CrossMark

Invited review

Is flow velocity important in tsunami empirical fragility modeling?

Raffaele De Risi [a],*, Katsuichiro Goda [a], Tomohiro Yasuda [b], Nobuhito Mori [b]

[a] Department of Civil Engineering, Queen's Building, University Walk, University of Bristol, BS8 1TR, UK
[b] Disaster Prevention Research Institute, Kyoto University, Kyoto 611-0011, Japan

ARTICLE INFO

ABSTRACT

The influence of flow velocity on structural damage induced by tsunami inundation is investigated to improve empirical fragility models considering flow velocity as explanatory hazard variable in addition to inundation depth. The analysis is based on extensive tsunami damage data for the 2011 Tohoku earthquake collected by the Ministry of Land, Infrastructure, and Transportation of the Japanese Government. Multivariate tsunami fragility curves are developed through multinomial logistic regression of un-binned data. This approach facilitates the flexible development of various nested models considering inundation depth alone or inundation depth and velocity altogether. Statistical diagnostic metrics, such as the Bayesian Information Criterion, the Akaike Information Criterion, and the residual deviance, are used to determine which model improves the predictability of tsunami damage. The significance and importance of including flow velocity in the vulnerability models are assessed quantitatively by examining the influence of different spatial resolutions in elevation model and different source models. Then, the effects of coastal topography have been investigated. Numerical results show that flow velocity generally improves the fragility models, particularly for severer structural damage states, and that it is important for sites along the coast where the inundation depth is not extremely high. Coarse digital elevation model and inaccurate source models have influence on the calculated values of flow velocity and thus they affect the accuracy of fragility modeling. Finally, two different fragility models are calibrated for plain-type and ria-type coasts by reflecting differences in hydrodynamic behavior and recorded damage on the structures.

Contents

* Corresponding author.
  E-mail addresses: raffaele.derisi@bristol.ac.uk (R. De Risi), katsu.goda@bristol.ac.uk (K. Goda), tomo@oceanwave.jp (T. Yasuda), mori@oceanwave.jp (N. Mori).

1. Introduction

Tsunamis pose a significant threat to coastal communities near active subduction zones. The events occurred in the last two decades highlight the importance of developing vulnerability models (e.g. tsunami fragility curves) to predict the tsunami impact on the built environment. A fragility curve evaluates the probability of reaching or exceeding specific damage states for a given hazard intensity (Porter et al., 2007). Such relationships between hazard and potential damage play a vital role in quantifying damage and losses associated with natural hazards. Several empirical tsunami fragility curves can be found in literature (e.g., Tarbotton et al., 2015). The majority of the existing studies adopt inundation depth as intensity measure, especially more recent ones (Reese et al., 2011; Suppasri et al., 2013; Charvet et al., 2014; De Risi et al., 2016). Only a few tsunami fragility models consider flow velocity (Koshimura et al., 2009; Koshimura and Kayaba, 2011; Suppasri et al., 2011). Nevertheless, it is generally accepted that the higher the flow velocity of the inundation, the greater the probability of structural damage (Kreibich et al., 2009). This is particularly true for the case of debris impact that is mainly governed by flow velocity (Ghobarah et al., 2006). In this context, Charvet et al. (2015) proposed a multivariate fragility formulation in order to take into account not only inundation depth, but also inundation velocity and debris impact.

There are many techniques to obtain tsunami flow velocity. They can be divided into direct and indirect methods, depending on whether the velocities are measured directly or are obtained from post-processing of other data. The direct methods include (i) the particle image velocimetry analysis based on videos and topographic data (Fritz et al., 2012; Hayashi and Koshimura, 2013), (ii) the coastal oceanographic radar tsunami system based on high-frequency radar technology (Lipa et al., 2012), and (iii) satellite altimetry based on recognition from the space (Song et al., 2012). Indirect methods are based on hydraulic or geological analyses. In the first case, flow velocity is estimated from Bernoulli principle applied to inundation depth before and after an obstacle or by combining measures of inundation depth and Froude number. In the second case, flow velocity is calculated through the analysis of soil sediments inland (Jaffe et al., 2011) and/or through the analysis of boulder transport (Etienne et al., 2011).

For empirical tsunami fragility modeling, the preceding methods are not applicable because the velocity information for numerous buildings is needed. In practice, the velocity information can be obtained through numerical tsunami simulation based on a given initial boundary condition, and many numerical tsunami codes are available in the literature (e.g., Dutykh et al., 2011). Generally, codes use initial tsunami conditions (based on an earthquake rupture model), bathymetry and elevation data, and roughness maps as input, and return time-histories of inundation depth and flow velocity at several locations of interest. The key issue is to ensure that the tsunami simulation, as a whole, is a reasonable representation of the actual phenomena, and for this reason, the tsunami model needs to be calibrated using observed off-shore wave as well as on-shore inundation data.

This study aims at providing an answer to a question: is flow velocity important for tsunami fragility modeling? A rigorous statistical analysis is carried out to assess the influence of accounting for flow velocity in tsunami vulnerability modeling systematically. The investigation is based on the tsunami damage data recorded from extensive post-event field surveys and tsunami damage inspections after the 2011 Tohoku, Japan tsunami. All the data are collected in the database by the Ministry of Land, Infrastructure, and Transportation (MLIT, 2014),

containing more than 200,000 buildings; each data entry includes building type, location, tsunami damage level, and inundation depth.

To obtain the tsunami velocity data at building sites, tsunami simulation is carried out using a calibrated tsunami source model by Satake et al. (2013) and fine-resolution (10 m) elevation data for Miyagi prefecture. The accuracy of the tsunami simulation model is evaluated by comparing observed and simulated tsunami inundation data. Moreover, to investigate the effect of different resolutions of Digital Elevation Model (DEM) and different source models on flow velocity, three simulation are considered additionally: (i) using the source model by Satake et al. (2013) and 50-m resolution elevation data, (ii) using the source model by Ammon et al. (2011) and 10-m resolution data, and (iii) using the source model by Iinuma et al. (2012) and 10-m resolution data. The source models by Satake et al. (2013), Ammon et al. (2011), and Iinuma et al. (2012) are calibrated based on tsunami data, geodetic and teleseismic data, and geodetic data, respectively; therefore modeling accuracy of flow velocities varies for the three source models considered.

Fragility curves are obtained adopting a multinomial logistic regression (e.g. Charvet et al., 2014). In total, five nested models are examined as linear predictor functions: the simplest one considers the inundation depth and the structural class, whereas the most complex one considers both inundation depth and flow velocity together with interaction terms between intensity measures and structural typologies. Diagnostic analyses for fragility model selection are performed considering three indicators: (i) the Bayesian Information Criterion, (ii) the Akaike Information Criterion, and (iii) the residual deviance. The diagnostic analysis for fragility model selection is carried out considering the entire MLIT database (i.e. all available data without distinction of coastal topography), using simulated flow velocity data based on the Satake et al. model and 10-m resolution DEM (i.e. the reference inundation model). Then, the diagnostic analysis is repeated considering the velocity values obtained by the three additional inundation cases to examine the effects of accuracy of the simulated velocity data on the fragility models. This sensitivity study of fragility curves to inundation model is a novel result for tsunami vulnerability.

Finally, since the topographic effects are known to be important, the influence of coastal topography on tsunami fragility is investigated via rigorous residual analysis by distinguishing the results for plain-type and ria-type coasts and by identifying systematic trends of the residuals. The residual analysis shows that it is important to consider the two coastal typologies separately when the flow velocity is incorporated in the vulnerability models. Insights gained through such residual analyses are valuable for improving the physical understanding of the correlation between damage and effective inundation intensity measure, and they are also particularly useful for developing future generations of tsunami vulnerability models.

The paper is organized as follows. Section 2 presents the available database for tsunami fragility analysis and the simulation procedures to obtain the flow velocity data. Section 3 presents the mathematical formulations for the multinomial regression analysis, and Section 4 shows the numerical results. Finally, key conclusions are drawn in Section 5.

2. MLIT database and velocity data

The tsunami data used in this study are obtained from the MLIT damage database (MLIT, 2014). Each building located in the affected areas is classified according to different attributes, such as geographical

location, structural material, tsunami inundation depth, and sustained damage level. The total number of the buildings that are contained in the database exceeds 200,000, and about 84% of the data are accompanied by the supplementary information that are necessary for tsunami fragility modeling. In this study, a subset of the MLIT data, which is located in Miyagi prefecture (Fig. 1(a)), is employed for the fragility analysis. The reason for this focus is that a high resolution (10-m) DEM is available for Miyagi prefecture and using the fine resolution data is critical in ensuring that the flow velocity data at building sites, obtained from tsunami simulations, are accurate. It is noted that the greatest part of the MLIT data fall in Miyagi prefecture (77.4%), thus the results from tsunami fragility analysis are representative of the MLIT database and the sustained tsunami damage during the 2011 Tohoku earthquake.

