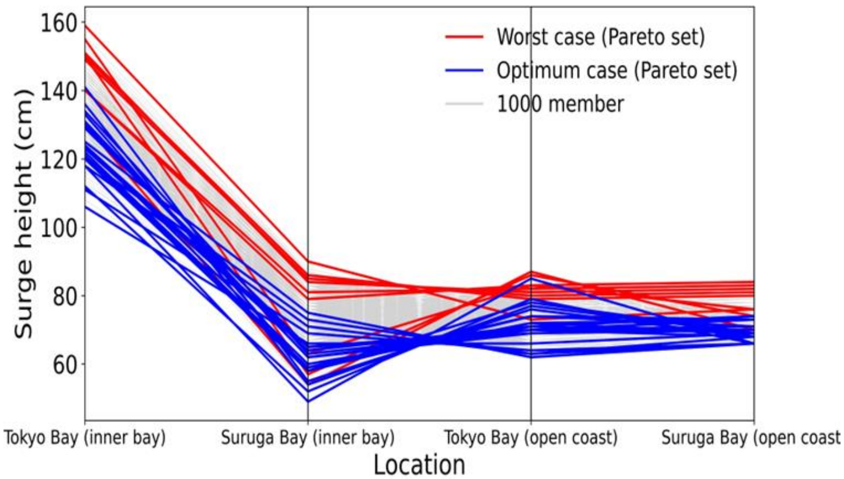

Figure 7. Parallel coordinate plot with Pareto optimal multi-scenarios determined for both Tokyo Bay and Suruga Bay [objective function (red lines): max surge height in Harumi, Shimizuminato, Mera, and Tago; objective function (blue lines): min surge height in Harumi, Shimizuminato, Mera, and Tago]

3.2.2 TC track and meteorological variables analysis of Pareto optimal solutions

It would be interesting to analyze the tracks and associated meteorological variables of the identified Pareto optimal solutions in Figures 5b and 6a. For example, Figure 8 reveals the strong sensitivity of the storm surge scenarios to the landfall location of TC Hagibis, leading to contrasting Pareto optimal solutions for Tokyo Bay (Figure 5b) and Suruga Bay (Figure 6a). It also demonstrates that the forecasted TC tracks for worst and optimum surge scenarios are significantly different from one other in both bays. For example, TC tracks (red lines; Figure 8a) that run parallel to the longitudinal axis of Tokyo Bay and pass over it would result in severe storm surges than TCs (blue lines; Figure 8a) that would travel 100 km or more to the west of the axis. Prior to landfall, under the worst case, easterly wind (Figure 9a) is forecasted to cause a buildup of water on the west coast (e.g., Yokohama) and initial draw-down in the north-eastern end of the Tokyo Bay (e.g., Chiba). Later, a surge level difference (0.6–0.8 m; Figure 5b) between the inner and lower ends of the bay is projected to occur (due to strong southerly winds) during the peak storm surge at the inner bays. During TC makes landfall under the optimum case (Figure 8a), the destructive right-side semicircle of the TC (Figure 9f) will interact with the vast land area rather than the ocean water, leading to less water being pushed towards Tokyo Bay (Figure 5b).

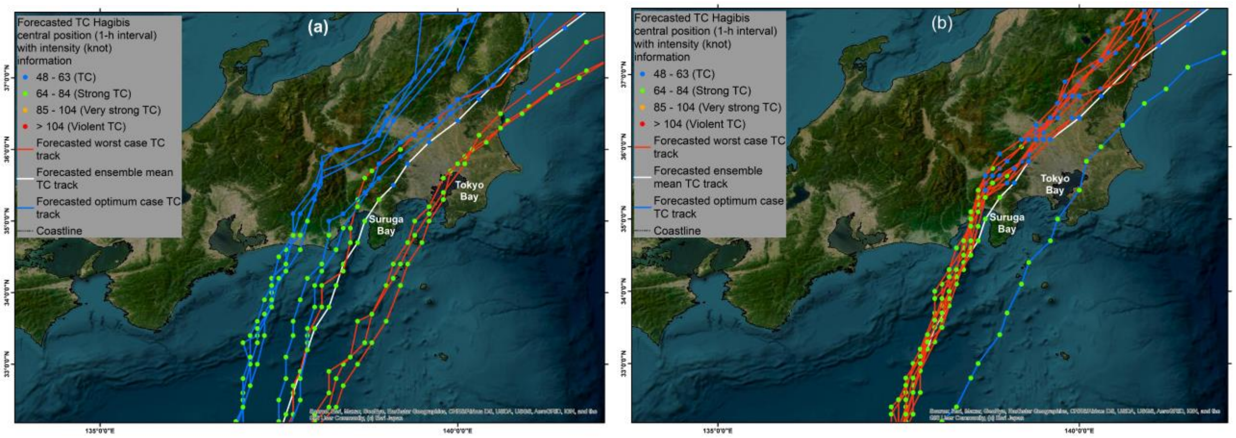

Figure 8. Forecasted TC Hagibis track in 39h lead time, respective to each Pareto frontier determined for (a) Tokyo Bay (as shown in Figure 5b); (b) Suruga Bay (as shown in Figure 6a). Red, blue, and white lines correspond to the forecasted worst, optimum, and ensemble mean TC track.

Notably, the optimized ensemble TC members for minimum surge scenarios in Tokyo Bay are forecasted to be stronger and larger than the members belonging to the worst surge cases until 24 hours prior to landfall, despite being centered in the same location. This is evident in Figure 9d, 9e, as opposed to Figure 9i, 9j. Despite both sets (worst and minimum) of optimized ensemble TC members weaken as they approach the mainland of Japan (Figure 9c, 9h and Figure 10), the worst TC members intensify by 7-kt (V_{max}) and remain large (R_{50}: ~120 NM) in the last 12 hours before landfall (Figure 9a, 9b and Figure 10). This large swath of strong winds is forecasted to affect a greater sea area and induce a motion in a greater quantity of water in Tokyo Bay.

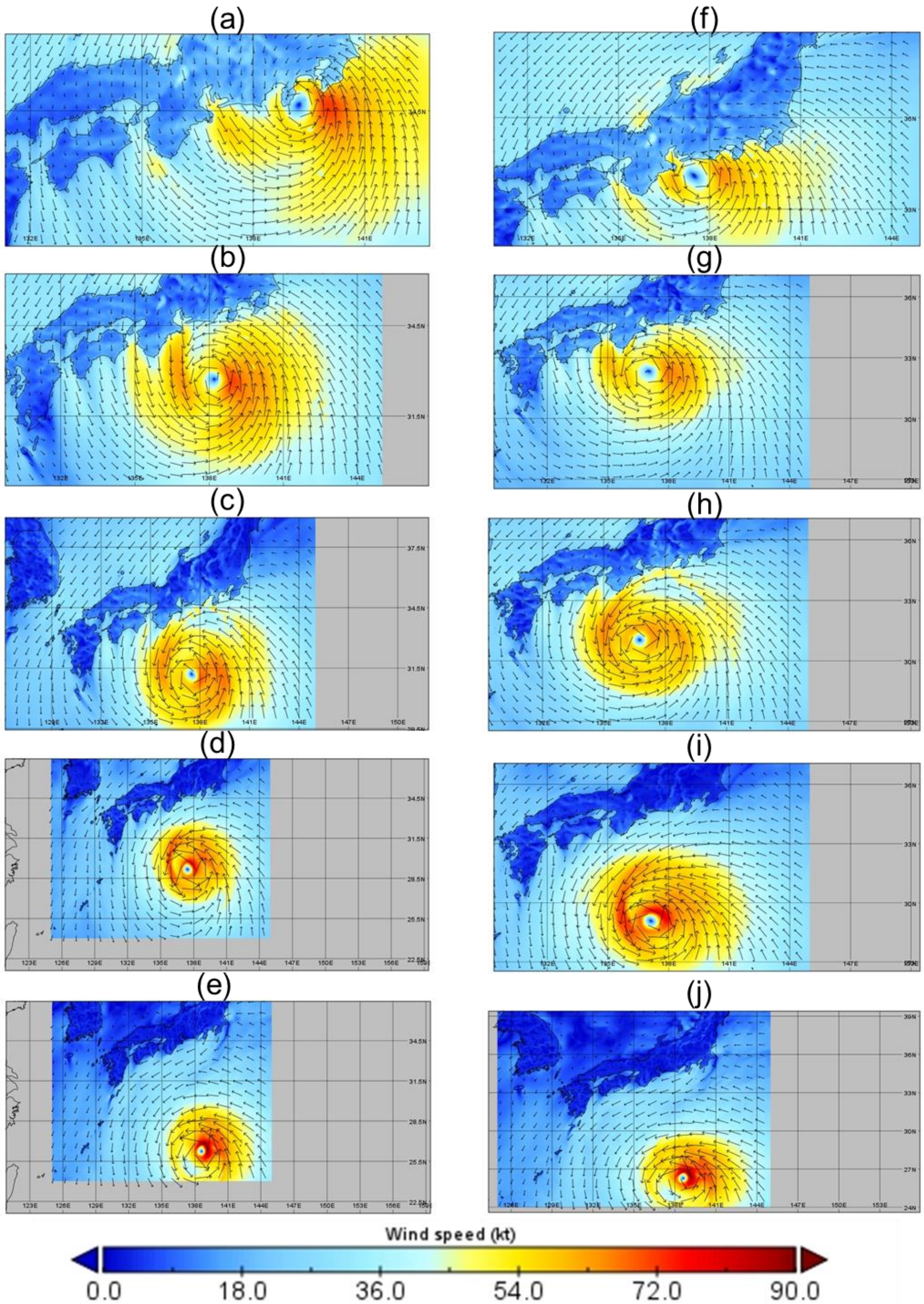

Figure 9. Forecasted composite 10 m wind field (kt-vectors), generated from optimized ensemble TC members for worst surge scenarios in Tokyo Bay (as shown in Figure 5b) during (a) landfall time (at 0900 UTC, 12 October 2019); (b) 6-h before landfall (at 0300 UTC, 12 October 2019); (c) 12-h before landfall (at 2100 UTC, 11 October 2019); (d) 24-h before landfall (at 0900 UTC, 11 October 2019); (e) 39-h before landfall (at 1800 UTC, 10 October 2019); (f-j) same as (a-e) but

generated from optimized ensemble TC members for minimum surge scenarios in Tokyo Bay (as shown in 5b).

Figure 8a highlights that in addition to TC tracks that pass directly over Tokyo Bay, one particular worst TC forecast (red line) makes landfall around 70 km west of the longitudinal axis of Tokyo Bay. This specific track is forecasted to cause severe storm surges in both Tokyo Bay (Figure 5b) and Suruga Bay (Figure 6a). This particular ensemble forecast has a wider range of intense winds (R_{50}: ~140 NM) across a larger area and a slower movement speed (S: ~32 km/h), despite having a similar landfall wind intensity (V_{max}: ~75-kt) compared to other worst-case TC forecasts. This unique phenomenon corroborates earlier numerical analyses that propose the likelihood of a severe storm surge scenario in the upper-bay region when a large and intense TC moves slowly, parallel to the longitudinal axis of Tokyo Bay, after making landfall 25 km southwest. (Islam & Takagi, 2020b).

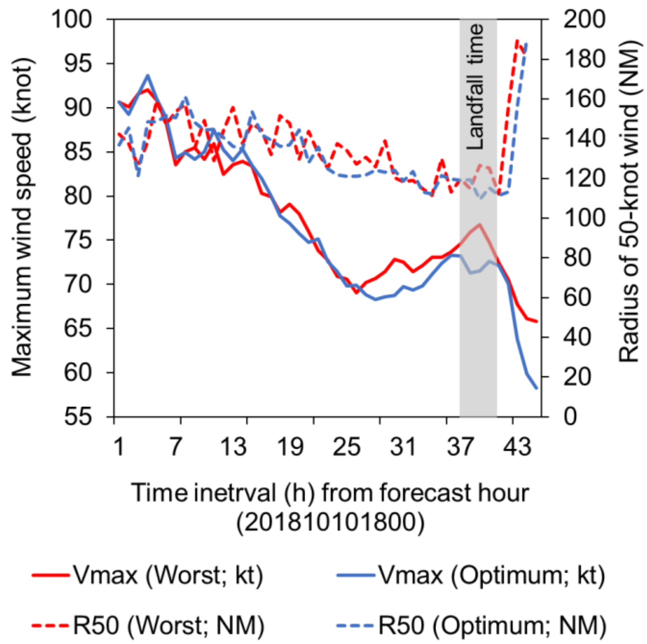

Figure 10. Evolution (composite) of maximum wind speed (V_{max}) and radius of 50-kt wind (R_{50}) for worst and optimum TC forecasts in Tokyo Bay (as shown in Figure 5b) in 39-h lead time.

4 Conclusions and discussion

The application of ensemble TC forecasting in hazard prediction, such as storm surge, has been greatly overlooked despite its use in forecasting TC track, intensity, and genesis. Enhanced analysis can unlock and maximize the benefit of ensemble forecasting. Here, we proposed Pareto optimality – a novel and practical way to identify potential ensemble TC (Hagibis) forecast from an extremely large ensemble (=1000 member) that can effectively assess storm surge multi-scenarios, including possible worst and optimum cases for a coastal location. The variability in storm surge intensity across the coastline makes it challenging for decision-makers to plan effective evacuation measures. To address this, we have demonstrated that a diversity of trade-off surge outcomes among coastal places can be identified by choosing the Pareto optimized forecasts. The in-depth evaluation of Pareto optimal solutions can shed light on how meteorological variables such as track, intensity, and size of TCs influence the worst and optimum surge scenarios, which are not well understood by emergency managers using current multi-scenario assessment methods (such as those used by JMA and NHC).