2.1. Recorded data in Miyagi prefecture

The MLIT database defines seven discrete levels of tsunami damage: no damage (DS1), minor damage (DS2), moderate damage (DS3), major damage (DS4), complete damage (DS5), collapse (DS6), and washed away (DS7). In engineering terms, DS1, DS2, and DS3 identify non-structural damage (i.e. minor flooding or slight damage to non-structural components), whereas DS4 to DS7 refer to structural damage, such as damage to frames/walls and overturning/translation of buildings. Fig. 1(a) shows the geographical distribution of damage states along the Miyagi coastline. The surveyed buildings in Miyagi prefecture experienced inundation depths ranging between 0.1 m and 27.3 m, as depicted in Fig. 1(b). Only 2.0% of the surveyed buildings suffered no damage, whereas many buildings (i.e. 30.3%) were washed away. The descriptive statistics of the damage levels can be seen in Fig. 1(c). For all damage states, wooden buildings were more affected, followed by masonry, steel, and RC structures. This reflects the distribution of structural typologies in Miyagi prefecture, i.e. 84.1% wooden buildings, 9.4% masonry buildings, 4.3% steel buildings, and 2.2% RC buildings.
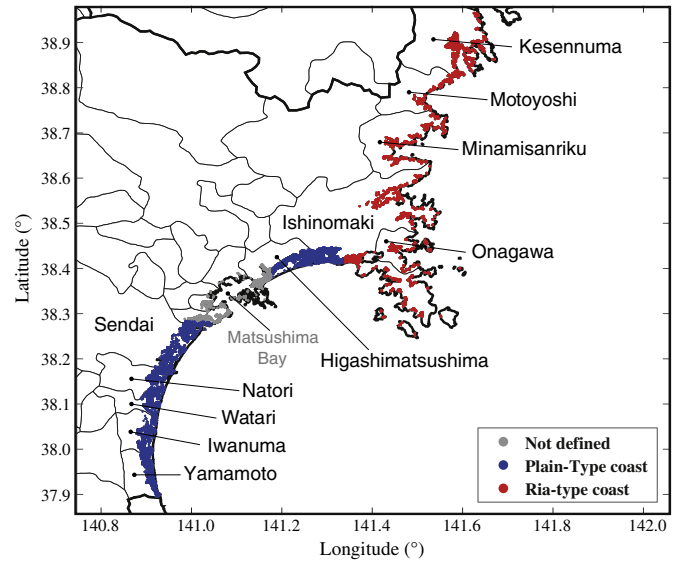

Fig. 2. Municipalities considered in this study for the analysis of the influence of topographical effect on the vulnerability model.

From a topographical point of view, it is interesting to note that, starting from South to North, about 57% of buildings are located along the plain-type coast, about 31% are along ria-type coast, and the remaining 12% are located around Matsushima Bay. Based on this, the Miyagi coast can be divided into three parts (Fig. 2): municipalities of Kesennuma, Motoyoshi, Minamisanriku, Onagawa, and part of Ishinomaki presenting ria-type coast, municipalities of Yamamoto, Iwanuma, Watari, Natori, Sendai, Higashimatsushima and the remaining part of Ishinomaki presenting plain-type coast, and municipalities surrounding Matsushima Bay. The data around Matsushima Bay are
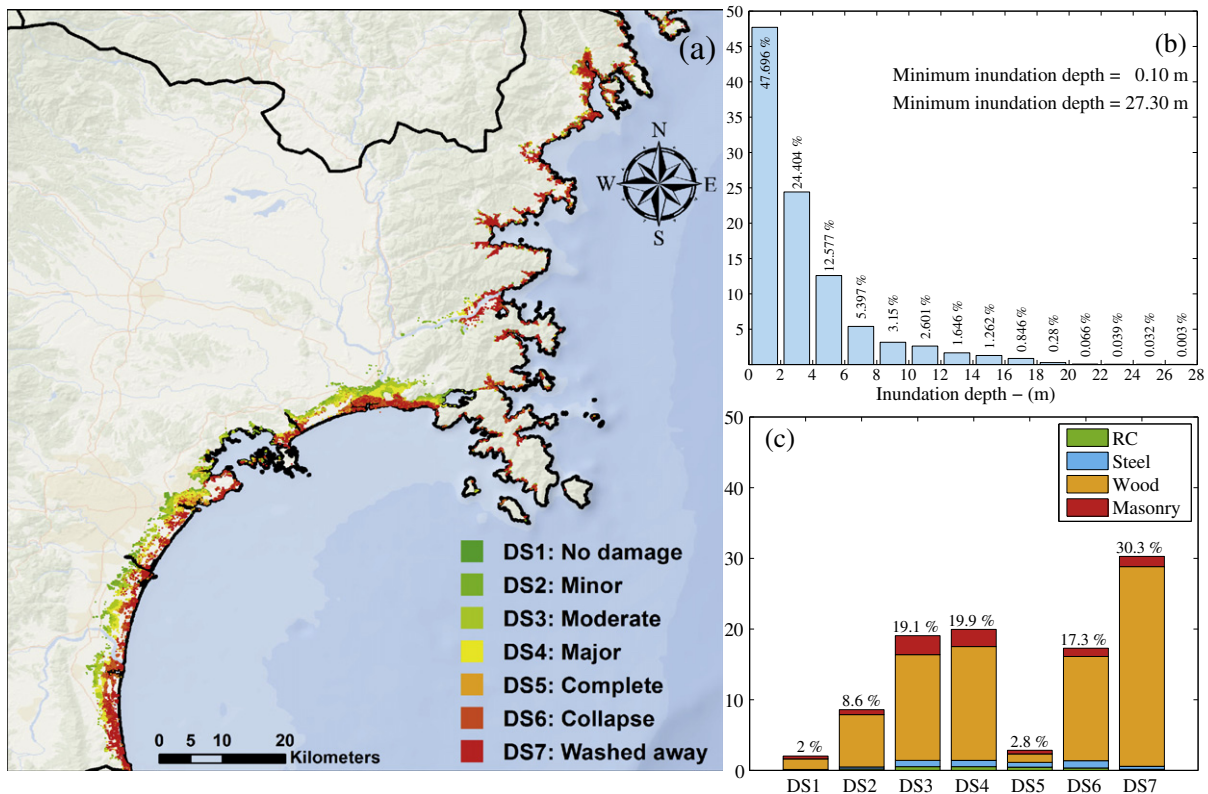

Fig. 1. (a) Spatial distribution of the MLIT damage data in Miyagi prefecture; (b) distribution of inundation depths; and (c) distribution of damage states for different structural typologies.
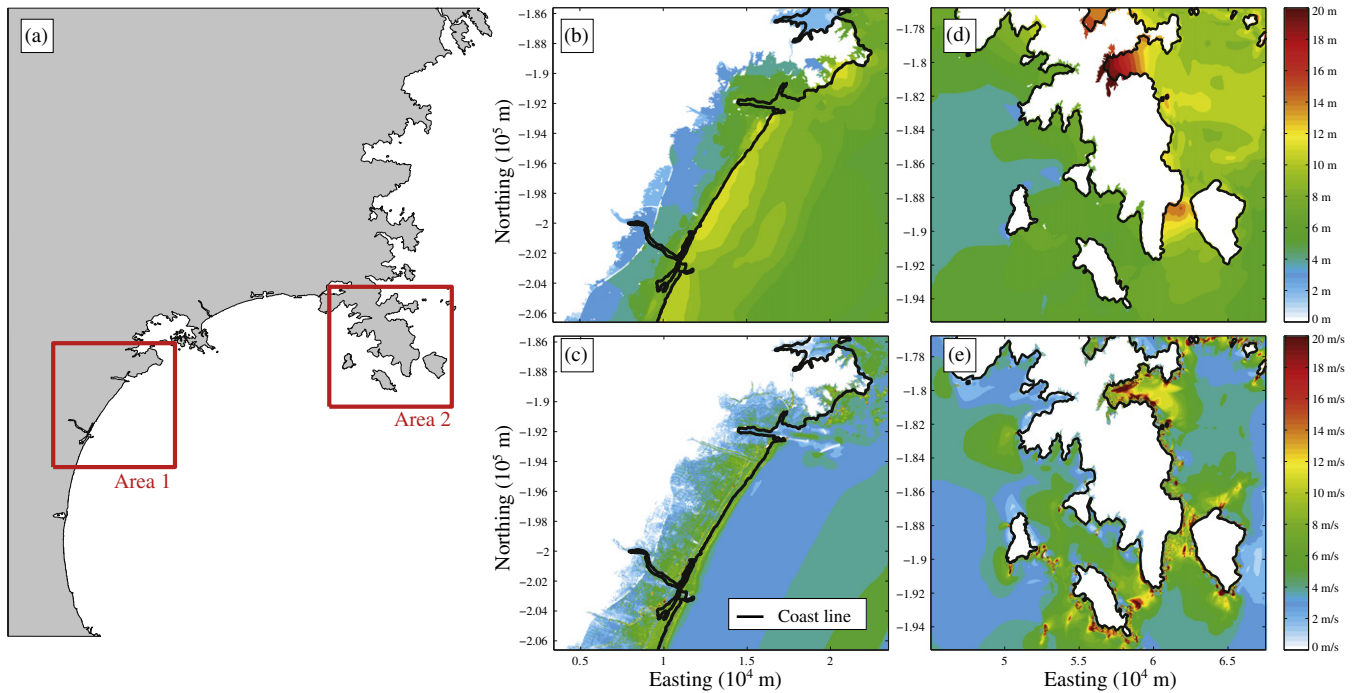
Fig. 3. (a) Considered areas in Miyagi prefecture; (b, c) maximum simulated inundation depths and velocities in Area 1; (d, e) maximum simulated inundation depths and velocities in Area 2. Results are based on the Satake et al.
(source model with 10-m resolution DEM.)

included in the tsunami fragility analysis considering the entire database (Section 4.2), but are excluded from the refined vulnerability models that account for topographical effects explicitly (Section 4.3). This is because the inner areas of Matsushima Bay are protected by numerous islands that are located near the mouth of the bay and thus tsunami inundation experienced in these areas is significantly less than other areas in Miyagi prefecture.