The significance of evaluating trade-offs based on Pareto optimization has long been acknowledged in the context of sustainable development goals and management of ecosystem services (Flecker et al., 2022). However, its application in disaster-risk communities is noticeably lacking. During a coastal storm surge event, effective evacuation planning and warning issuance involve multi-criteria problems such as storm surge intensity, coastal population vulnerability, and available evacuation resources. Traditionally, this decision-making process has relied on a hazard map, which typically depicts the severity of the predicted storm surge (e.g., exceeding a critical surge height; J. Hasegawa et al., 2017). However, this approach does not fully capture the diversity of potential storm surge scenarios across the coastlines, which can lead to ineffective evacuation planning. A recent TC Fani in 2019 serves as evidence to support this statement. TC Fani struck the southeastern part of India, approximately 450 km from the southwest coast of Bangladesh, as a Category 4 TC (in Saffir-Simpson Hurricane Wind Scale). Prior to TC Fani reaching the Bangladesh coast as a tropical storm, the Bangladesh Meteorological Department (BMD) issued 'danger' signal number seven (out of ten), which led to the evacuation of one million people (Bangladesh Meteorological Department, 2021). Later, the catastrophe, such as storm surge level (~1 m), did not hit the danger level as anticipated. BMD's evacuation order for the entire southwest coast was not based on a specific surge scenario (e.g., worst case) and forecasted meteorological conditions associated with it, leading to an excessive number of evacuees. Such a false alarm demotivated people to seek shelter when a Category 2 TC Amphan caused ~2.75 m storm surges and claimed at least 26 lives in 2020, despite the issuance of 'great danger' signal number ten (Raju, 2019; ReliefWeb, 2021; Alam et al., 2023). It seems that BMD took a safer and conservative decision during TC Fani by issuing signal no. 7, nevertheless, this cannot be considered effective decision-making. While such a complex decision-making process can certainly be improved by quantifying the uncertainty through an ensemble multi-scenario forecast, incorporating Pareto optimality can further maximize the benefits of it.

Pareto optimal solution provides an effective first filter to identify ensemble multi-scenario surge forecasts. This information can be presented visually to enhance the understanding of the uncertainty in the forecast. The median of Pareto optimal solutions could be utilized given a series of worst/minimum surge estimations for a specific location by several ensemble members. For example, Figure 5b identifies four worst scenarios for Harumi in Tokyo Bay where the median surge level is 150 cm. Although we stress the importance of diversity in trade-offs surge outcomes, a certain scenario (e.g., median of Pareto optimal solutions) can be given more weight depending on the values of society and decision-makers. In the decision-making process, a user-defined acceptable level of uncertainty or reference surge height (e.g., 25-year return period of surge) can be set for a specific location (e.g., Harumi). The forecaster can then determine if the height of the Pareto optimal solution exceeds this acceptable level. Subsequently, a relevant warning signal can be issued in a forecast horizon (e.g., 39-h lead time). The warning signal can be tailored to a specific location, if a diversity in the trade-off between surge outcomes exists among Pareto frontiers. For example, Pareto optimal solutions in Figure 5b predicted that TC Hagibis will bring worst surge level as maximum as 160 cm in the inner Tokyo Bay, requiring the issuance of an emergency warning, closing flood gates, and large-scale evacuation of the coastal population living below the storm surge height. On the other hand, those living along open coastlines are advised to stay indoors as the predicted worst surge level (90 cm) does not meet the criteria for issuing an emergency warning. Once this forecast becomes available, decision-makers (e.g., emergency managers) can start evacuation planning based on the forecasted worst TC track, wind intensity, and peak surge height. For instance, TC track that forecasted to bring severe storm surges of 160 cm in the inner Tokyo Bay, would make landfall 70 km west of the central bay axis with a landfall V_{max} of 75 kt, which is stronger by 11-kt from the historical mean (64-kt; (Islam et al., 2022)). Furthermore, it is projected to be twice as large as the historical average (65 NM) and move at a slower speed by 9 km/h compared to the average translation speed (41 km/h) in Tokyo Bay. The forecasted landfall location and meteorological conditions of the worst TC indicate that Tokyo Bay would be situated in the destructive right-side semicircle of the TC track, resulting in prolonged exposure to severe storm surges and strong winds. Emergency managers can utilize this information to disseminate surge warnings to residents and commence evacuation procedures with a 39-hour lead time. This evacuation can be done by dividing coastal regions into different zones depending on their vulnerability. Although disaster planning is not so straightforward as explained here, our proposed ensemble-based storm surge multi-scenario analysis is expected to motivate forecasters and risk management practitioners to explore new ways to assess storm surge hazards and reduce the associated risk.

Finally, we acknowledge that this study focuses exclusively on peak surge height while determining total sea water level that includes the influence of astronomic tide, wave set-up, and river discharge are also critical and can be done utilizing a full physical numerical model. Furthermore, several algorithms are currently available to determine a Pareto frontier. We encourage researchers from multiple disciplines to build on our approach to help us reach an improved understanding of Pareto optimality based multi-scenario analysis.

Acknowledgment

The authors declare no financial or conflict of interest. This work is supported by JST FOREST program (grant no. JPMJFR205Q) and JST Moonshot R&D project (grant no. JPMJMS2281). We acknowledge the support by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) through the Program for Promoting Researches on the Supercomputer Fugaku JPMXP1020351142 'Large Ensemble Atmospheric and Environmental Prediction for Disaster Prevention and Mitigation' (hp210166).

Open Research

Observed storm surge data can be downloaded from the JMA (https://www.data.jma.go.jp/kaiyou/db/tide/genbo/index.php) and JODC (https://jdoss1.jodc.go.jp/vpage/tide.html) websites. Predicted tide data can be obtained from the JMA (https://www.data.jma.go.jp/kaiyou/db/tide/suisan/index.php) website. TC best track data can be derived from the JMA (https://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/trackarchives.html) website. Ensemble forecast data may be available upon request.

References

Alam, Md. S., Chakraborty, T., Hossain, Md. Z., & Rahaman, K. R. (2023). Evacuation dilemmas of coastal households during cyclone Amphan and amidst the COVID-19 pandemic: a study of the Southwestern region of Bangladesh. *Natural Hazards*, *115*(1), 507–537. https://doi.org/10.1007/s11069-022-05564-9

Bangladesh Meteorological Department. (2021). Bangladesh Meteorological Department. Retrieved March 12, 2021, from http://live.bmd.gov.bd/

Bernier, N. B., & Thompson, K. R. (2015). Deterministic and ensemble storm surge prediction for Atlantic Canada with lead times of hours to ten days. *Ocean Modelling*, *86*, 114–127. https://doi.org/10.1016/j.ocemod.2014.12.002

Duc, L., Saito, K., & Hotta, D. (2020). Analysis and design of covariance inflation methods using inflation functions. Part 1: Theoretical framework. *Quarterly Journal of the Royal Meteorological Society*, *146*(733), 3638–3660. https://doi.org/10.1002/qj.3864

Duc, L., Kawabata, T., Saito, K., & Oizumi, T. (2021). Forecasts of the July 2020 Kyushu Heavy Rain Using a 1000-Member Ensemble Kalman Filter. *Sola*, *17*, 41–47. https://doi.org/10.2151/sola.2021-007

Flecker, A. S., Shi, Q., Almeida, R. M., Angarita, H., Gomes-Selman, J. M., García-Villacorta, R., et al. (2022). Reducing adverse impacts of Amazon hydropower expansion. *Science*, *375*(6582), 753–760. https://doi.org/10.1126/science.abj4017

Flowerdew, J., Mylne, K., Jones, C., & Titley, H. (2013). Extending the forecast range of the UK storm surge ensemble. *Quarterly Journal of the Royal Meteorological Society*, *139*(670), 184–197. https://doi.org/10.1002/qj.1950

Greenslade, D., Freeman, J., Sims, H., Allen, S., Colberg, F., Schulz, E., et al. (2017). An operational coastal sea level forecasting system. In *Australasian Coasts & Ports 2017: Working with Nature* (pp. 514–520). https://doi.org/10.3316/informit.925870785579893

Hasegawa, H., Kohno, N., Higaki, M., & Itoh, M. (2017). *Upgrade of JMA's Storm Surge Prediction for the WMO Storm Surge Watch Scheme (SSWS)*. RSMC Tokyo - Typhoon Center. Retrieved from https://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/techrev/text19-2.pdf

Hasegawa, J., Takagi, Y., & Tsuboi, Y. (2017). *Key techniques for Japan's Risk-based Warning Services for Heavy-rain Related Disasters*. Retrieved from http://www.typhooncommittee.org/50th/TECO/Parallel1A_2.pdf

Islam, M. R., & Takagi, H. (2020a). Statistical significance of tropical cyclone forward speed on storm surge generation: retrospective analysis of best track and tidal data in Japan. *Georisk*. https://doi.org/10.1080/17499518.2020.1756345

Islam, M. R., & Takagi, H. (2020b). Typhoon parameter sensitivity of storm surge in the semi-enclosed Tokyo Bay. *Frontiers of Earth Science*, *14*(3), 553–567. https://doi.org/10.1007/s11707-020-0817-1

Islam, M. R., Takagi, H., Anh, L. T., Takahashi, A., & Bowei, K. (2018). 2017 Typhoon Lan reconnaissance field survey in coasts of Kanto region, Japan. *Journal of Japan Society of Civil Engineers, Ser. B3 (Ocean Engineering)*, *74*(2), I_593-I_598. https://doi.org/10.2208/jscejoe.74.i_593

Islam, M. R., Lee, C.-Y., Mandli, K. T., & Takagi, H. (2021). A new tropical cyclone surge index incorporating the effects of coastal geometry, bathymetry and storm information. *Scientific Reports*, *11*(1), 16747. https://doi.org/10.1038/s41598-021-95825-7

Islam, M. R., Satoh, M., & Takagi, H. (2022). Tropical Cyclones Affecting Japan Central Coast and Changing Storm Surge Hazard since 1980. *Journal of the Meteorological Society of Japan. Ser. II*, *100*(3), 493–507. https://doi.org/10.2151/jmsj.2022-024

Japan Oceanographic Data Center. (2020). 500m Gridded Bathymetry Data. Retrieved January 8, 2021, from https://www.jodc.go.jp/jodcweb/JDOSS/infoJEGG.html

Japan Oceanography Data Center. (2021). Tide Stations. Retrieved January 8, 2021, from https://jdoss1.jodc.go.jp/vpage/tide.html

JMA. (2021). Japan Meteorological Agency | RSMC Tokyo - Typhoon Center | Best Track Data. Retrieved March 13, 2021, from https://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/trackarchives.html

JMA. (2022). Tidal observation data. Retrieved June 30, 2022, from https://www.data.jma.go.jp/gmd/kaiyou/db/tide/genbo/index.php

JMA. (2023). Japan Meteorological Agency｜Numerical Weather Prediction Activities. Retrieved March 10, 2023, from https://www.jma.go.jp/jma/en/Activities/nwp.html

Kobayashi, K., Duc, L., Apip, Oizumi, T., & Saito, K. (2020). Ensemble flood simulation for a small dam catchment in Japan using nonhydrostatic model rainfalls – Part 2: Flood forecasting using 1600-member 4D-EnVar-predicted rainfalls. *Natural Hazards and Earth System Sciences*, *20*(3), 755–770. https://doi.org/10.5194/nhess-20-755-2020

Kochenderfer, M. J., & Wheeler, T. A. (2019). Multiobjective Optimization. In *Algorithms for Optimization* (pp. 211–233). The MIT Press.

Kohno, N., Dube, S. K., Entel, M., Fakhruddin, S. H. M., Greenslade, D., Leroux, M.-D., et al. (2018). Recent Progress in Storm Surge Forecasting. *Tropical Cyclone Research and Review*, *7*(2), 128–139. https://doi.org/10.6057/2018TCRR02.04

Kristensen, N. M., Røed, L. P., & Sætra, Ø. (2022). A forecasting and warning system of storm surge events along the Norwegian coast. *Environmental Fluid Mechanics*. https://doi.org/10.1007/s10652-022-09871-4

Liu, Y., Liu, Z., Lamont, S., Hirpa, F., Zhang, Y., Li, T., et al. (2021). *Simulation of Storm Surge during Typhoon Hagibis, 2019 with a Large-scale Japan-wide Coastal Hydrodynamic Model*. Presented at the AGU Fall Meeting 2021, New Orleans, LA, USA. Retrieved from https://ui.adsabs.harvard.edu/abs/2021AGUFMOS33B..05L

Lombardi, G., Ceppi, A., Ravazzani, G., Davolio, S., & Mancini, M. (2018). From Deterministic to Probabilistic Forecasts: The 'Shift-Target' Approach in the Milan Urban Area (Northern Italy). *Geosciences*, *8*(5), 181. https://doi.org/10.3390/geosciences8050181

Ma, W., Ishitsuka, Y., Takeshima, A., Hibino, K., Yamazaki, D., Yamamoto, K., et al. (2021). Applicability of a nationwide flood forecasting system for Typhoon Hagibis 2019. *Scientific Reports*, *11*(1), 10213. https://doi.org/10.1038/s41598-021-89522-8

Mel, R., & Lionello, P. (2014). Storm Surge Ensemble Prediction for the City of Venice. *Weather and Forecasting*, *29*(4), 1044–1057. https://doi.org/10.1175/WAF-D-13-00117.1

Needham, H. F., Keim, B. D., & Sathiaraj, D. (2015). A review of tropical cyclone-generated storm surges: Global data sources, observations, and impacts. *Reviews of Geophysics*, *53*(2), 545–591. https://doi.org/10.1002/2014RG000477

NHC. (n.d.). Sea, Lake, and Overland Surges from Hurricanes (SLOSH). Retrieved January 19, 2023, from https://www.nhc.noaa.gov/surge/slosh.php

Raju, M. N. A. (2019, November 12). Cyclone Bulbul: Views from the ground. Retrieved February 8, 2023, from https://www.thedailystar.net/opinion/environment/news/cyclone-bulbul-views-the-ground-1825732

Rappaport, E. N. (2014). Fatalities in the United States from Atlantic Tropical Cyclones: New Data and Interpretation. *Bulletin of the American Meteorological Society*, *95*(3), 341–346. https://doi.org/10.1175/BAMS-D-12-00074.1