The coastal topography has major influence on inundation depth and velocity. For the same inundation depth, the ria-type coast leads to a faster flow velocity in the first hundreds of meters from the shoreline, and consequently greater damage to buildings is expected (Suppasri et al., 2013). Therefore, in the following, the topographical aspect is considered explicitly and its effect on the fragility modeling is analyzed in detail. Fig. 1(a) shows that along the plain-type coast, there is a gradual transition from the severest damage state (i.e. red dots corresponding to DS7) to no damage (i.e. green dots corresponding to DS1), meanwhile it is difficult to observe such a trend along the ria-type coast. This is because the plain-type coast allows a gradual inundation with similar characteristics in depth and velocity for points at the same distance from the coast, whereas the ria-type coast experiences different inundation depth and velocity, depending on the local bathymetry and topography.

Finally, it is important to underline that the damage states must be completely exhaustive and mutually exclusive in the multinomial regression analysis for developing tsunami fragility models. Furthermore, they should be increasing with respect to a tsunami intensity measure. In this context, the original DS6 and DS7 are two different descriptions of a collapse mode. For these reasons, in the following, these two damage states are combined and six damage states are considered (DS6 and DS7 are integrated into D6&7).

2.2. Simulation of flow velocity

This section summarizes a numerical procedure to obtain the simulated flow velocity data at the building locations. Firstly, key features of the tsunami input data (e.g. DEM and surface roughness) are described. Subsequently, salient information of the tsunami source models that are adopted in this study is given. The accuracy of the simulated tsunami intensities (i.e. inundation depth and flow velocity) is evaluated by comparing with the inundation depth data compiled in the MLIT database. The accuracy check of the flow velocity is done based on that of inundation depth because no reliable flow velocity data were observed at the building locations. Several studies carried out detailed comparisons of the calculated flow velocity and the observed flow velocity using recorded videos at specific locations (Fritz et al., 2012; Hayashi and Koshimura, 2013). However, it is not feasible to extend such investigations to all building locations in the MLIT database.

2.2.1. Input data for tsunami simulation

A complete dataset of bathymetry/elevation, coastal/riverside structures (e.g. breakwater and levees), and surface roughness is obtained from the Miyagi prefectural government. The data are provided in the form of nested grids (1350-m–450-m–150-m–50-m–10-m), covering the geographical regions of Tohoku. The low-lying land areas along the coast are covered by the 10-m grids for accurate inundation modeling.

The ocean-floor topography data are based on the 1:50,000 bathymetric charts and JTOPO30 database developed by Japan Hydrographic Association and based on the nautical charts developed by Japan Coastal Guard. The raw data are gridded using triangulated irregular network. The land elevation data are based on the 5-m grid DEM developed by the Geospatial Information Authority of Japan. The raw data are based on airborne laser surveys and aerial photographic surveys. These data have measurement errors of less than 1.0 m horizontally and of 0.3 m to 0.7 m vertically (as standard deviation). The tidal fluctuation is not taken into account in this study.

The elevation data of the coastal/riverside structures are primarily provided by municipalities. In the coastal/riverside structural dataset, structures having dimensions less than 10 m only are represented, noting that those having dimensions greater than 10 m are included in the DEM. In the tsunami simulation, the coastal/riverside structures are represented by a vertical wall at one or two sides of the computational cells. To evaluate the volume of water that overpasses these walls, Homma's overflowing formulae are employed.
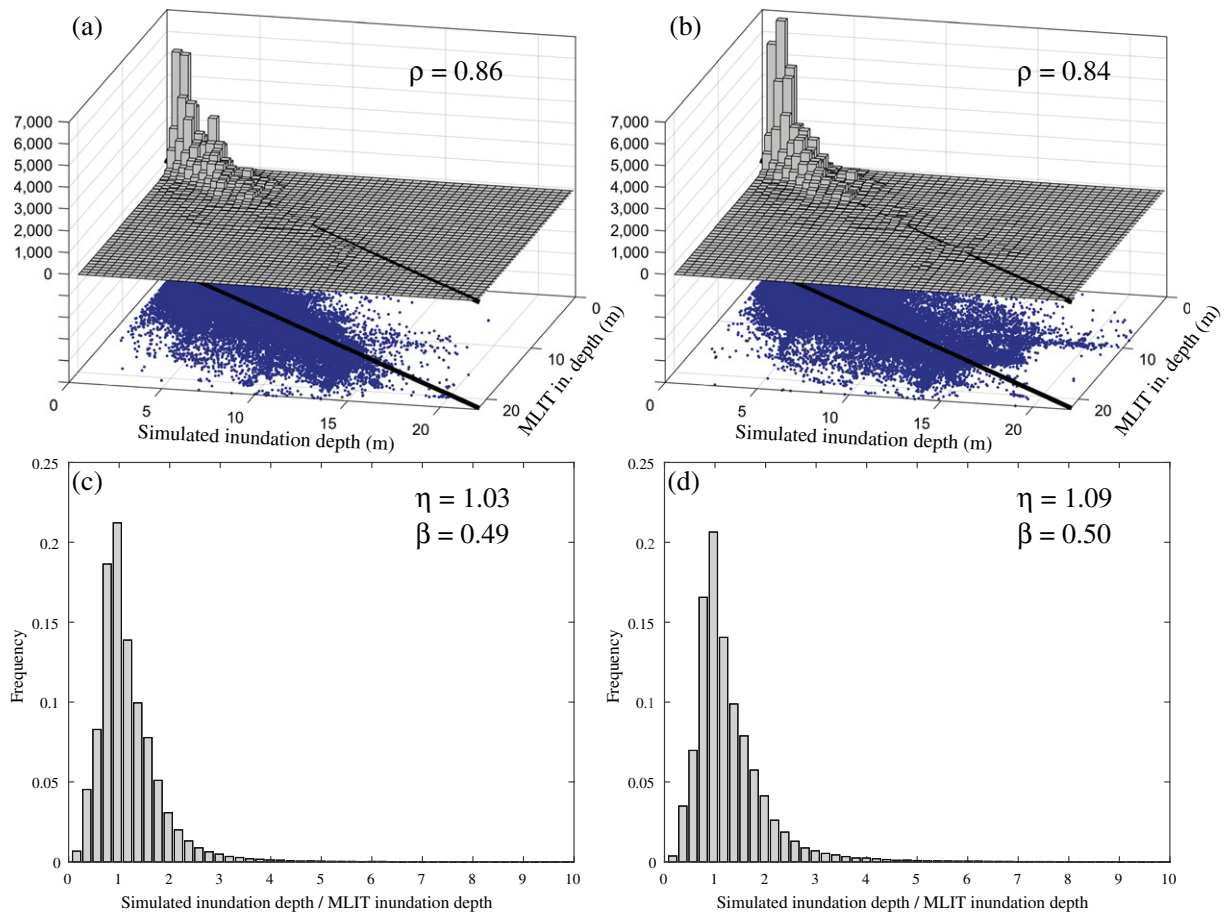
Fig. 4. Comparison of the MLIT and simulated inundation data based on the Satake et al. model: scatter plot and 3D histogram of the MLIT and simulated inundation depth data for (a) 10-m and (b) 50-m resolution, respectively, and histograms of the ratio between the MLIT and simulated inundation depth data for (c) 10-m and (d) 50-m resolution, respectively.

In the tsunami simulation, the bottom friction is evaluated using Manning's formula. The Manning's coefficients are assigned to computational cells based on national land use data in Japan: $0.02 \text{ m}^{-1/3} \text{ s}$ for agricultural land, $0.025 \text{ m}^{-1/3} \text{ s}$ for ocean/water, $0.03 \text{ m}^{-1/3} \text{ s}$ for forest vegetation, $0.04 \text{ m}^{-1/3} \text{ s}$ for low density residential areas, $0.06 \text{ m}^{-1/3} \text{ s}$ for moderate density residential areas, and $0.08 \text{ m}^{-1/3} \text{ s}$ for high density residential areas. The Manning's coefficients depend on the resolution of the available DEM because using refined grids more detailed roughness conditions can be represented in the simulation, resulting in more reliable intensity measure estimation can be obtained (Kaiser et al., 2011).

2.2.2. Source models and tsunami simulation for the 2011 Tohoku earthquake

Using an appropriate source model in tsunami modeling is important as this affects the simulated tsunami results significantly (Goda et al., 2014). This issue becomes even more critical when the simulated flow velocity is used as input to develop tsunami fragility models in terms of inundation depth as well as flow velocity (note: inundation depth data are available from the MLIT database). In this study, three inversion-based source models for the 2011 Tohoku earthquake are considered. The base model is set to the source inversion by Satake et al. (2013), which was developed using tsunami data. The Satake et al. model performs best among the eleven source models for the 2011 Tohoku earthquake (Goda et al., 2014; see also Section 2.2.3). As alternatives to the Satake et al. model, two additional models by Ammon et al. (2011) and Iinuma et al. (2012) are considered. The

former was developed based on geodetic-teleseismic joint source inversion, whereas the latter was developed based on geodetic source inversion. The aim of using different source models is to evaluate the robustness of conclusions obtained from the tsunami fragility analyses regarding the accuracy of calculated flow velocity data. It is noteworthy that the effects of biased input data have not been investigated in the literature; quantifying such biases is useful for the cases where the observed inundation data are not well constrained and hence performance assessment of different source models involves greater uncertainty.

Kinematic rupture processes are considered for the Satake et al. and Ammon et al. models, while instantaneous rupture process is adopted for the Iinuma et al. model. Subsequently, vertical seafloor displacements are calculated using Okada (1985) and Tanioka and Satake (1996) formulations. Assuming incompressibility of water, the vertical seafloor displacement is identically translated to the sea surface, representing the initial condition of the tsunami propagation model.

Tsunami propagation is carried out using a well-tested numerical code (Goto et al., 1997). It solves nonlinear shallow water equations using a leap-frog staggered-grid finite difference scheme and is capable of generating off-shore tsunami propagation and inundation/run-up. The run-up calculation is based on a moving boundary approach, where a dry/wet condition of a computational cell is determined based on total water depth relative to its elevation. The numerical tsunami calculation is performed for duration of 2 h, time sufficient to model most critical phases of tsunami waves. The integration time step is determined by satisfying the C.F.L. condition; it depends on the
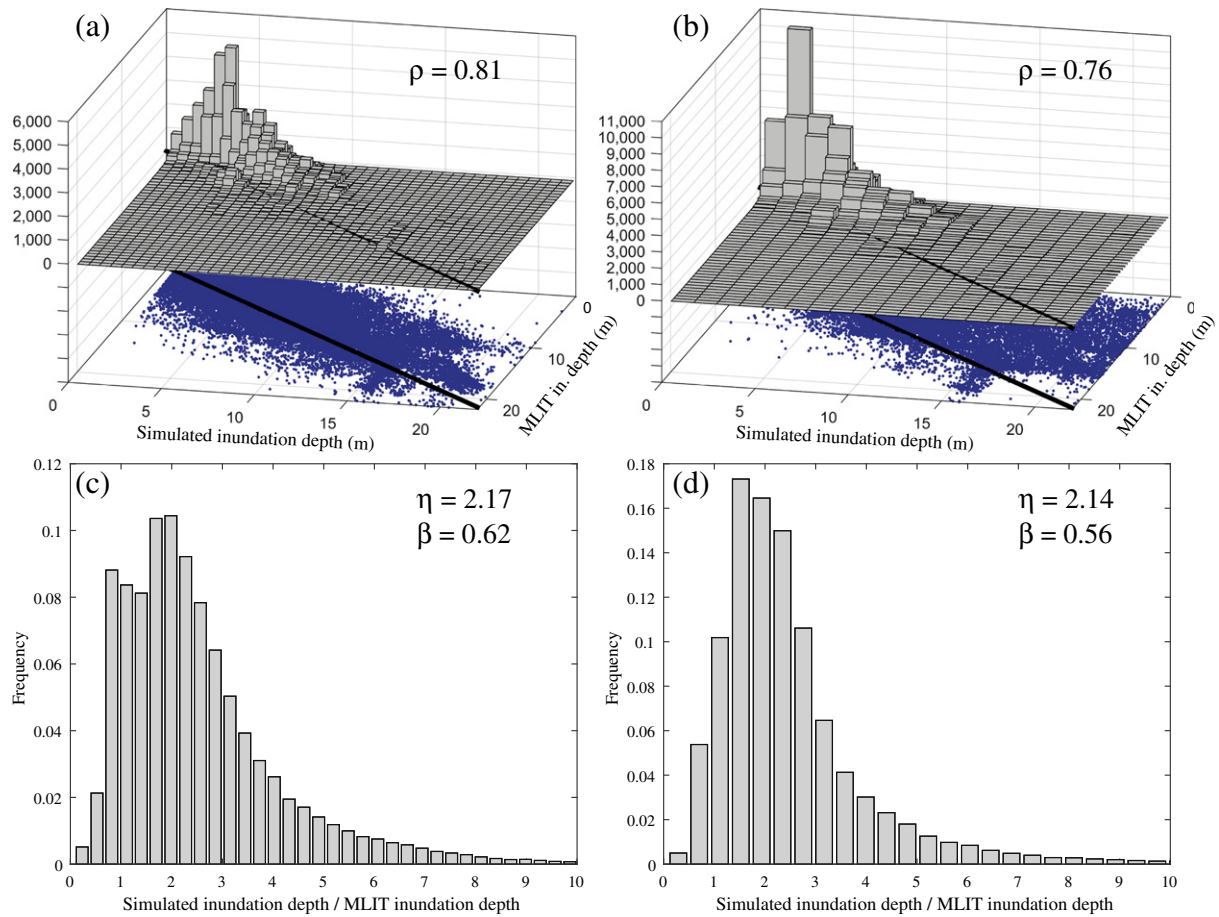
Fig. 5. Comparison of the MLIT and simulated inundation data based on the 10-m resolution DEM: scatter plot and 3D histogram of the MLIT and simulated inundation depth data based on the (a) Ammon et al. model and (b) Iinuma et al. model, respectively, and histograms of the ratio between the MLIT and simulated inundation depth data for the (c) Ammon et al. model and (d) Iinuma et al. model, respectively.

bathymetry/elevation data and their grid sizes and is typically between 0.1 s and 0.5 s.

Through the aforementioned procedure, the maximum inundation depth and velocity are calculated for all surveyed buildings in the MLIT database. As an example, Fig. 3 shows the results in terms of maximum inundation depth and maximum flow velocity (i.e. the square root of sum of squares (SRSS) of the two velocity components) for two areas in Miyagi prefecture; area 1 and area 2 are representative of plain-type coast and ria-type coast, respectively. The results are based on the Satake et al. source model and 10-m resolution DEM. It can be observed that in ria-type coast, there is an evident increase of maximum velocity with respect to the case of plain-type coast, as observed by Suppasri et al. (2011). This increase can be justified considering that for the same value of flow discharge, nearer to the coast, there is a reduction of the cross section due to the particular "V" shape of the bays.

2.2.3. Comparison of the simulated and observed tsunami intensity parameters

The simulated flow velocity data at the building locations of the MLIT database depend on the data resolutions and the source models in tsunami modeling. In Section 4.2.2, the effects of the resolution and source model on the tsunami fragility results will be investigated. Prior to these, influence of the resolution and source model on the calculated flow velocity (i.e. input data to the fragility analysis) is evaluated in this section. More specifically, four cases are considered: (i) the Satake et al. source model with 10-m resolution DEM (base case), (ii) the Satake et al. source model with 50-m resolution DEM, (iii) the

Ammon et al. source model with 10-m resolution DEM, and (iv) the Iinuma et al. source model with 10-m resolution DEM.

Figs. 4 and 5 show the comparisons between recorded inundation data and maximum simulated inundation depths. For both figures, the top two plots (i.e. a and b) represent the scatter plots between recorded and simulated inundation depth; the 3D histogram gives a more clear idea about the density of the scatter plot. The bottom two plots (i.e. c and d) represent the histograms of ratio between simulated and observed inundation depth. It is observed that this ratio is well fitted by a lognormal distribution having the median η and the logarithmic standard deviation β.