ReliefWeb. (2021). Counting the Costs: Cyclone Amphan One Year On - Bangladesh | ReliefWeb. Retrieved February 9, 2023, from https://reliefweb.int/report/bangladesh/counting-costs-cyclone-amphan-one-year

Saito, K., Fujita, T., Yamada, Y., Ishida, J., Kumagai, Y., Aranami, K., et al. (2006). The Operational JMA Nonhydrostatic Mesoscale Model. *Monthly Weather Review*, *134*(4), 1266–1298. https://doi.org/10.1175/MWR3120.1

Sawada, Y., Kanai, R., & Kotani, H. (2022). Impact of cry wolf effects on social preparedness and the efficiency of flood early warning systems. *Hydrology and Earth System Sciences*, *26*(16), 4265–4278. https://doi.org/10.5194/hess-26-4265-2022

Sharma, M., Lamers, A., Devi, S., Schumacher, A., Earl-Spurr, C., Fritz, C., et al. (2022). *Forecasting Tropical Cyclone Hazards and Impacts*. 10th International Workshop on Tropical Cyclones (IWTC-10). Retrieved from https://community.wmo.int/en/iwtc-10-reports

Shimozono, T., Tajima, Y., Kumagai, K., Arikawa, T., Oda, Y., Shigihara, Y., et al. (2020). Coastal impacts of super typhoon Hagibis on Greater Tokyo and Shizuoka areas, Japan. *Coastal Engineering Journal*, *62*(2), 129–145. https://doi.org/10.1080/21664250.2020.1744212

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T. M., Hewson, T. D., et al. (2016). The TIGGE Project and Its Achievements. *Bulletin of the American Meteorological Society*, *97*(1), 49–67. https://doi.org/10.1175/BAMS-D-13-00191.1

Takagi, H., Xiong, Y., & Furukawa, F. (2018). Track analysis and storm surge investigation of 2017 Typhoon Hato: were the warning signals issued in Macau and Hong Kong timed appropriately? *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, *12*(4), 297–307. https://doi.org/10.1080/17499518.2018.1465573

Takagi, H., Anh, L. T., Islam, R., & Hossain, T. T. (2022). Progress of disaster mitigation against tropical cyclones and storm surges: a comparative study of Bangladesh, Vietnam, and Japan. *Coastal Engineering Journal*, *0*(0), 1–15. https://doi.org/10.1080/21664250.2022.2100179

Titley, H. A., Yamaguchi, M., & Magnusson, L. (2019). Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey. *Tropical Cyclone Research and Review*, *8*(3), 166–180. https://doi.org/10.1016/j.tcrr.2019.10.005

Tommy. (2021). paretoset (Version 1.2.0). Retrieved from https://github.com/tommyod/paretoset

Wilson, K. A., Heinselman, P. L., Skinner, P. S., Choate, J. J., & Klockow-McClain, K. E. (2019). Meteorologists' Interpretations of Storm-Scale Ensemble-Based Forecast Guidance. *Weather, Climate, and Society*, *11*(2), 337–354. https://doi.org/10.1175/WCAS-D-18-0084.1

Yamaguchi, M., Vitart, F., Lang, S. T. K., Magnusson, L., Elsberry, R. L., Elliott, G., et al. (2015). Global Distribution of the Skill of Tropical Cyclone Activity Forecasts on Short- to Medium-Range Time Scales. *Weather and Forecasting*, *30*(6), 1695–1709. https://doi.org/10.1175/WAF-D-14-00136.1

Assessing Storm Surge Multi-Scenarios based on Ensemble Tropical Cyclone Forecasting

Md. Rezuanul Islam[1*], Le Duc[1,2], and Yohei Sawada[1,2]

[1]Institute of Engineering Innovation, The University of Tokyo, Tokyo, 113-0032, Japan

[2]Meteorological Research Institute, Japan Meteorological Agency

Corresponding author: Md. Rezuanul Islam (fahiemislam@gmail.com)

Key Points:

- The potential of ensemble tropical cyclone forecasting for assessing storm surge multi-scenarios is shown.
- Pareto optimized solutions from an ensemble storm surge forecast can efficiently illustrate potential worst and minimum storm surge scenarios.
- Analyses of meteorological variables of ensemble members in Pareto frontiers help understand the impact of a tropical cyclone on predicted storm surge multi-scenarios.

Abstract

Ensemble forecasting is a promising tool to aid in making informed decisions against risks of coastal storm surges. Although tropical cyclone (TC) ensemble forecasts are commonly used in operational numerical weather prediction systems, their potential for disaster prediction has not been maximized. Here we present a novel, efficient, and practical method to utilize a large ensemble forecast of 1000 members to analyze storm surge scenarios toward effective decision making such as evacuation planning and issuing surge warnings. We perform the simulation of TC Hagibis (2019) using the Japan Meteorological Agency's (JMA) non-hydrostatic model. The simulated atmospheric predictions were utilized as inputs for a statistical surge model named the Storm Surge Hazard Potential Index (SSHPI) to estimate peak surge heights along the central coast of Japan. We show that Pareto optimized solutions from an ensemble storm surge forecast can describe potential worst (maximum) and optimum (minimum) storm surge scenarios while exemplifying a diversity of trade-off surge outcomes among different coastal places. For example, some of the Pareto optimized solutions that illustrate worst surge scenarios for inner bay locations are not necessarily accountable for bringing severe surge cases in open coasts. We further emphasize that an in-depth evaluation of Pareto optimal solutions can shed light on how meteorological variables such as track, intensity, and size of TCs influence the worst and optimum surge scenarios, which is not clearly quantified in current multi-scenario assessment methods such as those used by JMA/National Hurricane Center in the United States.

Plain Language Summary

Ensemble forecasting generates multiple predictions of a weather event with various possible outcomes based on varying initial conditions, model parameters, and physics. The potential of ensemble tropical cyclone (TC) forecasting for assessing storm surge multi-scenarios has largely been overlooked previously. Enhanced analysis can unlock and maximize the benefit of ensemble forecasting. This study simulated an extremely large ensemble (=1000 members) to reforecast past TC Hagibis which hit the central coast of Japan in 2019 and utilized the results to predict storm surges. We propose that Pareto optimality can identify good ensemble members that reasonably represent potential worst/minimum storm surge scenarios, meaning no other ensemble members can represent better than those. Comprehensive analyses of Pareto members can give forecasters and decision makers a better understanding of how the predicted track, wind intensity, and size of a TC can impact the worst and best storm surge scenarios. This type of analysis is expected to improve the planning of evacuations and the issuing of storm surge warnings.

1 Introduction

Since 1737, 29 coastal storm surge events have claimed at least 5,000 people globally. Two of these events happened in the 21st century and ranked as two of the five worst coastal disasters in the running millennium (Needham et al., 2015; Takagi et al., 2022). Rappaport (2014) has shown that 49% of tropical cyclone (TC)-induced deaths are directly attributed to storm surges. Hence, it is crucially important to improve the understanding of storm surge and their associated risk as it is among the deadliest and most destructive natural disasters.

In recent years, forecast services have likely reduced TC-induced deaths relative to historical standards. For example, several countries have already adopted a dynamical TC ensemble prediction system (EPS) to capture forecast uncertainties and reduce sampling errors in the three-dimensional meteorological simulation (Sharma et al., 2022). Numerical weather prediction centers such as Japan Meteorological Agency (JMA), National Centers for Environmental Prediction in the United States (US), European Centre for Medium-Range Weather Forecasts generate TC track forecasts from their ensemble forecast models and utilize them in their operational settings (Swinbank et al., 2016). Yamaguchi et al. (2015) have shown that EPS can provide skillful guidance of TC genesis forecasts with a forecast lead time extending to two weeks in seven TC basins. Nevertheless, there is a great potential to maximize the use of this EPS not only in TC activity (e.g., track, intensity) forecast but also in forecasting hazards (e.g., storm surge), aiding end users to be prepared better before the dangerous situation (Kobayashi et al., 2020; Duc et al., 2021).

Titley et al. (2019) have recently conducted a questionnaire survey at operational TC forecast centers worldwide to understand the current and potential use of EPS in operational TC forecasting. They reported that over 90% of respondents used an ensemble forecast for TC track forecast, followed by genesis and intensity forecasts. In contrast, less than 10% of surveyed forecasters use ensemble products for hazard (e.g., storm surge) forecasting. Deterministic forecasts are often used for hazard forecasting as it is produced using the best available TC data and unperturbed models. In some cases, ensemble mean (e.g., track and intensity of TC) is used as inputs for hazard forecast to compare the result with the deterministic forecasts, although the full use of EPS in hazard forecasting remains challenging (Titley et al., 2019). A lack of detailed analysis of ensemble members (beyond ensemble mean/median analysis) and less technical expertise on ensemble-based hazard forecasts hinder its' application among hazard forecasters. Wilson et al. (2019) reported that a deterministic mindset resulted in tendencies to modify understanding of probabilistic concepts when presented with different meteorological variables. Furthermore, local authorities responsible for hazard forecasting avoid EPS information as citizens and emergency managers habitually trust a single forecast only, and they are not sufficiently educated to deal with the probabilistic prediction (Lombardi et al., 2018). These findings highlight that ensemble-based hazard (e.g., storm surge) forecast is unfamiliar in disaster risk management communities.

Notwithstanding the challenges mentioned above, ensemble surge prediction system (ESPS) has recently received considerable attention from both the research and operational communities. For instances, Flowerdew et al. (2013), Greenslade et al. (2017), and Kristensen et al. (2022) have successfully developed and evaluated the performance of an operational ESPS for United Kingdom, Australia, and Norway, respectively. Along the coastline of Canada, it was found that 20-member ESPS could reasonably estimate both the uncertainty in peak surge height and timing of surge events resulting from imperfectly forecast atmospheric conditions six days before (Bernier & Thompson, 2015). A 50-member ensemble simulation of 10 surge events during 2010 in Venice by Mel & Lionello (2014) has shown that the distribution of maximum sea level is acceptably realistic with respect to the deterministic forecast. They also found that the uncertainty became its maximum during storm surge peaks and increased linearly with the forecasting lead time. Although these ensemble simulation studies paved the way for a robust surge hazard assessment over a single forecast-based assessment, they considered ensemble TC forecast information only for developing and evaluating the performance (skill and accuracy) of an ESPS. In addition to quantifying the uncertainty of surge height, ensemble-based storm surge multi-scenario (e.g., worst/optimum case) analysis is equally important, aiding disaster risk managers in evacuation planning (Kohno et al., 2018).

To the best of our knowledge, the potential of ensemble TC forecasting for assessing storm surge multi-scenarios has largely been overlooked previously. However, recent developments have seen the introduction of multi-scenario storm surge predictions, such as the worst-case scenario from six typical TC tracks by the JMA (H. Hasegawa et al., 2017) and the maximum storm tide height by the National Hurricane Center in the US (NHC, n.d.). These worst-case scenarios are composite products, representing the maxima among all scenarios. Therefore, it is possible that the worst-case values for two adjacent locations may have come from two different ensemble TC track run. Therefore, the users (e.g., emergency managers) cannot understand which forecasted TC track or which combination of forecasted TC meteorological variables (track, intensity, size, translation speed) may trigger the worst surge scenario for a particular location based on a composite product. This can make it difficult for decision-makers to determine the appropriate level of storm surge warning and evacuation orders. In addition, storm surge is spatially heterogeneous because of its' dependency on a TC characteristic and coastal geometry. It is entirely plausible that the worst case scenario may not occur everywhere within a forecasted TC threat zone (Islam & Takagi, 2020a, 2020b). If the decision makers in cities/tourist districts with highly valuable economies issue a higher warning level without any concrete understating over a worst event, they will inevitably suffer significant economic losses because of false alarming (in case the area has not affected by worst storm surge) and eventually can lower citizens trust over official warning (Sawada et al., 2022; Takagi et al., 2018).

Here we present Pareto optimality - a novel way of assessing storm surge multi-scenarios based on ensemble TC forecasts. Our approach is more advanced than existing assessments. We employed a multi-objective function to determine possible worst/optimum cases to quantify the hazards in a large region. Our approach involved a comprehensive analysis of Pareto optimal solutions in understanding the combination of forecasted TC meteorological variables - such as track, intensity, size, and translation speed of TC - that could result in the worst/optimum surge scenario. We utilized an extremely large ensemble (=1000 member) forecasts of TC Hagibis that made landfall in central Japan in 2019. Our Pareto-based optimal solutions provide an instantaneous overall assessment of storm surge multi-scenarios without any computational burdens. The proposed method will allow forecasters to predict storm surge multi-scenarios harnessing ensemble TC forecasts efficiently and help emergency responders as means of quantifying surge hazards effectively.

2 Data and Methods

2.1 TC Hagibis and ensemble forecast

TC Hagibis in 2019, one of the most destructive and deadliest TC that hit Japan in decades (Shimozono et al., 2020; Ma et al., 2021), has been chosen to demonstrate our multi-scenario storm surge assessment. Hagibis was formed in the western North Pacific Ocean on 2 October 2019 and made landfall in central Japan on 12 October 2019 (around 0900 UTC), as depicted in Figure 1. At the landfall time, its maximum wind speeds sustained at 80 kt. This combined with heavy rainfall, resulted in high storm surges and severe flooding in the area (Shimozono et al., 2020; Ma et al., 2021; JMA, 2021).

The atmospheric ensemble forecasts of TC Hagibis were obtained by running JMA's former operational limited-area model called NHM (non-hydrostatic model; Saito et al., 2006). The integration domain (see Figure S1) had a grid spacing of 5 km consisting of 817 × 661 horizontal grid points and 50 vertical levels. Boundary conditions were interpolated to the NHM domain from JMA's global model forecasts and the forecast perturbations of JMA's operational one-week EPS.