Fig. 4 shows the results based on the Satake et al. model, using 10-m (a and c) and 50-m (b and d) resolution DEM data. The ratios between simulated and observed inundation depths shown in Fig. 4(c and d) indicate that the results are relatively unbiased. A high coefficient of correlation (i.e. 0.86 and 0.84 for the two cases) suggests a general consistency of the simulations based on the Satake et al. model with the observations. Moreover, adopting finer resolution data, performance of the tsunami simulation model is improved (i.e. less dispersion and greater correlation). Fig. 5 shows the results based on the Ammon et al. and Iinuma et al. source models, using 10-m resolution DEM. It can be observed that in both cases, the median ratios are biased with respect to the observations, and there is a reduction of the coefficient of correlation and an increase of dispersion with respect to those for the Satake et al. model (Fig. 4). The model proposed by Iinuma et al. (2012) is less correlated than the model proposed by Ammon et al. (2011), but on average is closer to the MLIT data.
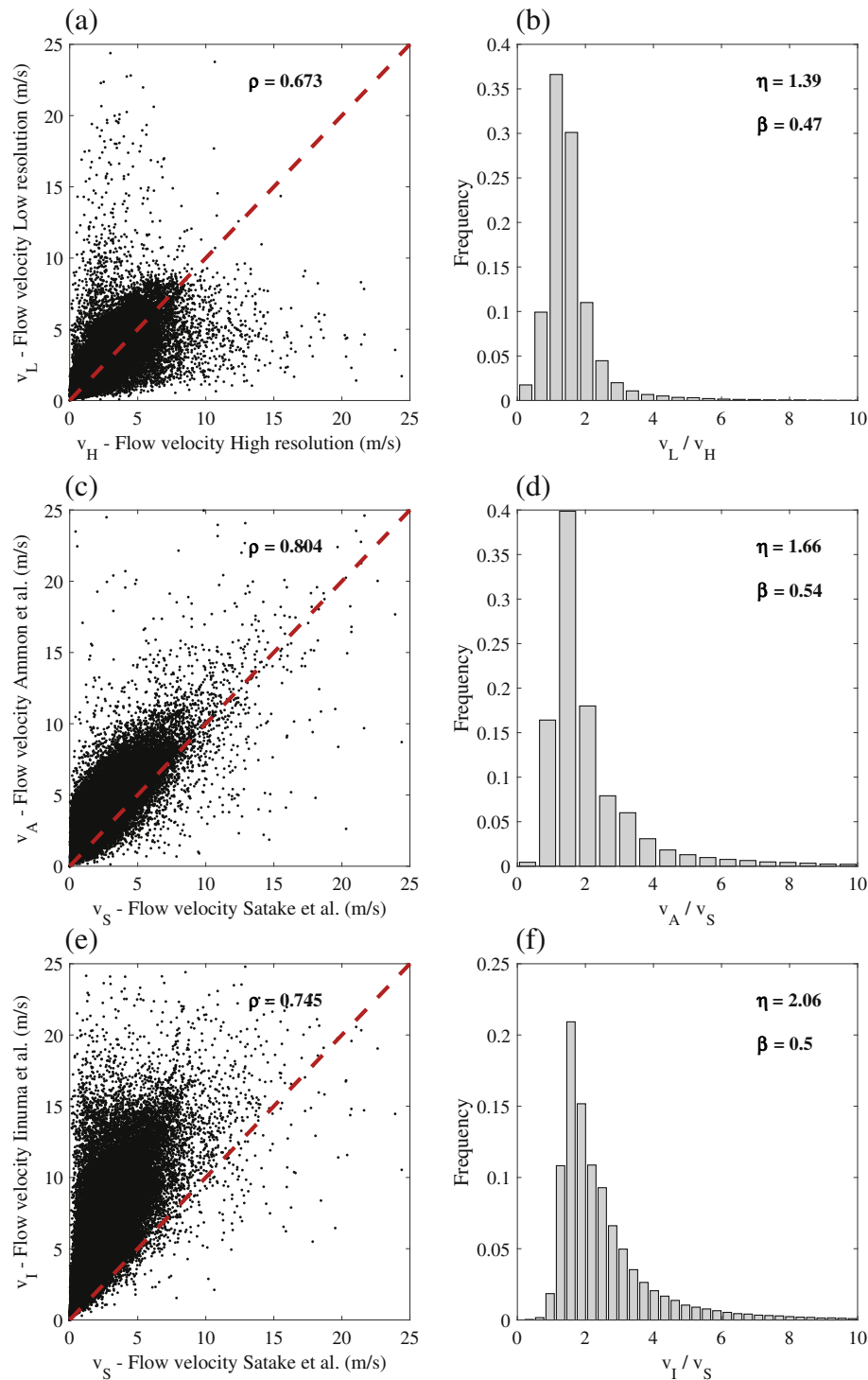
Fig. 6. (a, c, e) Comparison between the maximum flow velocity values obtained with the three additional tsunami simulations (Satake et al. model with 50-m resolution DEM, Ammon et al. model with 10-m resolution DEM, and Iinuma et al. model with 10-m resolution DEM) and the reference case (Satake et al. model with 10-m resolution DEM). (b, d, f) Histograms of the ratio of flow velocity for the three aforementioned cases with respect to the reference case. ρ is the coefficient of correlation between the two sets of flow velocity data. η and β are the median and logarithmic standard deviation of the ratio, respectively.

The results show that using different input source models or different DEM resolutions, variations with respect to the observed data occur. These variations may be used to characterize probability distributions describing the input data uncertainty that can be eventually considered in fragility modeling, but this is out of the scope of this paper. Finally, it is noted that inundation data from tsunami damage database are not used to carry out source inversion. The source model can be further calibrated against observed inundation data to improve the matching between observed and calculated flow intensity measures. Such a source model is more suitable to obtain the simulated flow velocity dataset required for tsunami fragility modeling.

2.2.4. Comparison of the simulated tsunami flow velocities

Results from the previous subsection suggest that the inundation model calculated through the Satake et al. model and 10-m resolution DEM is more reliable and thus can be used as reference. This subsection

Table 1
Goodness-of-fit metrics based on diagnostic analysis (Roman numbers in parenthesis indicate the resulting model order according to the considered parameter).

Model	AIC		BIC		Deviance		Number of parameters
M1	183,273	(V)	183,515	(V)	183,223	(V)	25
M2	178,948	(III)	179,239	(III)	178,888	(III)	30
M3	182,142	(IV)	182,530	(IV)	182,062	(IV)	40
M4	177,983	(II)	178,419	(I)	177,893	(II)	45
M5	177,851	(I)	178,432	(II)	177,731	(I)	60

compares the simulated flow velocity data obtained from the reference inundation model and the three additional inundation models.

Fig. 6(a and b) compares the maximum flow velocity values obtained from the simulations carried out using the same source model, but with different DEM datasets. Despite the high correlation between the two sets of flow velocity data, the values obtained using the 50-m resolution DEM show a larger dispersion and are on average greater than those obtained using the more refined DEM.

Fig. 6(c–f) shows the comparisons between the simulated maximum flow velocity values based on the two alternative source models and the velocity data based on the reference model. It can be observed that in both cases, there is a good correlation between simulated values. Moreover, there is an increase, on average, of flow velocity values and an increase of dispersion. The correlation between the velocity values is almost the same for the two cases, but the Ammon et al. case presents a greater bias with respect to the Iinuma et al. model, as observed for inundation depth in Fig. 5.

3. Bivariate tsunami fragility modeling using flow velocity

To examine the significance of incorporating flow velocity in tsunami fragility modeling, rigorous regression analyses are carried out, developing bivariate fragility curves based on inundation depth and flow velocity (as well as other parameters, such as structural material type). Detailed explanations of the adopted analytical methods are given in the following.