Since we used NHM for all forecast members, the only source of uncertainty stemmed from initial conditions. This uncertainty is encapsulated in error covariances of current atmospheric states (analysis error covariances), estimated using a data assimilation system. An ensemble Kalman filter (EnKF) was employed to sample from these error covariances and generate an analysis ensemble. While operational forecast centers generally use around 100 ensemble members, a state-of-the-art data assimilation system with 1000 ensemble members, called the four-dimensional variational-ensemble assimilation technique (4DEnVAR), was utilized in this study (Kobayashi et al., 2020). Our 4DEnVAR system only applied horizontal localization, with the horizontal localization length scales derived from the JMA's operational four-dimensional variational assimilation system's climatological horizontal correlation length scales. This helped to remove sampling noise in estimating forecast error covariances and maintain the coherent vertical structure between atmospheric fields, which is critical in predicting tropical cyclones. As the ensemble member count was large (=1000), localization was relaxed by retaining vertical correlations and removing horizontal correlations at distant locations (Duc et al., 2021).

Unlike EnKF, EnVAR solely estimates the means of analysis ensembles and not the analysis ensembles themselves, even though this method heavily relies on forecast ensembles to estimate these means. To solve this issue, a common approach is to run a separate EnKF in parallel to generate analysis ensembles. However, our 4DEnVAR system is unique in that an EnKF is not necessary. Instead, the same EnVAR program was used to generate analysis perturbations, as suggested in the context of inflation functions (Duc et al., 2020), where we demonstrated that using quadratic inflation functions implies using the Kalman gain to generate analysis perturbations. Using the same program for analysis means and analysis perturbations is essential because it ensures consistency between the two when the same background error covariance, localization, and observations are utilized in both cases. The assimilation system commenced at 00UTC on 7 October 2019, with a 3-hour assimilation cycle and continued until 18:00UTC 10 October 2019. The final analysis ensemble was then used as initial conditions for 39h forecasts with NHM. The assimilation domain was chosen the same as the forecast domain in Figure S1 and we assimilated all routine observations obtained from JMA's database. Here, we opted for a 39h forecast horizon because JMA's operational Meso-scale Ensemble Prediction System (MEPS) also generates 39h forecasts at 6-hour intervals (JMA, 2023).

2.2 Ensemble storm surge forecast

We used storm surge hazard potential index (SSHPI; eq. 1), a statistical model to compute peak storm surge height. While the coastal engineers and ocean modelers are interested in the forecast of storm surge hydrograph, most of the decision makers responsible for issuing surge warning and relief measures have a primary interest in the value of predicted peak surge height. The SSHPI uses meteorological variables sensitive to storm surge, including TC intensity (V_{max}), size (radius of 50-kt wind; R_{50}), and translation speed (S). In addition, the SSHPI considers coastal geometry (a = 1 = open coasts and a = −1 = bays), landfall location sensitivity (D_L), and regional scale bathymetry (L_{30}). The SSHPI does not incorporate factors associated with wave set-up and astronomic tide to keep the configuration simple. TC Hagibis ensemble forecasts (=1000 member; see Section 2.1), particularly during landfall, was used as meteorological forcing of the SSHPI. We produced corresponding 1000 perturbed surge forecasts with a lead time of 39h. The bathymetry of the target region was obtained from the Japan Oceanographic Data Center (Japan Oceanographic Data Center, 2020). The effectiveness of the SSHPI for predicting peak surge hazard potential was discussed in Islam et al. (2021, 2022). The formulation of the SSHPI is the following:

$$\text{SSHPI} = \left(\frac{V_{max}}{V_{ref}}\right)^2 \left(\frac{R_{50}}{R_{ref}}\right) \left(\frac{S}{S_{ref}}\right)^a \left(\frac{L_{30}}{L_*}\right) (D_L) \qquad (1)$$

$$\frac{R_{50}}{R_{ref}} = \begin{cases} 1.5 & if\ \frac{R_{50}}{R_{ref}} \geq 1.5 \\ \frac{R_{50}}{R_{ref}} & if\ 0.5 < \frac{R_{50}}{R_{ref}} < 1.5 \\ 0.5 & if\ \frac{R_{50}}{R_{ref}} \leq 0.5 \end{cases};\quad \left(\frac{S}{S_{ref}}\right)^a = \begin{cases} 1.5 & if\ \left(\frac{S}{S_{ref}}\right)^a \geq 1.5 \\ \left(\frac{S}{S_{ref}}\right)^a & if\ 0.5 < \left(\frac{S}{S_{ref}}\right)^a < 1.5 \\ 0.5 & if\ \left(\frac{S}{S_{ref}}\right)^a \leq 0.5 \end{cases};\quad \frac{L_{30}}{L_*} = \begin{cases} \frac{L_{30}}{L_*},\ if\ \frac{L_{30}}{L_*} \geq 1 \\ 1,\ if\ \frac{L_{30}}{L_*} \leq 1 \end{cases}$$

$$D_L = \begin{cases} 1 & if\ the\ surge\ estimated\ point\ falls\ right\ side\ of\ TC\ track\ and\ x \leq 20 \\ & OR \\ & if\ the\ surge\ estimated\ point\ falls\ left\ side\ of\ TC\ track\ and\ x \leq 10 \\ 1 - \dfrac{0.03(x-20)}{20} & if\ the\ surge\ estimated\ point\ falls\ right\ side\ of\ TC\ track\ and\ x > 20 \\ 1 - \dfrac{0.05(x-10)}{10} & if\ the\ surge\ estimated\ point\ falls\ left\ side\ of\ TC\ track\ and\ x > 10 \end{cases}$$

V_{ref}, R_{ref}, and S_{ref}, are reference constants as follows: 50-kt equivalents of the tropical storm category, 95 NM (historical mean R_{50} at the time of landfall in Japan mainland), and 35 km/h (historical mean S at the time of landfall in Japan), respectively (Islam et al., 2021). L_{30} is the horizontal distance (km) between the shoreline and the 30-m depth contour. L^* was chosen to be 10 km. D_L is defined by different expressions depending on the surge estimated points' (e.g., tidal station) position (right/left) respective to the TC track and horizontal distance (x in km) between the TC landfall location and a surge estimated point. Compared to V_{max}, the upper and lower bounds of R_{50}, S, and L_{30} in eq. 1 restrict their contribution in generating surge hazards and, thus, prevents discrete jumps in the SSHPI.

Figure 1 shows a storm surge modeling domain and the position of tide gauges used for validating surge model and predicting surge hazards in this study. There are two domains, covering Tokyo Bay and Suruga Bay individually. Each domain has tide gauges located both in inner bays and open coasts. It should be noted that the tide gauges chosen for this study are the only stations that possess recorded (historical) storm surge data, which is kept by JMA (JMA, 2022) and Japan Coast Guard (Japan Oceanography Data Center, 2021). The empirical relationship for expected storm surge in each tide gauge was determined in Islam et al. (2021, 2022) by drawing a line of best fit through the historical surge data and the SSHPI and thus, used for the surge forecasts in this study.

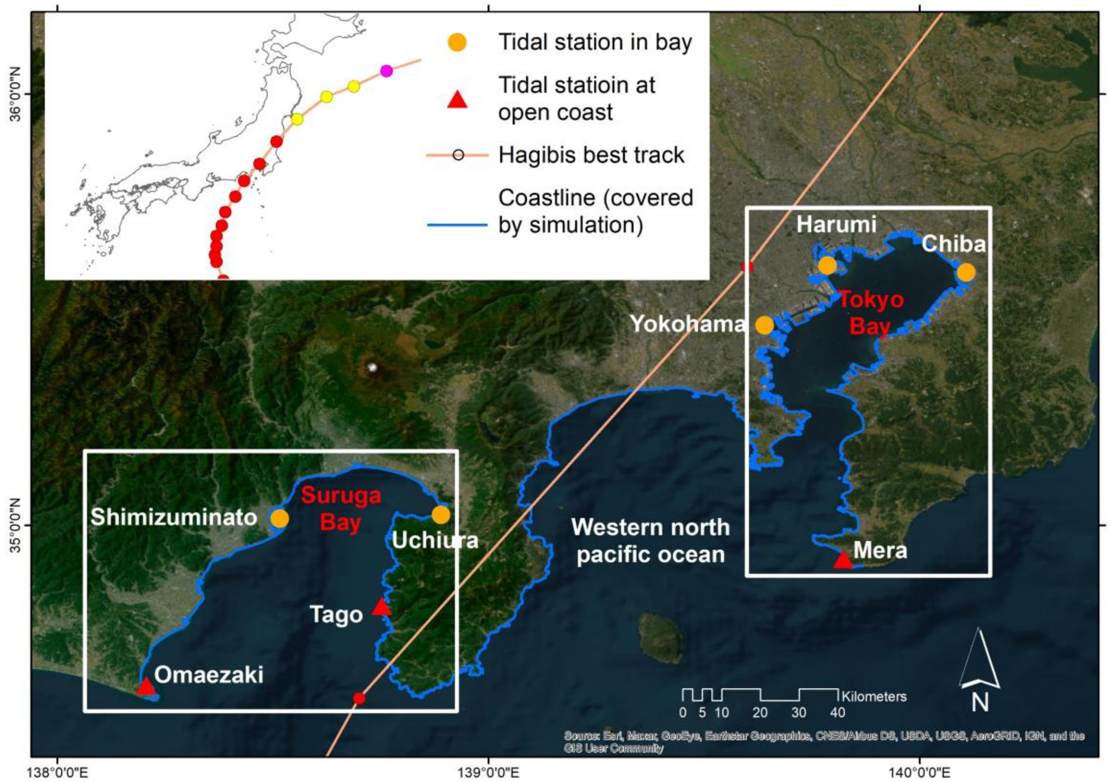

Figure 1. Domain of the storm surge forecasts model and the locations of the tide gauges used for model validation and surge forecasts.

2.3 Pareto optimality and assessing storm surge multi-scenarios

It is unrealistic to anticipate a "nice" forecast scenario that accurately predicts the exact intensity of a hazard at all locations within a given domain for a particular condition (e.g., worst/optimum). An improved forecast at one location is usually accompanied by a deterioration of forecast at another location and vice versa. The best we can do is to quantify the trade-off between different objectives. Here, we conducted multi-objective optimization to select ensemble forecast members (among 1000 ensemble forecasts; see Section 2.1 and 2.2) that reasonably characterize the potential worst/optimum storm surge case for a particular location (e.g., Tokyo Bay) by computing the Pareto frontier. The Pareto frontier captures the trade-offs between objectives. It is the set of all Pareto-optimum solutions where a single Pareto optimal solution denotes a solution that is not dominated by any other solution (Kochenderfer & Wheeler, 2019).

In this study, we analyzed a subset of 1000 forecasted surge scenarios for each tide gauge in a specific domain, referred to as solution z. Each scenario is evaluated based on d objectives, represented by the values $y^1(z), y^2(z), \ldots\ldots, y^d(z)$. As an example, we considered a scenario in Tokyo Bay where the objectives are to maximize the forecasted peak surge at four tide gauges: Harumi, Chiba, Yokohama, and Mera (as shown in Figure 1). This objective function considers the potential worst-case scenario in Tokyo Bay for a set of 1000 ensemble surge forecasts. We compared the given two solutions z and z', if for every objective i, $y^i(z) \geq y^i(z')$ and the strict inequality holds for at least for one objective, we considered that solution z dominates z'. In other words, if one solution (e.g., ensemble member no. 10) provides 160 cm, 155 cm, 140 cm, 110 cm of forecasted peak surge in Harumi, Chiba, Yokohama, and Mera, respectively, but another solution (e.g., ensemble member no. 20) yields 160 cm, 155 cm, 140 cm, 90 cm of forecasted peak surge for the same tide gauges, then the solution given by ensemble member no. 10 dominates the solution provided by ensemble member no. 20. This is because a smaller storm surge (90 cm) in Mera is undesirable as stated in the objective function. A similar objective function but minimizing the forecasted peak surge at four tide gauges: Harumi, Chiba, Yokohama, and Mera (as shown in Figure 1) was considered to determine ensemble forecasts member that characterizes a potential optimum surge case. In order to select the most appropriate ensemble TC member that may cause the worst surge case in the inner bay, but also results in the optimum surge at the open coast, we further assume that objectives are to be maximized for the inner bay tide gauges (Harumi, Chiba, and Yokohama), but minimized for open coast tide gauges (Mera) at the same time.

The details of the implementation of the algorithm used here are described in Tommy (2021). This algorithm can compute the Pareto frontier for four objectives within a minute, meaning the runtime should be acceptable to any operational hazard forecast settings.

3 Results

3.1 Model evaluation

3.1.1 TC ensemble forecasts validation

From the forecasts of atmospheric fields given by NHM, TC tracks and intensities were detected. Here, TC centers are defined as the average of the mean sea level pressure minima, geopotentials at 850 hPa and 700 hPa. Figure 2(b) shows 1000 track forecasts generated from 1000 initial conditions obtained from the 4DEnVAR data assimilation system, along with the ensemble mean forecast, control forecast, and best track. The ellipses in the figure illustrate uncertainties of the TC centers, which are determined by the forecast error covariances of the TC centers. The arrows show the distance errors between the observations (i.e., the best track) and the ensemble mean. For comparison, the operational 20-member JMA ensemble forecast (MEPS) is included in Figure 2a. As shown in Figure 2, 4DEnVAR outperforms MEPS in terms of track forecasts and its ensemble mean is almost identical to the best track with only minor distance errors.

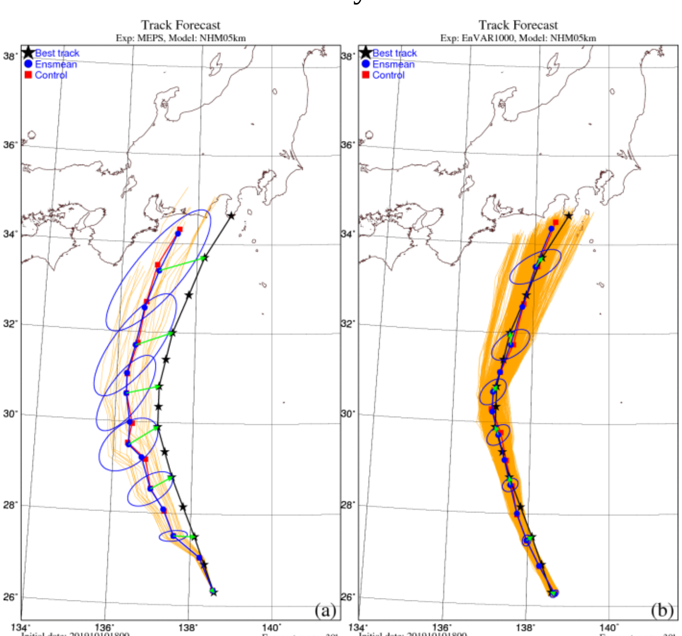

Figure 2. A comparison of 39h (at 1800 UTC on 10 October 2019) ensemble track forecasts for TC Hagibis issued by (a) the operational ensemble prediction system MEPS of JMA and (b) the 4DEnVAR data assimilation system. The ellipses represent forecast error covariances of the TC centers. The arrows denote the distance errors between the ensemble mean and best track.

Figure 3 presents the forecasted intensity of TC Hagibis as indicated by its central pressure. It is evident from the figure that the 4DEnVAR (Figure 3b) surpasses JMA's operational MEPS (Figure 3a) in predicting the intensity. Even though both ensemble forecasts show overestimation of intensity, the tendency of overestimation becomes more apparent with increasing forecast ranges in MEPS (Figure 3a). Despite having a smaller number of ensemble members, MEPS exhibits greater uncertainty in intensity forecast as compared to 4DEnVAR. This can be attributed to the fact that JMA's operational MEPS employs singular vectors to generate initial conditions for ensemble members, which maximizes their spread (JMA, 2023).

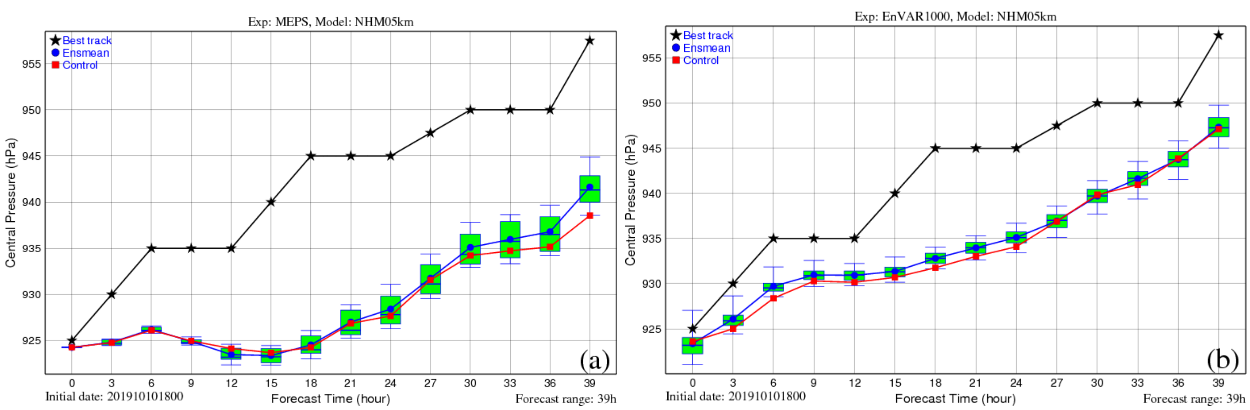

Figure 3. A comparison of 39h (at 1800 UTC on 10 October 2019) ensemble intensity forecast for TC Hagibis issued by (a) the operational ensemble prediction system MEPS of JMA and (b) the 4DEnVAR data assimilation system. The distributions of ensemble intensities are represented as box-and-whisker plots.

3.1.2 Storm surge ensemble forecasts validation

A comparison of the forecasted and measured peak storm surge (Japan Oceanography Data Center, 2021; JMA, 2022) at eight different tide gauges is shown in Figure 4a. It is noted that a significant deviation from the climatological mean surge height is observed at all stations during TC Hagibis. We evaluate JMA best track (JMA, 2021) as an ideal meteorological forcing input as well as our 39h ensemble TC forecasts. Both ensemble median forecasts and best track estimates systematically underestimate the observed peak levels. Nevertheless, the observed peak surge values are enveloped by the full ensemble of the forecasts in most tide gauges, implying that the ensemble spread is large enough to represent the uncertainty in the prediction. The average mean absolute error of the eight stations is 11.3 cm. In the case of Mera and Omaezaki, wave set-up is often the dominant driver for generating storm surges (Islam et al., 2018, 2022), which is not considered in the SSHPI. Therefore, the ensemble surge forecast underestimated observed surges.

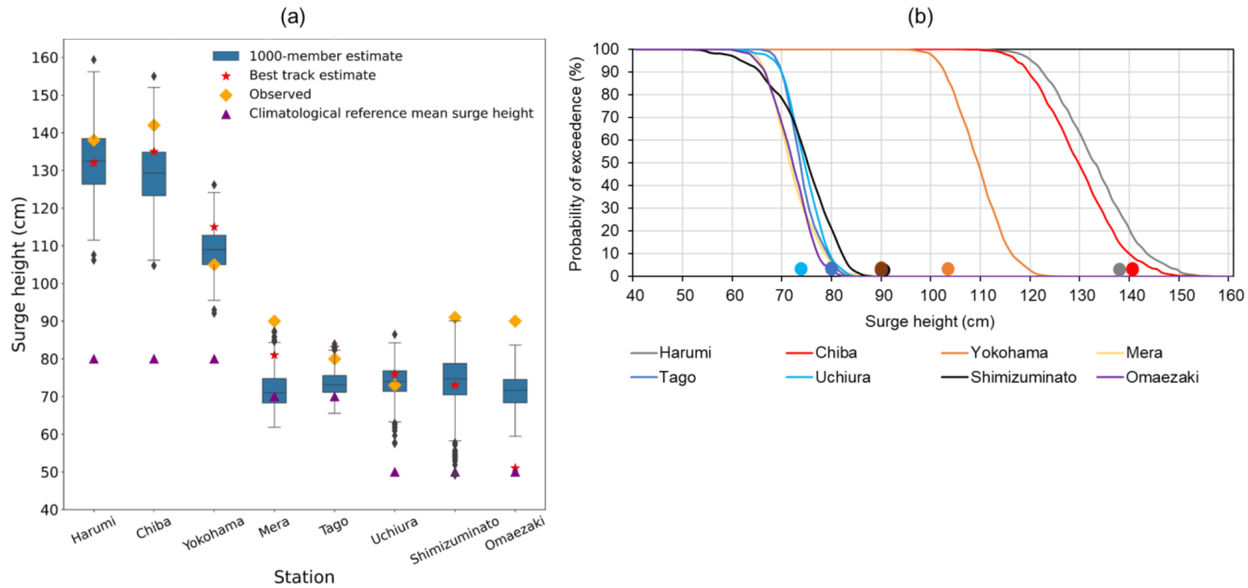

Figure 4. (a) A comparison of the 39h (at 1800 UTC on 10 October 2019) ensemble peak surge forecasts and the observed peak surge height for TC Hagibis at eight tide gauges. (b) Probability of exceeding (Y-axis) for a given storm surge threshold (X-axis) during TC Hagibis landfall time (at 0900 UTC, 12 October 2019). It was estimated by fitting ensemble forecasts empirically. A circle that shares the same color as a line represents the peak surge height recorded at a specific tide gauge.

The probability of surpassing a specific surge threshold during the landfall time of TC Hagibis is illustrated in Figure 4b, as determined by the 39h ESPS. The observed peak surge levels are within the predicted range, except for Mera and Omaezaki. For example, the probability of surpassing the observed peak surge for Harumi (138 cm) is 26.4%. In general, Figure 4 indicates that the SSHPI and its corresponding 39h peak surge forecasts are comparable in quality to those produced by numerical surge models such as Liu et al. (2021). The latter study reported an RMSE of ~10 cm when predicting the maximum total water level in Tokyo Bay during TC Hagibis with a 72h forecast horizon, using atmospheric forcing fields from the Global Forecast System and a hydrodynamic model known as the Semi-implicit Cross-scale Hydroscience Integrated System, which has a nearshore resolution of ~150 m.

3.2 Multi-scenario analysis

3.2.1 Pareto optimal multi-scenarios

The Pareto-optimal frontier, as shown in Figure 5, illustrates a group of solutions that depict the forecasted potential worst and optimum storm surge scenarios for TC Hagibis in Tokyo Bay. The two-dimensional Pareto frontiers (Figure 5a), allow for a straightforward evaluation of trade-offs among the forecasted peak storm surge levels. The results identify the best one or two members from the 1000 TC forecasts to represent the potential worst (Harumi, Chiba: ~155 cm) or optimum (Harumi, Chiba: ~106 cm) surge scenario in the inner Tokyo Bay. Here, the tide gauges (Harumi, Chiba) possess similar coastal geometry features, including bathymetry, and are situated in close proximity to each other (Figure 1). Therefore, the predicted surge response (from 1000 TC forecasts; Figure 5a) between them is almost linear.

Figure 5b reveals a diversity of trade-off surge outcomes in the Pareto frontier for Tokyo Bay. It includes four Pareto frontiers for worst-surge scenarios and six Pareto frontiers for optimum surge scenarios. This diversity is due to the emergence of trade-offs among tide gauges with distinct coastal geometry characteristics, including bathymetry. For instance, some of the identified most severe surge scenarios for open coastlines in Tokyo Bay do not result in high surge levels in the inner bay. Specifically, a Pareto optimal solution in Figure 5b predicts that Mera would experience the worst surge levels of 90 cm (<1% exceedance probability; Figure 5c), while Harumi and Chiba would experience approximately 135 cm (>25% exceedance probability; Figure 5c) of highest surge levels under the same scenario. This storm surge level (~135 cm) in Harumi and Chiba is substantially less than the worst surge levels (~155 cm) predicted by other optimal solutions. Additionally, the surge intensity may vary across Tokyo Bay, depending on the characteristics of the approaching TC and the impact can be much more severe in some places compared to others. For example, several Pareto optimal solutions shown in Figure 5d predict that the inner Tokyo Bay such as Harumi would witness surge levels higher than 150 cm, while it would keep as minimum as 70 cm along the open coastline (e.g., Mera). Owing to such surge incongruence among the coastal locations, creating multiple scenarios for different coastal places can lead to multiple optimal solutions, as seen in Figures 5b and 5d. This emphasizes the importance of considering multiple scenarios when issuing warnings and assessing the risks posed by extreme weather events like storm surges.

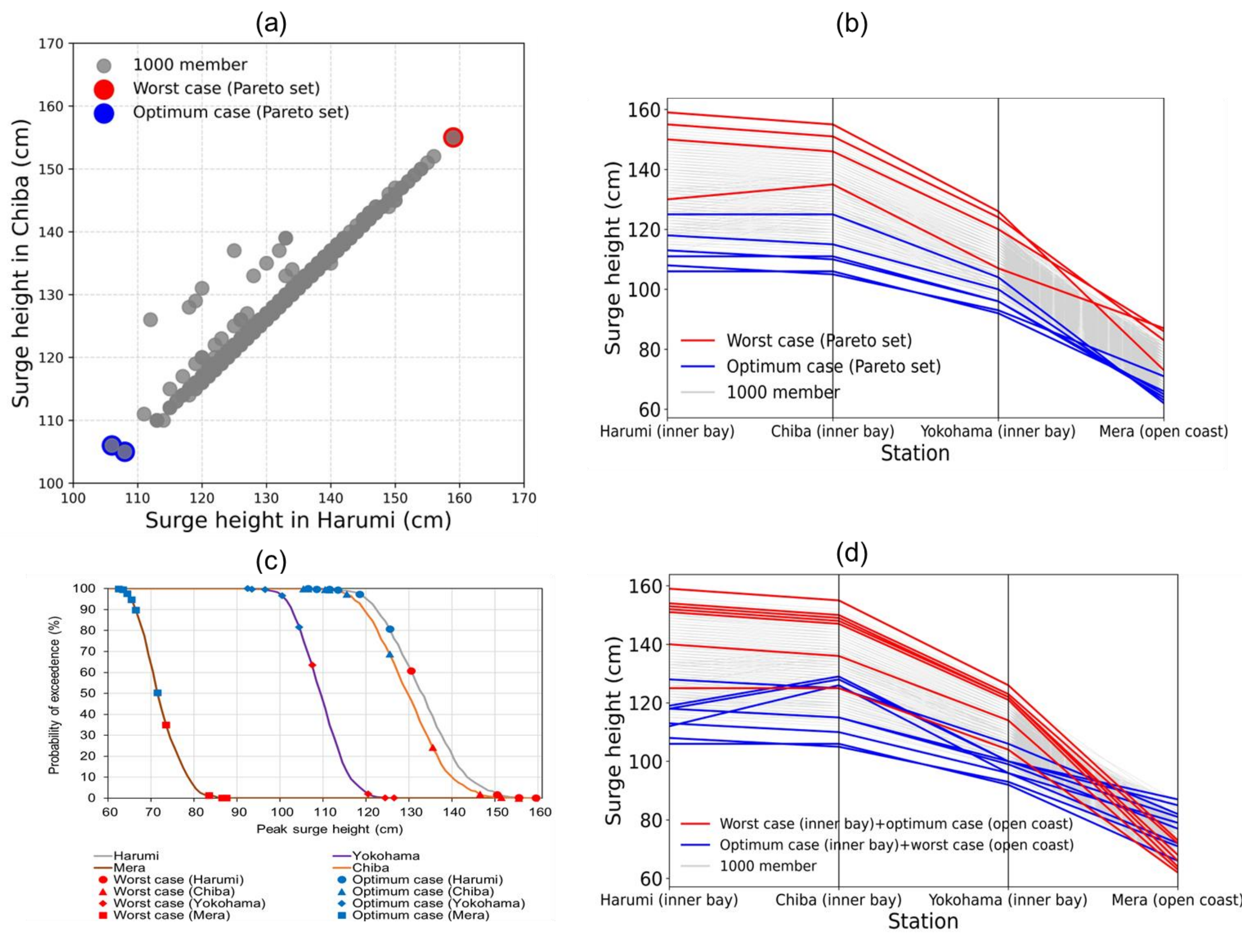

Figure 5. Forecasted Pareto optimal multi-scenarios due to TC Hagibis, apply for (a) Harumi and Chiba in Tokyo Bay [objective function (red dot): max surge height in Harumi and Chiba; objective function (blue dots): min surge height in Harumi and Chiba]; (b) Tokyo Bay (Harumi, Chiba, Yokohama and Mera) using the parallel coordinate plot [objective function (red lines): max surge height in Harumi, Chiba, Yokohama, and Mera; objective function (blue lines): min surge height in Harumi, Chiba, Yokohama, and Mera]; (c) Probability of exceeding (Y-axis) a storm surge threshold (X-axis) determined from each Pareto optimal solution in Figure 5b, at TC Hagibis landfall time (at 0900 UTC, 12 October 2019). It was estimated by fitting ensemble forecasts empirically; (d) Parallel coordinate plot including forecasted Pareto optimal solutions where forecasted peak surge intensity varies across Tokyo Bay [objective function (red lines): max surge height in Harumi, Chiba, and Yokohama + min surge height in Mera; objective function (blue lines): min surge height in Harumi, Chiba, and Yokohama + max surge height in Mera].