3.1. Fragility modeling

Empirical tsunami fragility functions relate building damage to a tsunami intensity measure (i.e. inundation depth), taking into account other explanatory variables, such as structural typology, number of floors, and topographical indicators. In this study, multinomial logistic regression is adopted for developing tsunami fragility models. The method is an extension of binomial logistic regression to the case of multiclass problems (i.e. *polytomous* response), belonging to the family of Generalized Linear Models. This type of regression can handle more than two damage states simultaneously, and does not require binning of the data. It also allows the consideration of the ordered and hierarchical nature of damage states, avoiding inconsistent results, such as intersection of fragility functions. It is noted that previous studies typically aggregate the damage data in bins having similar tsunami intensity values, and then they fit a simple or generalized linear statistical model to the binned data. This approach is not considered to avoid biases in estimated regression parameters due to binning. For instance, it is well-known that the grouping in bins can affect fragility curves
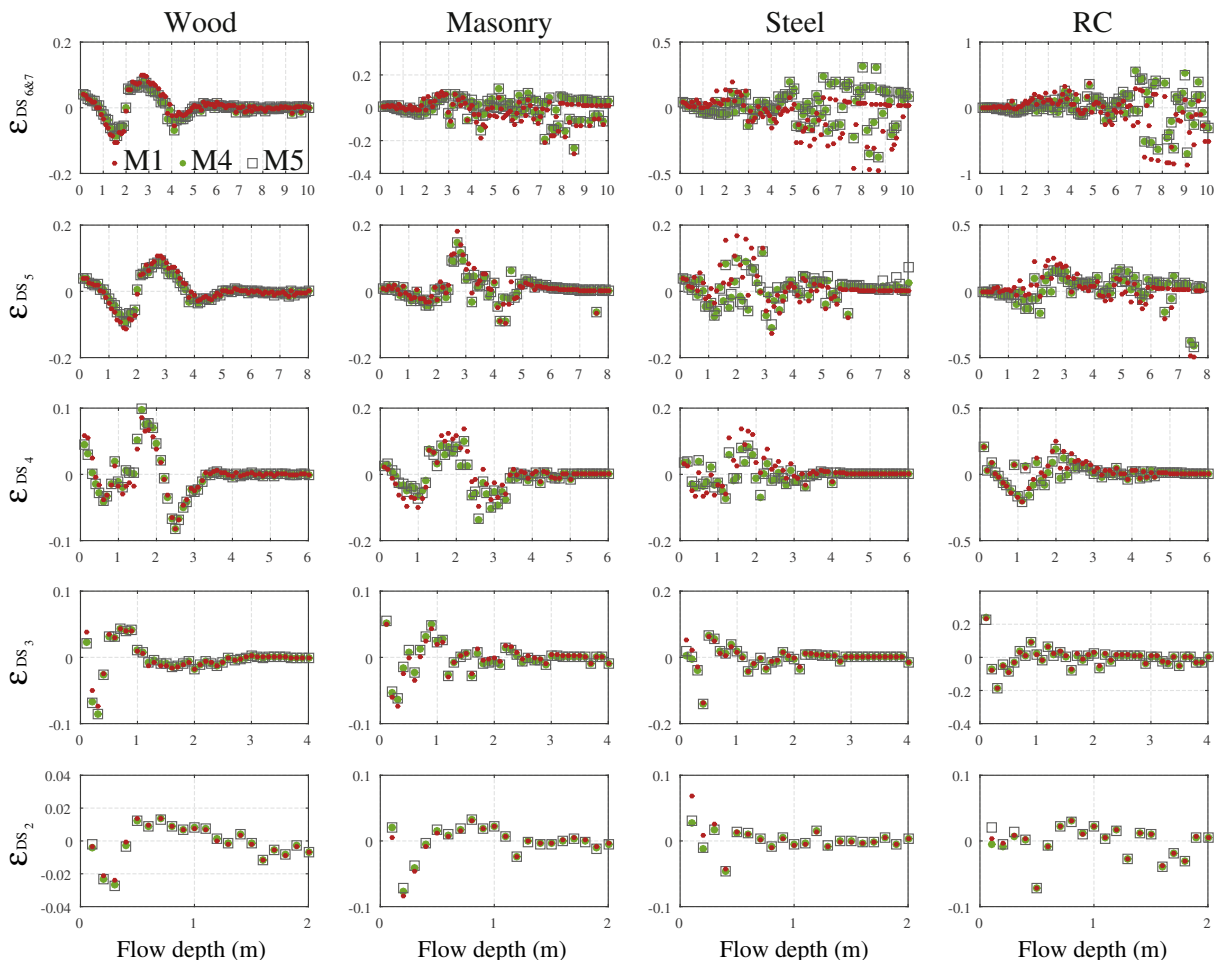

Fig. 7. Raw residuals (observation minus prediction) based on the five models for all damage states and for all structural typologies.

Table 2
Deviances (G^2) from all structural typologies and all damage states based on **M3** and **M4**.

Damage state	Wood		Masonry		Steel		RC	
	M3	**M4**	**M3**	**M4**	**M3**	**M4**	**M3**	**M4**
DS1	7935	7934	1798	1782	364	362	207	204
DS2	21,335	20,765	2496	2488	879	865	553	547
DS3	35,459	34,599	5064	5053	1854	1783	1112	1081
DS4	41,516	40,351	5545	5471	2514	2466	1512	1499
DS5	10,748	10,774	2936	2896	2520	2478	1369	1340
DS6&7	26,913	25,907	3816	3723	2427	2362	1189	1162

especially at the tails and that bins that present extreme probabilities (i.e. 0 and 1, which correspond to no damage and complete damage, respectively) are systematically dismissed during the classical statistical linear fitting (Charvet et al., 2014). Furthermore, the classical linear least squares fitting is not recommended in the case of discrete and ordinal variables, such as tsunami damage states (Charvet et al., 2014).

Consider that the damage state (DS) that takes one of the six discrete values, i.e. ds1, ds2, …, and ds6&7. Let

$$\pi_{ij} = P(DS_i = ds_j) \tag{1}$$

denote the probability that the ith observation falls in the jth category. As the damage states are mutually exclusive and completely exhaustive, the sum of the damage state probabilities equals one for each observation. The probability that all buildings of the ith bin fall in the respective damage state is given by the multinomial probability distribution:

$$P(DS_{i1} = ds_1, \ldots, DS_{i6\&7} = ds_{6\&7}) = \frac{m_i!}{\prod_{j=1}^{6} y_{ij}!} \cdot \prod_{j=1}^{6} \pi_{ij}^{y_{ij}} \tag{2}$$

where m_i denotes the total number of structures corresponding to the ith bin (always equal to 1 in this application) and y_{ij} denotes the number of structures from the ith bin attaining the jth damage state ds_j (i.e. 1 or 0 for damage state attainment or not, respectively). The distribution shown in Eq. (2) represents the random component of the model,

which describes the distribution of the response around the central value.

The systematic component of the model relates the probability π_{ij} and a vector of explanatory variables x, and is represented by a link function. Usually, it is a linear function of explanatory variables:

$$f(\pi_{ij}) = \theta_{j,0} + \sum_{k=i}^{p} \theta_{j,k} \cdot x_k \tag{3}$$

where θ is the vector of the model regression parameters $\theta_{j,k}$, and p is the number of explanatory variables. When the model regression parameters are the same for all damage states (except for the intercept $\theta_{j,0}$), the resultant fragility curves having the same slope and the model is referred to as ordered. Relaxing the previous constraint, the model is called as partially ordered. In the following, only partially ordered models are considered to avoid over-constraining the fitting.

Several link functions are suggested in the literature (Hosmer et al., 2013). In this study, probit and logit models are adopted:

$$\text{Probit}: f(\pi_{ij}) = \Phi^{-1}(\pi_{ij}) \text{ and } \text{Logit}: f(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \sum_{k=1}^{j} \pi_{ik}}\right) \tag{4}$$

where $\Phi(\bullet)$ represents the standard normal cumulative distribution function. Depending on the link function, the regression procedure is commonly known as multinomial logit or multinomial probit regression.

The point estimates for the model parameters are obtained based on the maximum likelihood estimation (MLE) approach by computing the first and second derivatives of the likelihood function that is expressed as follows:

$$L(\theta|x, y) = \prod_{i=1}^{n} \prod_{j=1}^{6} \pi_{ij}^{y_{ij}} \tag{5}$$

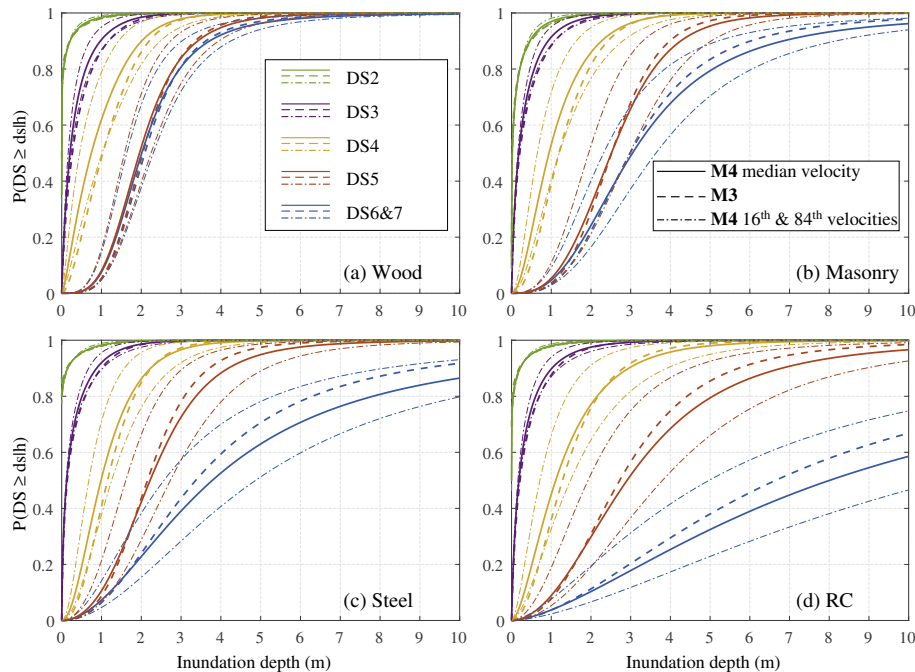where n is the number of data points.

Fig. 8. Empirical fragility functions for four different structural typologies according to **M3** (dashed lines) and **M4** (solid lines for median velocity value, i.e. 1.84 m/s, and dashed-dotted lines for the 16th and 84th percentile velocity values, i.e. 1.00 m/s and 5.80 m/s, respectively).

Table 3
Regression parameters for **M3** considering the entire Miyagi prefecture.

Parameter	DS2	DS3	DS4	DS5	DS6&7
θ_0	1.545	1.429	−0.823	−3.109	−3.251
θ_1	0.053	0.951	2.079	2.666	1.716
θ_2	0.224	−0.392	0.638	−2.517	0.307
θ_3	−0.808	0.007	0.280	−2.231	−0.177
θ_4	−0.189	0.168	0.207	−0.532	0.562
θ_5	0.079	−0.025	−0.362	1.032	2.281
θ_6	0.138	−0.019	0.203	1.654	1.418
θ_7	−0.093	−0.128	0.261	1.061	0.499

Table 5
Regression parameters for **M4** considering the entire Miyagi prefecture.