The results displayed in Figure 6a are comparable to those in Figure 5b but for Suruga Bay. It includes nine Pareto frontiers for worst-surge scenarios and one Pareto frontier for optimum surge scenario. Figure 6a predicts that all selected tide gauges would experience the worst surge levels of ~85 cm (<1% exceedance probability; Figure 6b), while the minimum surge height would be ~55 cm (>99% exceedance probability; Figure 6b). It is noteworthy that only one worst-surge scenario is found to be shared by both the Pareto frontiers of Tokyo Bay (Figure 5b) and Suruga Bay (Figure 6a), which represents a potential worst-case across the Japanese coastline during TC Hagibis. This indicates that among 1000 TC ensemble forecasts, a particular TC ensemble member has the potential to bring severe surge levels to both Bays. The reason behind this commonality will be discussed in Section 3.2.2.

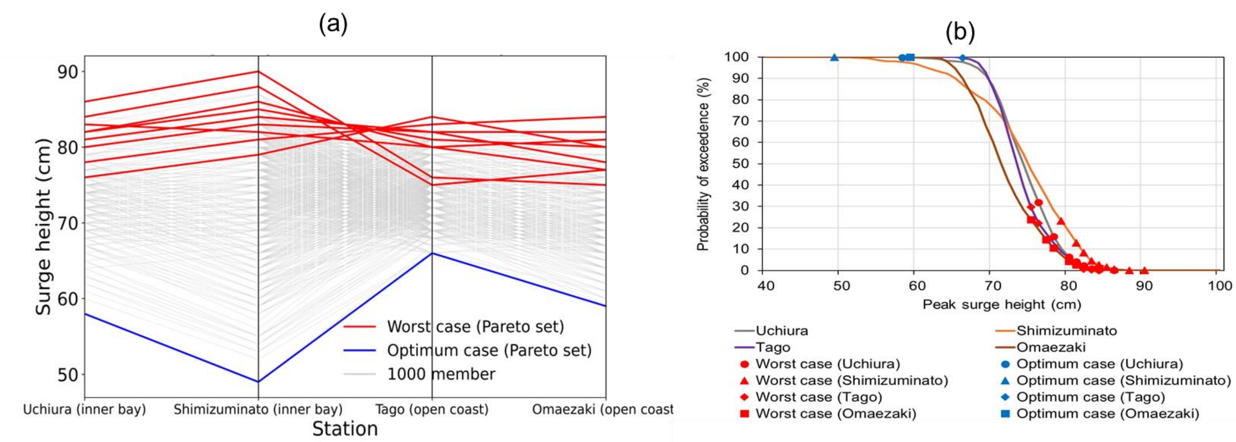

Figure 6. (a) Parallel coordinate plot with Pareto optimal multi-scenarios for Suruga Bay, at TC Hagibis landfall time (at 0900 UTC, 12 October 2019) [objective function (red lines): max surge height in Uchiura, Shimizuminato, Tago, and Omaezaki; objective function (blue lines): min surge height in Uchiura, Shimizuminato, Tago, and Omaezaki]; (b) Probability of exceeding (Y-axis) a storm surge threshold (X-axis) determined from each Pareto optimal solution in Figure 6a. It was estimated by fitting ensemble forecasts empirically.

In addition to Pareto Frontiers illustrated in Figures 5b and 6a, we further noticed many distinct trade-offs when both bays were considered together (Figure 7). For instance, we incorporated the representative tide gauges for each category of coastal geometry, such as inner bay (Harumi in Tokyo Bay and Shimizuminato in Suruga Bay) and open coast (Mera in Tokyo Bay and Tago in

Suruga Bay) in the objective function. This resulted in a large number of Pareto frontiers (worst case: nine; optimum case: twenty-two), with multiple overlapping solutions between worst and optimum cases. Thus, 33% of Pareto optimal solutions (red lines) predict that inner Tokyo Bay will experience the worst surge levels of ~150 cm, while inner Suruga Bay will witness ~60 cm, equivalent to minimum surge cases predicted by 41% of Pareto optimal solutions (blue lines; Figure 7). Although we maximize the potential of the large (i.e., 1000) ensemble forecasts, our proposed method is also found to be useful with the small ensemble size. We repeated the same analysis with 36 ensemble size. Although some worst and minimum storm surge scenarios were missed, a meaningful set of Pareto optimal solutions was still obtained (see Figure S2 in the supplementary section).

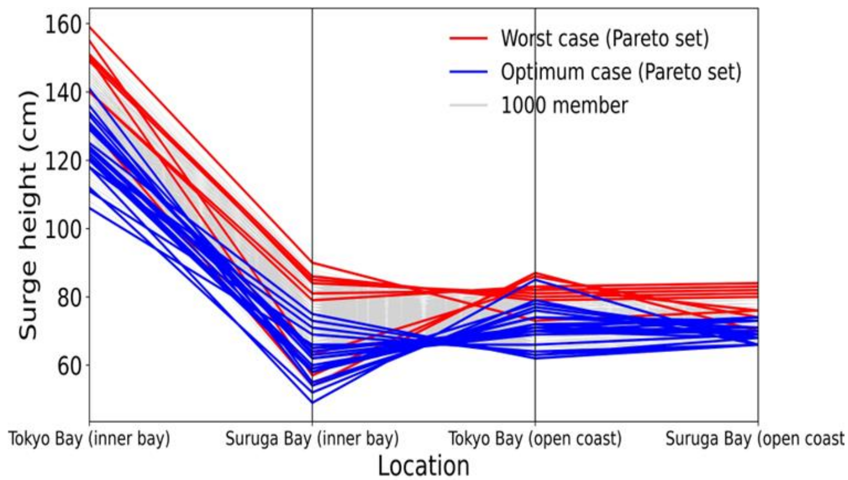

Figure 7. Parallel coordinate plot with Pareto optimal multi-scenarios determined for both Tokyo Bay and Suruga Bay [objective function (red lines): max surge height in Harumi, Shimizuminato, Mera, and Tago; objective function (blue lines): min surge height in Harumi, Shimizuminato, Mera, and Tago]

3.2.2 TC track and meteorological variables analysis of Pareto optimal solutions

It would be interesting to analyze the tracks and associated meteorological variables of the identified Pareto optimal solutions in Figures 5b and 6a. For example, Figure 8 reveals the strong sensitivity of the storm surge scenarios to the landfall location of TC Hagibis, leading to contrasting Pareto optimal solutions for Tokyo Bay (Figure 5b) and Suruga Bay (Figure 6a). It also demonstrates that the forecasted TC tracks for worst and optimum surge scenarios are significantly different from one other in both bays. For example, TC tracks (red lines; Figure 8a) that run parallel to the longitudinal axis of Tokyo Bay and pass over it would result in severe storm surges than TCs (blue lines; Figure 8a) that would travel 100 km or more to the west of the axis. Prior to landfall, under the worst case, easterly wind (Figure 9a) is forecasted to cause a buildup of water on the west coast (e.g., Yokohama) and initial draw-down in the north-eastern end of the Tokyo Bay (e.g., Chiba). Later, a surge level difference (0.6–0.8 m; Figure 5b) between the inner and lower ends of the bay is projected to occur (due to strong southerly winds) during the peak storm surge at the inner bays. During TC makes landfall under the optimum case (Figure 8a), the destructive right-side semicircle of the TC (Figure 9f) will interact with the vast land area rather than the ocean water, leading to less water being pushed towards Tokyo Bay (Figure 5b).

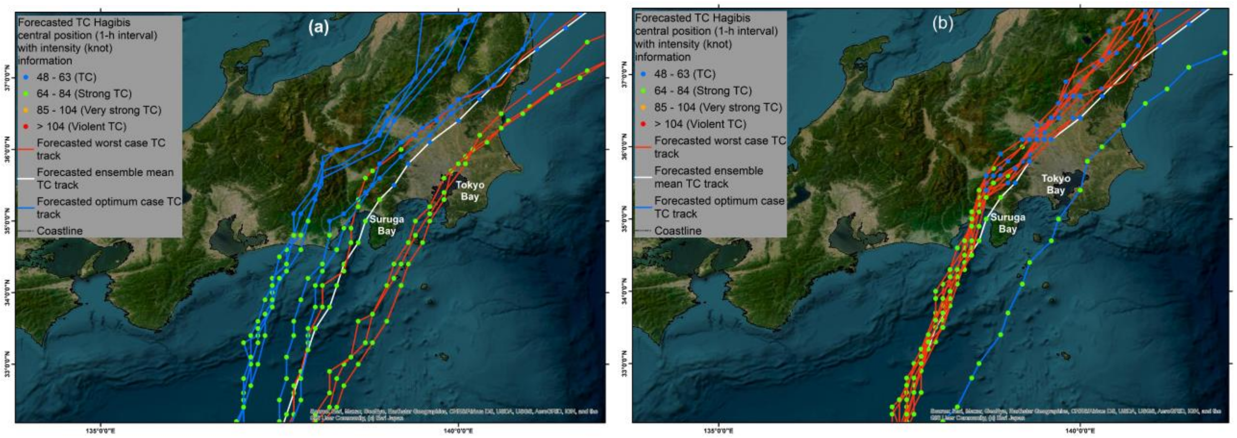

Figure 8. Forecasted TC Hagibis track in 39h lead time, respective to each Pareto frontier determined for (a) Tokyo Bay (as shown in Figure 5b); (b) Suruga Bay (as shown in Figure 6a). Red, blue, and white lines correspond to the forecasted worst, optimum, and ensemble mean TC track.

Notably, the optimized ensemble TC members for minimum surge scenarios in Tokyo Bay are forecasted to be stronger and larger than the members belonging to the worst surge cases until 24 hours prior to landfall, despite being centered in the same location. This is evident in Figure 9d, 9e, as opposed to Figure 9i, 9j. Despite both sets (worst and minimum) of optimized ensemble TC members weaken as they approach the mainland of Japan (Figure 9c, 9h and Figure 10), the worst TC members intensify by 7-kt (V_{max}) and remain large (R_{50}: ~120 NM) in the last 12 hours before landfall (Figure 9a, 9b and Figure 10). This large swath of strong winds is forecasted to affect a greater sea area and induce a motion in a greater quantity of water in Tokyo Bay.

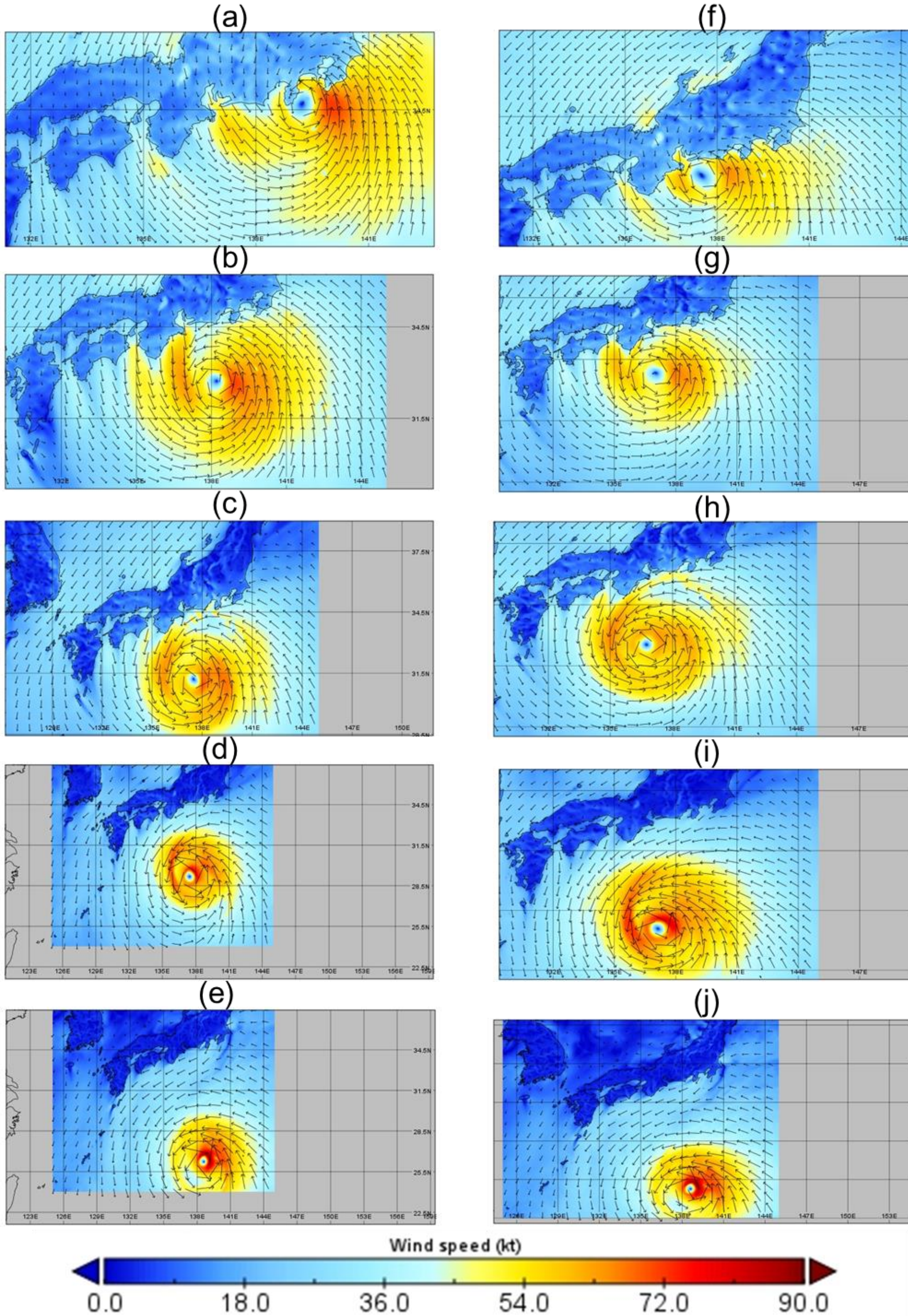

Figure 9. Forecasted composite 10 m wind field (kt-vectors), generated from optimized ensemble TC members for worst surge scenarios in Tokyo Bay (as shown in Figure 5b) during (a) landfall time (at 0900 UTC, 12 October 2019); (b) 6-h before landfall (at 0300 UTC, 12 October 2019); (c) 12-h before landfall (at 2100 UTC, 11 October 2019); (d) 24-h before landfall (at 0900 UTC, 11 October 2019); (e) 39-h before landfall (at 1800 UTC, 10 October 2019); (f-j) same as (a-e) but

generated from optimized ensemble TC members for minimum surge scenarios in Tokyo Bay (as shown in 5b).

Figure 8a highlights that in addition to TC tracks that pass directly over Tokyo Bay, one particular worst TC forecast (red line) makes landfall around 70 km west of the longitudinal axis of Tokyo Bay. This specific track is forecasted to cause severe storm surges in both Tokyo Bay (Figure 5b) and Suruga Bay (Figure 6a). This particular ensemble forecast has a wider range of intense winds (R_{50}: ~140 NM) across a larger area and a slower movement speed (S: ~32 km/h), despite having a similar landfall wind intensity (V_{max}: ~75-kt) compared to other worst-case TC forecasts. This unique phenomenon corroborates earlier numerical analyses that propose the likelihood of a severe storm surge scenario in the upper-bay region when a large and intense TC moves slowly, parallel to the longitudinal axis of Tokyo Bay, after making landfall 25 km southwest. (Islam & Takagi, 2020b).

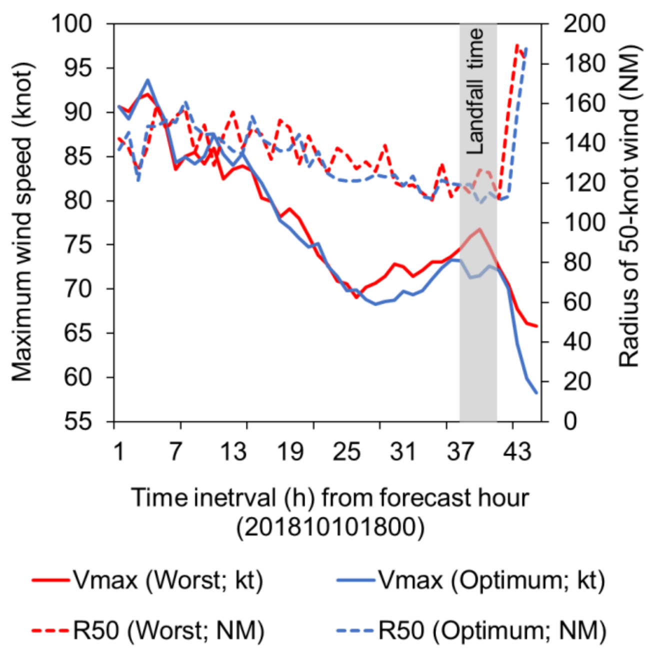

Figure 10. Evolution (composite) of maximum wind speed (V_{max}) and radius of 50-kt wind (R_{50}) for worst and optimum TC forecasts in Tokyo Bay (as shown in Figure 5b) in 39-h lead time.

4 Conclusions and discussion

The application of ensemble TC forecasting in hazard prediction, such as storm surge, has been greatly overlooked despite its use in forecasting TC track, intensity, and genesis. Enhanced analysis can unlock and maximize the benefit of ensemble forecasting. Here, we proposed Pareto optimality – a novel and practical way to identify potential ensemble TC (Hagibis) forecast from an extremely large ensemble (=1000 member) that can effectively assess storm surge multi-scenarios, including possible worst and optimum cases for a coastal location. The variability in storm surge intensity across the coastline makes it challenging for decision-makers to plan effective evacuation measures. To address this, we have demonstrated that a diversity of trade-off surge outcomes among coastal places can be identified by choosing the Pareto optimized forecasts. The in-depth evaluation of Pareto optimal solutions can shed light on how meteorological variables such as track, intensity, and size of TCs influence the worst and optimum surge scenarios, which are not well understood by emergency managers using current multi-scenario assessment methods (such as those used by JMA and NHC).

The significance of evaluating trade-offs based on Pareto optimization has long been acknowledged in the context of sustainable development goals and management of ecosystem services (Flecker et al., 2022). However, its application in disaster-risk communities is noticeably lacking. During a coastal storm surge event, effective evacuation planning and warning issuance involve multi-criteria problems such as storm surge intensity, coastal population vulnerability, and available evacuation resources. Traditionally, this decision-making process has relied on a hazard map, which typically depicts the severity of the predicted storm surge (e.g., exceeding a critical surge height; J. Hasegawa et al., 2017). However, this approach does not fully capture the diversity of potential storm surge scenarios across the coastlines, which can lead to ineffective evacuation planning. A recent TC Fani in 2019 serves as evidence to support this statement. TC Fani struck the southeastern part of India, approximately 450 km from the southwest coast of Bangladesh, as a Category 4 TC (in Saffir-Simpson Hurricane Wind Scale). Prior to TC Fani reaching the Bangladesh coast as a tropical storm, the Bangladesh Meteorological Department (BMD) issued 'danger' signal number seven (out of ten), which led to the evacuation of one million people (Bangladesh Meteorological Department, 2021). Later, the catastrophe, such as storm surge level (~1 m), did not hit the danger level as anticipated. BMD's evacuation order for the entire southwest coast was not based on a specific surge scenario (e.g., worst case) and forecasted meteorological conditions associated with it, leading to an excessive number of evacuees. Such a false alarm demotivated people to seek shelter when a Category 2 TC Amphan caused ~2.75 m storm surges and claimed at least 26 lives in 2020, despite the issuance of 'great danger' signal number ten (Raju, 2019; ReliefWeb, 2021; Alam et al., 2023). It seems that BMD took a safer and conservative decision during TC Fani by issuing signal no. 7, nevertheless, this cannot be considered effective decision-making. While such a complex decision-making process can certainly be improved by quantifying the uncertainty through an ensemble multi-scenario forecast, incorporating Pareto optimality can further maximize the benefits of it.

Pareto optimal solution provides an effective first filter to identify ensemble multi-scenario surge forecasts. This information can be presented visually to enhance the understanding of the uncertainty in the forecast. The median of Pareto optimal solutions could be utilized given a series of worst/minimum surge estimations for a specific location by several ensemble members. For example, Figure 5b identifies four worst scenarios for Harumi in Tokyo Bay where the median surge level is 150 cm. Although we stress the importance of diversity in trade-offs surge outcomes, a certain scenario (e.g., median of Pareto optimal solutions) can be given more weight depending on the values of society and decision-makers. In the decision-making process, a user-defined acceptable level of uncertainty or reference surge height (e.g., 25-year return period of surge) can be set for a specific location (e.g., Harumi). The forecaster can then determine if the height of the Pareto optimal solution exceeds this acceptable level. Subsequently, a relevant warning signal can be issued in a forecast horizon (e.g., 39-h lead time). The warning signal can be tailored to a specific location, if a diversity in the trade-off between surge outcomes exists among Pareto frontiers. For example, Pareto optimal solutions in Figure 5b predicted that TC Hagibis will bring worst surge level as maximum as 160 cm in the inner Tokyo Bay, requiring the issuance of an emergency warning, closing flood gates, and large-scale evacuation of the coastal population living below the storm surge height. On the other hand, those living along open coastlines are advised to stay indoors as the predicted worst surge level (90 cm) does not meet the criteria for issuing an emergency warning. Once this forecast becomes available, decision-makers (e.g., emergency managers) can start evacuation planning based on the forecasted worst TC track, wind intensity, and peak surge height. For instance, TC track that forecasted to bring severe storm surges of 160 cm in the inner Tokyo Bay, would make landfall 70 km west of the central bay axis with a landfall V_{max} of 75 kt, which is stronger by 11-kt from the historical mean (64-kt; (Islam et al., 2022)). Furthermore, it is projected to be twice as large as the historical average (65 NM) and move at a slower speed by 9 km/h compared to the average translation speed (41 km/h) in Tokyo Bay. The forecasted landfall location and meteorological conditions of the worst TC indicate that Tokyo Bay would be situated in the destructive right-side semicircle of the TC track, resulting in prolonged exposure to severe storm surges and strong winds. Emergency managers can utilize this information to disseminate surge warnings to residents and commence evacuation procedures with a 39-hour lead time. This evacuation can be done by dividing coastal regions into different zones depending on their vulnerability. Although disaster planning is not so straightforward as explained here, our proposed ensemble-based storm surge multi-scenario analysis is expected to motivate forecasters and risk management practitioners to explore new ways to assess storm surge hazards and reduce the associated risk.

Finally, we acknowledge that this study focuses exclusively on peak surge height while determining total sea water level that includes the influence of astronomic tide, wave set-up, and river discharge are also critical and can be done utilizing a full physical numerical model. Furthermore, several algorithms are currently available to determine a Pareto frontier. We encourage researchers from multiple disciplines to build on our approach to help us reach an improved understanding of Pareto optimality based multi-scenario analysis.

Acknowledgment

The authors declare no financial or conflict of interest. This work is supported by JST FOREST program (grant no. JPMJFR205Q) and JST Moonshot R&D project (grant no. JPMJMS2281). We acknowledge the support by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) through the Program for Promoting Researches on the Supercomputer Fugaku JPMXP1020351142 'Large Ensemble Atmospheric and Environmental Prediction for Disaster Prevention and Mitigation' (hp210166).

Open Research

Observed storm surge data can be downloaded from the JMA (https://www.data.jma.go.jp/kaiyou/db/tide/genbo/index.php) and JODC (https://jdoss1.jodc.go.jp/vpage/tide.html) websites. Predicted tide data can be obtained from the JMA (https://www.data.jma.go.jp/kaiyou/db/tide/suisan/index.php) website. TC best track data can be derived from the JMA (https://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/trackarchives.html) website. Ensemble forecast data may be available upon request.

References

Alam, Md. S., Chakraborty, T., Hossain, Md. Z., & Rahaman, K. R. (2023). Evacuation dilemmas of coastal households during cyclone Amphan and amidst the COVID-19 pandemic: a study of the Southwestern region of Bangladesh. *Natural Hazards*, *115*(1), 507–537. https://doi.org/10.1007/s11069-022-05564-9

Bangladesh Meteorological Department. (2021). Bangladesh Meteorological Department. Retrieved March 12, 2021, from http://live.bmd.gov.bd/

Bernier, N. B., & Thompson, K. R. (2015). Deterministic and ensemble storm surge prediction for Atlantic Canada with lead times of hours to ten days. *Ocean Modelling*, *86*, 114–127. https://doi.org/10.1016/j.ocemod.2014.12.002

Duc, L., Saito, K., & Hotta, D. (2020). Analysis and design of covariance inflation methods using inflation functions. Part 1: Theoretical framework. *Quarterly Journal of the Royal Meteorological Society*, *146*(733), 3638–3660. https://doi.org/10.1002/qj.3864

Duc, L., Kawabata, T., Saito, K., & Oizumi, T. (2021). Forecasts of the July 2020 Kyushu Heavy Rain Using a 1000-Member Ensemble Kalman Filter. *Sola*, *17*, 41–47. https://doi.org/10.2151/sola.2021-007

Flecker, A. S., Shi, Q., Almeida, R. M., Angarita, H., Gomes-Selman, J. M., García-Villacorta, R., et al. (2022). Reducing adverse impacts of Amazon hydropower expansion. *Science*, *375*(6582), 753–760. https://doi.org/10.1126/science.abj4017

Flowerdew, J., Mylne, K., Jones, C., & Titley, H. (2013). Extending the forecast range of the UK storm surge ensemble. *Quarterly Journal of the Royal Meteorological Society*, *139*(670), 184–197. https://doi.org/10.1002/qj.1950

Greenslade, D., Freeman, J., Sims, H., Allen, S., Colberg, F., Schulz, E., et al. (2017). An operational coastal sea level forecasting system. In *Australasian Coasts & Ports 2017: Working with Nature* (pp. 514–520). https://doi.org/10.3316/informit.925870785579893

Hasegawa, H., Kohno, N., Higaki, M., & Itoh, M. (2017). *Upgrade of JMA's Storm Surge Prediction for the WMO Storm Surge Watch Scheme (SSWS)*. RSMC Tokyo - Typhoon Center. Retrieved from https://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/techrev/text19-2.pdf

Hasegawa, J., Takagi, Y., & Tsuboi, Y. (2017). *Key techniques for Japan's Risk-based Warning Services for Heavy-rain Related Disasters*. Retrieved from http://www.typhooncommittee.org/50th/TECO/Parallel1A_2.pdf