Parameter	DS2	DS3	DS4	DS5	DS6&7
θ_0	1.540	1.430	−0.904	−3.462	−3.722
θ_1	0.105	0.924	1.650	2.299	1.557
θ_2	0.186	−0.375	0.768	−2.255	0.738
θ_3	−0.845	0.023	0.437	−1.989	0.215
θ_4	−0.186	0.169	0.213	−0.417	0.696
θ_5	0.058	−0.019	−0.257	0.906	2.023
θ_6	0.108	−0.010	0.270	1.486	1.160
θ_7	−0.118	−0.122	0.319	0.878	0.354
θ_8	−0.137	0.062	0.698	0.848	0.696

3.2. Model selection

Different explanatory variables (and their combinations) can be taken into account for the systematic component, leading to several statistical models. In particular, it is necessary to identify the model that provides the best fit in comparison with the available alternatives. A diagnostic analysis or model selection can be performed by assessing the relative goodness-of-fit of the candidate models. In this study, the model selection is conducted at two levels, identifying: 1) the best link function for the same model (i.e. two models with the same linear predictors but different link functions), and 2) the best model among several candidate nested models for the same link function (i.e. the more complex model includes at least all the parameters of its simpler counterpart).

Herein, two criteria, i.e. the Bayesian Information Criterion (BIC; Schwarz, 1978), and the Akaike Information Criterion (AIC; Akaike, 1974), are considered:

$$BIC = -2 \ln[L(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})] + k \cdot \ln n \tag{6}$$

$$AIC = -2 \ln[L(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})] + 2k \tag{7}$$

where $L(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})$ denotes the data likelihood under the MLE of a candidate model, k is the number of regression parameters θ, and n is the number of observations. The model that presents the smallest value of BIC or AIC is considered to provide a better fit to the available data. It is noted that for selecting a suitable nested model, several statistical tests can be used to assess the relative goodness-of-fit (e.g. the likelihood-ratio test, the F test, and the analysis of variance). However, these tests are unreliable for un-binned binary data (McCullagh and Nelder, 1989; Hosmer et al., 2013). For this reason, the preceding diagnostic criteria are used consistently. In the context of model selection, generally BIC tends to choose models that are more parsimonious than those favored by AIC.

In addition, another statistic, i.e. residual deviance G^2, is considered to measure the overall model performance with respect to the data. This measure compares the proposed model (i.e. a model with a small number of parameters) to a saturated one (i.e. a model with parameters equal to the number of observations), and is expressed as follows:

$$G^2 = 2 \sum_{i=1}^{n} \sum_{j=1}^{6} y_{ij} \ln \frac{y_{ij}}{\pi_{ij}}. \tag{8}$$

A model with smaller deviance is preferred (i.e. the model is close to the saturated one, presenting zero residual deviance). It is worth noting that for un-binned data there are no over-dispersion issues (i.e. model dispersion is greater than data dispersion). In fact, when over-dispersion occurs, the overall goodness-of-fit is distorted. This problem has never been treated rigorously in the previous studies related to tsunami fragility model selection.

4. Numerical results

In this section, various tsunami vulnerability models are developed by considering different functional forms (Section 4.1); the goodness-of-fit of the model is evaluated based on the three diagnostic criteria. In Section 4.2, a base vulnerability model is developed using the entire MLIT data in Miyagi prefecture and the flow velocity data obtained from the reference inundation model (i.e. Satake et al. model and 10-m resolution DEM). The robustness of the vulnerability model with respect to the accuracy of simulated flow velocity data using different inundation models is examined. Furthermore, through detailed residual analysis, the effects due to coastal topography (plain-type versus ria-type) are investigated in Section 4.3. Finally, based on the results of the residual analysis, two vulnerability models accounting for coastal topographical effects are proposed.

4.1. Functional forms

Five different models of the linear predictor function are investigated (Eq. (3)). The first model (**M1**) considers the logarithm of inundation depth h and structural classes only:

M1

$$f(\pi_{ij}) = \theta_{j,0} + \theta_{j,1} \log(h_i) + \theta_{j,2} \cdot d_W + \theta_{j,3} \cdot d_M + \theta_{j,4} \cdot d_S \tag{9}$$

where d_W, d_M, and d_S are the dummy variables for the wooden, masonry, and steel structures, respectively and take a value of 1 when the ith observation falls in the respective category, and 0 otherwise. There are three dummy variables, not four (i.e. the number of structural classes), to avoid over-parametrization. The second model (**M2**) expands by accounting for inundation velocity v:

Table 4
Fragility statistics for all damage states and for all structural typologies based on **M3**.

DS	Wood η (m)	Wood β (−)	Masonry η (m)	Masonry β (−)	Steel η (m)	Steel β (−)	RC η (m)	RC β (−)
DS2	0.00	4.50	0.01	4.25	0.00	4.50	0.00	4.50
DS3	0.27	1.42	0.21	1.56	0.14	1.82	0.21	1.54
DS4	1.03	0.76	1.20	0.63	1.15	0.60	1.28	0.68
DS5	2.04	0.40	2.52	0.40	2.15	0.47	2.72	0.61
DS6&7	2.09	0.41	2.99	0.53	3.37	0.75	6.65	0.97

Table 6
Regression parameters for **M4** considering plain-type coast only.

Parameter	DS2	DS3	DS4	DS5	DS6&7
θ_0	1.795	1.742	−1.283	−3.346	−3.826
θ_1	0.307	1.355	1.925	1.564	1.701
θ_2	−0.159	−0.803	1.105	−3.300	0.548
θ_3	−1.352	−0.196	0.754	−1.679	0.144
θ_4	−0.696	−0.089	0.406	−0.697	0.589
θ_5	−0.051	−0.499	−0.819	1.415	2.001
θ_6	−0.088	−0.343	−0.271	0.375	1.263
θ_7	−0.332	−0.576	−0.212	1.152	0.296
θ_8	−0.211	−0.037	0.782	1.604	0.850

Table 7
Regression parameters for **M4** considering ria-type coast only.

Parameter	DS2	DS3	DS4	DS5	DS6&7
θ_0	81.396	1.685	−0.195	−2.294	−2.758
θ_1	1.638	0.827	0.458	1.774	1.063
θ_2	−76.758	−0.867	0.459	−3.692	0.432
θ_3	105.544	−0.606	0.767	−2.403	1.039
θ_4	21.370	0.223	0.486	−2.618	0.849
θ_5	−2.677	0.052	0.976	1.470	2.491
θ_6	−110.334	0.088	0.808	2.090	0.595
θ_7	−1.709	0.776	0.702	1.730	0.304
θ_8	−0.974	0.004	0.767	0.234	0.773

M2

$$f(\pi_{ij}) = \theta_{j,0} + \theta_{j,1} \log(h_i) + \theta_{j,2} \cdot d_W + \theta_{j,3} \cdot d_M + \theta_{j,4} \cdot d_S + \theta_{j,5} \log(v_i). \tag{10}$$

The third, fourth, and fifth models (**M3**, **M4**, and **M5**, respectively), are expansion of **M1** and **M2** and are aimed to account for the interaction between the intensity measures and the structural classes:

M3

$$f(\pi_{ij}) = \theta_{j,0} + \theta_{j,1} \log(h_i) + \theta_{j,2} \cdot d_W + \theta_{j,3} \cdot d_M + \theta_{j,4} \cdot d_S + \dots \\ + \theta_{j,5} \log(h_i) \cdot d_W + \theta_{j,6} \log(h_i) \cdot d_M + \theta_{j,7} \log(h_i) \cdot d_S \tag{11}$$

M4

$$f(\pi_{ij}) = \theta_{j,0} + \theta_{j,1} \log(h_i) + \theta_{j,2} \cdot d_W + \theta_{j,3} \cdot d_M + \theta_{j,4} \cdot d_S + \dots \\ + \theta_{j,5} \log(h_i) \cdot d_W + \theta_{j,6} \log(h_i) \cdot d_M + \theta_{j,7} \log(h_i) \cdot d_S + \dots \\ + \theta_{j,8} \log(v_i) \tag{12}$$

M5

$$f(\pi_{ij}) = \theta_{j,0} + \theta_{j,1} \log(h_i) + \theta_{j,2} \cdot d_W + \theta_{j,3} \cdot d_M + \theta_{j,4} \cdot d_S + \dots \\ + \theta_{j,5} \log(h_i) \cdot d_W + \theta_{j,6} \log(h_i) \cdot d_M + \theta_{j,7} \log(h_i) \cdot d_S + \dots \\ + \theta_{j,8} \log(v_i) + \dots \\ + \theta_{j,9} \log(v_i) \cdot d_W + \theta_{j,10} \log(v_i) \cdot d_M + \theta_{j,11} \log(v_i) \cdot d_S. \tag{13}$$

In particular, **M3** is a complete extension of **M1**, **M4** is a partial extension of **M2**, and **M5** is a complete extension of **M2**. Passing from **M1** to **M5**, the number of model regression parameters increases. The number of the required parameters for the five models is 25, 30, 40, 45, and 60, respectively.