Islam, M. R., & Takagi, H. (2020a). Statistical significance of tropical cyclone forward speed on storm surge generation: retrospective analysis of best track and tidal data in Japan. *Georisk*. https://doi.org/10.1080/17499518.2020.1756345

Islam, M. R., & Takagi, H. (2020b). Typhoon parameter sensitivity of storm surge in the semi-enclosed Tokyo Bay. *Frontiers of Earth Science*, *14*(3), 553–567. https://doi.org/10.1007/s11707-020-0817-1

Islam, M. R., Takagi, H., Anh, L. T., Takahashi, A., & Bowei, K. (2018). 2017 Typhoon Lan reconnaissance field survey in coasts of Kanto region, Japan. *Journal of Japan Society of Civil Engineers, Ser. B3 (Ocean Engineering)*, *74*(2), I_593-I_598. https://doi.org/10.2208/jscejoe.74.i_593

Islam, M. R., Lee, C.-Y., Mandli, K. T., & Takagi, H. (2021). A new tropical cyclone surge index incorporating the effects of coastal geometry, bathymetry and storm information. *Scientific Reports*, *11*(1), 16747. https://doi.org/10.1038/s41598-021-95825-7

Islam, M. R., Satoh, M., & Takagi, H. (2022). Tropical Cyclones Affecting Japan Central Coast and Changing Storm Surge Hazard since 1980. *Journal of the Meteorological Society of Japan. Ser. II*, *100*(3), 493–507. https://doi.org/10.2151/jmsj.2022-024

Japan Oceanographic Data Center. (2020). 500m Gridded Bathymetry Data. Retrieved January 8, 2021, from https://www.jodc.go.jp/jodcweb/JDOSS/infoJEGG.html

Japan Oceanography Data Center. (2021). Tide Stations. Retrieved January 8, 2021, from https://jdoss1.jodc.go.jp/vpage/tide.html

JMA. (2021). Japan Meteorological Agency | RSMC Tokyo - Typhoon Center | Best Track Data. Retrieved March 13, 2021, from https://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/trackarchives.html

JMA. (2022). Tidal observation data. Retrieved June 30, 2022, from https://www.data.jma.go.jp/gmd/kaiyou/db/tide/genbo/index.php

JMA. (2023). Japan Meteorological Agency｜Numerical Weather Prediction Activities. Retrieved March 10, 2023, from https://www.jma.go.jp/jma/en/Activities/nwp.html

Kobayashi, K., Duc, L., Apip, Oizumi, T., & Saito, K. (2020). Ensemble flood simulation for a small dam catchment in Japan using nonhydrostatic model rainfalls – Part 2: Flood forecasting using 1600-member 4D-EnVar-predicted rainfalls. *Natural Hazards and Earth System Sciences*, *20*(3), 755–770. https://doi.org/10.5194/nhess-20-755-2020

Kochenderfer, M. J., & Wheeler, T. A. (2019). Multiobjective Optimization. In *Algorithms for Optimization* (pp. 211–233). The MIT Press.

Kohno, N., Dube, S. K., Entel, M., Fakhruddin, S. H. M., Greenslade, D., Leroux, M.-D., et al. (2018). Recent Progress in Storm Surge Forecasting. *Tropical Cyclone Research and Review*, *7*(2), 128–139. https://doi.org/10.6057/2018TCRR02.04

Kristensen, N. M., Røed, L. P., & Sætra, Ø. (2022). A forecasting and warning system of storm surge events along the Norwegian coast. *Environmental Fluid Mechanics*. https://doi.org/10.1007/s10652-022-09871-4

Liu, Y., Liu, Z., Lamont, S., Hirpa, F., Zhang, Y., Li, T., et al. (2021). *Simulation of Storm Surge during Typhoon Hagibis, 2019 with a Large-scale Japan-wide Coastal Hydrodynamic Model*. Presented at the AGU Fall Meeting 2021, New Orleans, LA, USA. Retrieved from https://ui.adsabs.harvard.edu/abs/2021AGUFMOS33B..05L

Lombardi, G., Ceppi, A., Ravazzani, G., Davolio, S., & Mancini, M. (2018). From Deterministic to Probabilistic Forecasts: The 'Shift-Target' Approach in the Milan Urban Area (Northern Italy). *Geosciences*, *8*(5), 181. https://doi.org/10.3390/geosciences8050181

Ma, W., Ishitsuka, Y., Takeshima, A., Hibino, K., Yamazaki, D., Yamamoto, K., et al. (2021). Applicability of a nationwide flood forecasting system for Typhoon Hagibis 2019. *Scientific Reports*, *11*(1), 10213. https://doi.org/10.1038/s41598-021-89522-8

Mel, R., & Lionello, P. (2014). Storm Surge Ensemble Prediction for the City of Venice. *Weather and Forecasting*, *29*(4), 1044–1057. https://doi.org/10.1175/WAF-D-13-00117.1

Needham, H. F., Keim, B. D., & Sathiaraj, D. (2015). A review of tropical cyclone-generated storm surges: Global data sources, observations, and impacts. *Reviews of Geophysics*, *53*(2), 545–591. https://doi.org/10.1002/2014RG000477

NHC. (n.d.). Sea, Lake, and Overland Surges from Hurricanes (SLOSH). Retrieved January 19, 2023, from https://www.nhc.noaa.gov/surge/slosh.php

Raju, M. N. A. (2019, November 12). Cyclone Bulbul: Views from the ground. Retrieved February 8, 2023, from https://www.thedailystar.net/opinion/environment/news/cyclone-bulbul-views-the-ground-1825732

Rappaport, E. N. (2014). Fatalities in the United States from Atlantic Tropical Cyclones: New Data and Interpretation. *Bulletin of the American Meteorological Society*, *95*(3), 341–346. https://doi.org/10.1175/BAMS-D-12-00074.1

ReliefWeb. (2021). Counting the Costs: Cyclone Amphan One Year On - Bangladesh | ReliefWeb. Retrieved February 9, 2023, from https://reliefweb.int/report/bangladesh/counting-costs-cyclone-amphan-one-year

Saito, K., Fujita, T., Yamada, Y., Ishida, J., Kumagai, Y., Aranami, K., et al. (2006). The Operational JMA Nonhydrostatic Mesoscale Model. *Monthly Weather Review*, *134*(4), 1266–1298. https://doi.org/10.1175/MWR3120.1

Sawada, Y., Kanai, R., & Kotani, H. (2022). Impact of cry wolf effects on social preparedness and the efficiency of flood early warning systems. *Hydrology and Earth System Sciences*, *26*(16), 4265–4278. https://doi.org/10.5194/hess-26-4265-2022

Sharma, M., Lamers, A., Devi, S., Schumacher, A., Earl-Spurr, C., Fritz, C., et al. (2022). *Forecasting Tropical Cyclone Hazards and Impacts*. 10th International Workshop on Tropical Cyclones (IWTC-10). Retrieved from https://community.wmo.int/en/iwtc-10-reports

Shimozono, T., Tajima, Y., Kumagai, K., Arikawa, T., Oda, Y., Shigihara, Y., et al. (2020). Coastal impacts of super typhoon Hagibis on Greater Tokyo and Shizuoka areas, Japan. *Coastal Engineering Journal*, *62*(2), 129–145. https://doi.org/10.1080/21664250.2020.1744212

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T. M., Hewson, T. D., et al. (2016). The TIGGE Project and Its Achievements. *Bulletin of the American Meteorological Society*, *97*(1), 49–67. https://doi.org/10.1175/BAMS-D-13-00191.1

Takagi, H., Xiong, Y., & Furukawa, F. (2018). Track analysis and storm surge investigation of 2017 Typhoon Hato: were the warning signals issued in Macau and Hong Kong timed appropriately? *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, *12*(4), 297–307. https://doi.org/10.1080/17499518.2018.1465573

Takagi, H., Anh, L. T., Islam, R., & Hossain, T. T. (2022). Progress of disaster mitigation against tropical cyclones and storm surges: a comparative study of Bangladesh, Vietnam, and Japan. *Coastal Engineering Journal*, *0*(0), 1–15. https://doi.org/10.1080/21664250.2022.2100179

Titley, H. A., Yamaguchi, M., & Magnusson, L. (2019). Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey. *Tropical Cyclone Research and Review*, *8*(3), 166–180. https://doi.org/10.1016/j.tcrr.2019.10.005

Tommy. (2021). paretoset (Version 1.2.0). Retrieved from https://github.com/tommyod/paretoset

Wilson, K. A., Heinselman, P. L., Skinner, P. S., Choate, J. J., & Klockow-McClain, K. E. (2019). Meteorologists' Interpretations of Storm-Scale Ensemble-Based Forecast Guidance. *Weather, Climate, and Society*, *11*(2), 337–354. https://doi.org/10.1175/WCAS-D-18-0084.1

Yamaguchi, M., Vitart, F., Lang, S. T. K., Magnusson, L., Elsberry, R. L., Elliott, G., et al. (2015). Global Distribution of the Skill of Tropical Cyclone Activity Forecasts on Short- to Medium-Range Time Scales. *Weather and Forecasting*, *30*(6), 1695–1709. https://doi.org/10.1175/WAF-D-14-00136.1

Journal of Geophysical Research: Atmospheres

Supporting Information for

Assessing Storm Surge Multi-Scenarios based on Ensemble Tropical Cyclone Forecasting

Md. Rezuanul Islam[1*], Le Duc[1,2], and Yohei Sawada[1,2]

[1]Institute of Engineering Innovation, The University of Tokyo, Tokyo, 113-0032, Japan

[2]Meteorological Research Institute, Japan Meteorological Agency

Corresponding author: Md. Rezuanul Islam (fahiemislam@gmail.com)

Contents of this file

Figures S1 to S2

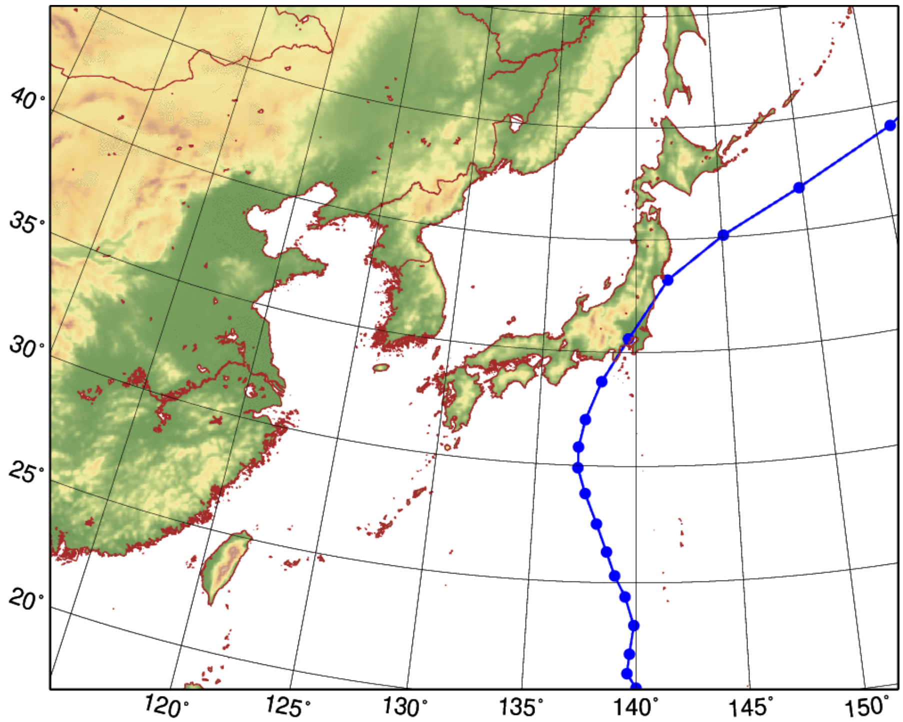

Figure S1. The analysis and forecast domain of the data assimilation system 4DEnVAR and the forecast model NHM. The best track of the TC Hagibis is also plotted.

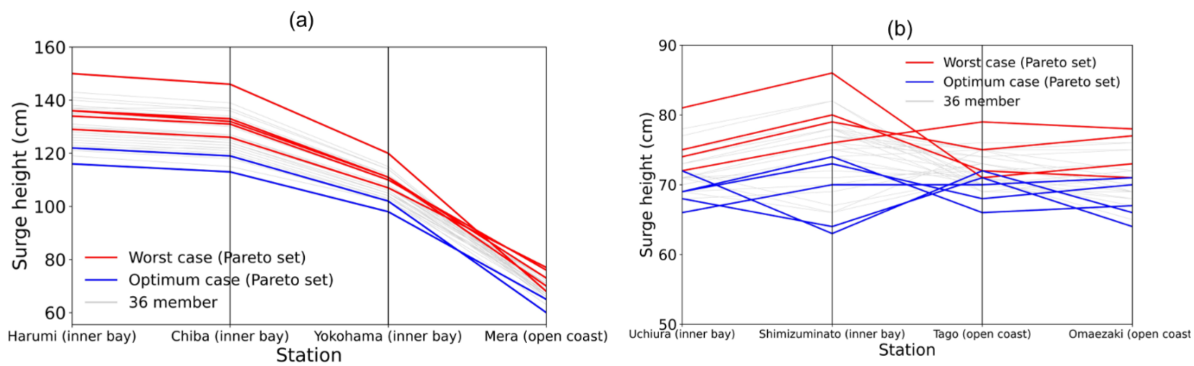

Figure S2. Parallel coordinate plot with Pareto optimal multi-scenarios based on 36 ensemble forecasts for (a) Tokyo Bay (Harumi, Chiba, Yokohama and Mera) [objective function (red lines): max surge height in Harumi, Chiba, Yokohama, and Mera; objective function (blue lines): min surge height in Harumi, Chiba, Yokohama, and Mera]; (b) Suruga Bay (Uchiura, Shimizuminato, Tago, and Omaezaki) [objective function (red lines): max surge height in Uchiura, Shimizuminato, Tago, and Omaezaki; objective function (blue lines): min surge height in Uchiura, Shimizuminato, Tago, and Omaezaki].