4.2. Base fragility models and their variations

4.2.1. Base models

The goodness-of-fit to different link functions is first evaluated. The results indicate for all considered models, the smallest values of AIC and BIC are obtained for the logit function. Therefore, only the multinomial logit model is considered for further analyses.
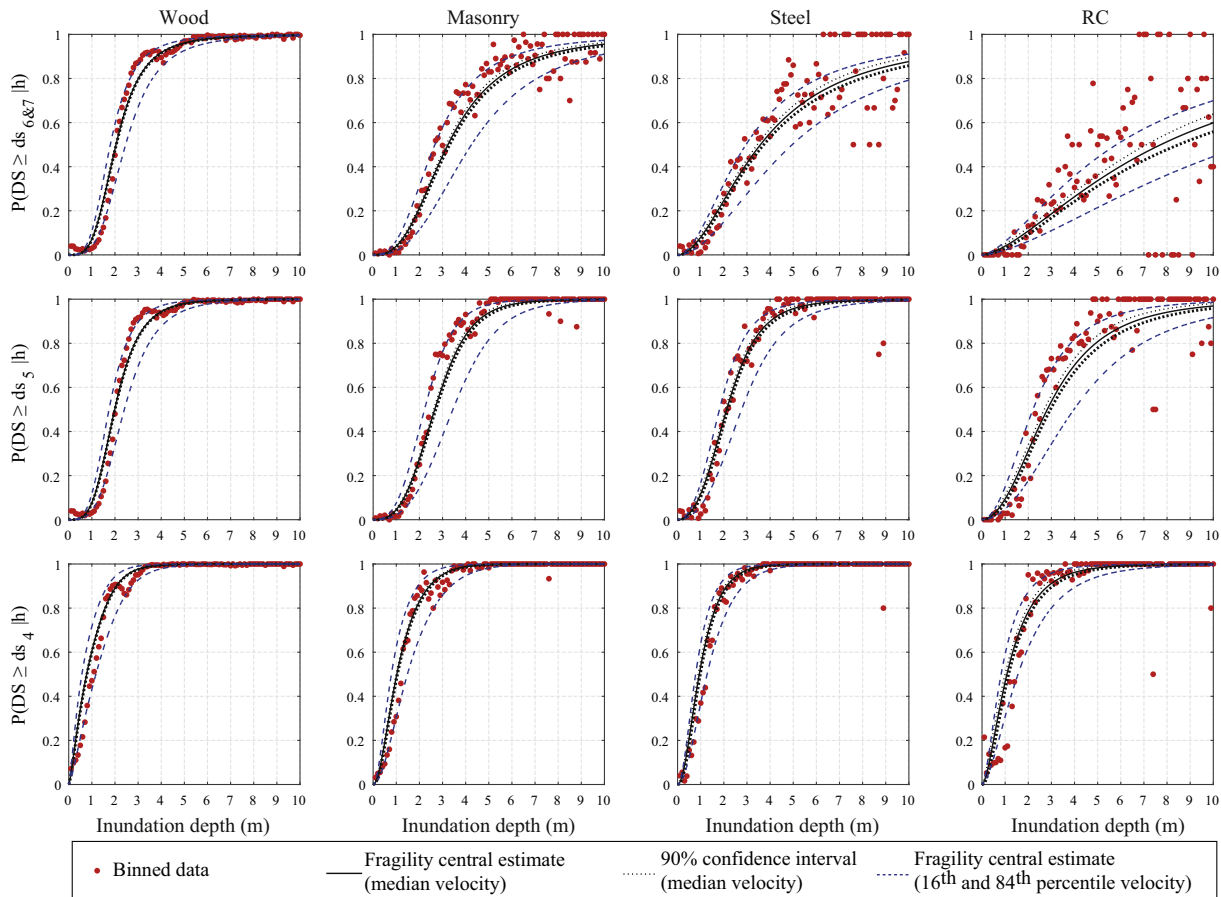

Fig. 9. Empirical fragility functions based on **M4** for DS4, DS5, and DS6&7 (and their corresponding 90% confidence intervals represented by black dotted lines) for four different structural typologies using median velocity for each structural category. The blue lines refer to empirical fragility functions based on **M4** considering the 16th and 84th percentile velocity values.
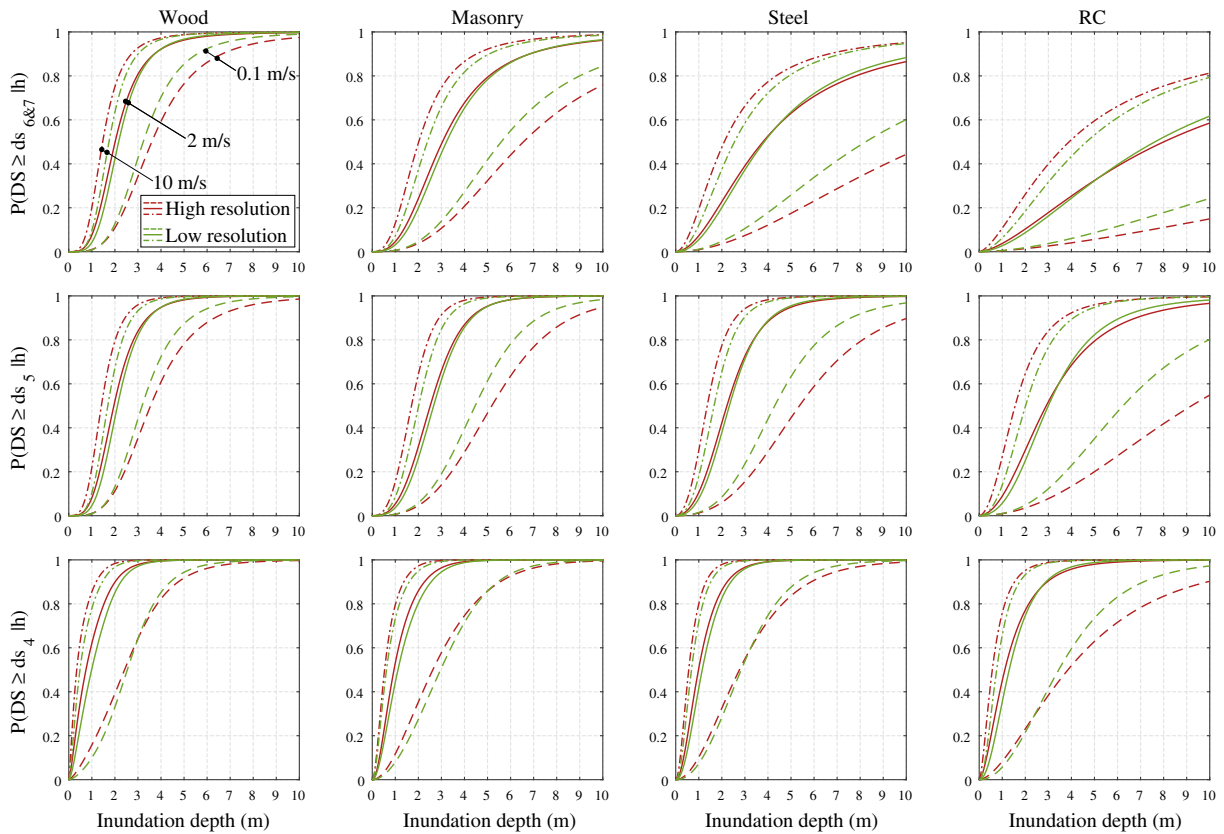
Fig. 10. Empirical fragility functions for four different structural typologies according to **M4** considering a velocity of 0.1 m/s (dashed line), 2 m/s (solid line), and 10 m/s (dashed-dotted line), and using data from simulations at 10-m resolution (red line) and 50-m resolution (green line).

Table 1 presents the results related to the model selection among the five models. The diagnostic analysis results based on AIC and deviance indicate that **M5** is the best model, followed in order by **M4**, **M2**, **M3**, and **M1**. On the other hand, the analysis based on BIC indicates that **M4** is more informative, followed in order by **M5**, **M2**, **M3**, and **M1**. In all considered cases, the models considering flow velocity (**M2**, **M4**, and **M5**) improve the goodness-of-fit in comparison with those neglecting flow velocity (**M1** and **M3**). It is therefore evident that **M4** and **M5** perform better in this first screening.

It is considered that **M4** may be preferred over **M5** for three reasons: (*i*) the BIC, more parsimonious among the chosen diagnostic metrics (since it prefers models with a less number of parameters), indicates **M4** as the best choice; (*ii*) the differences of all considered indicators (i.e. AIC, BIC, and deviance) between **M4** and **M5** are less than 1‰, thus the potential improvement from **M4** to **M5** is negligible despite the high increase in number of regression parameters (i.e. 45 versus 60); (*iii*) looking at the raw residuals in Fig. 7, it can be observed that residuals related to **M4** (green circle) and **M5** (grey square) overlap, and both are close to zero at a large inundation depth. Moreover, both **M4** and **M5** present residuals lower than **M1**, represented with red dots in Fig. 7, for a large part of inundation depths. Note that the differences decrease for high values of inundation depth. Concerning structural typologies, Fig. 7 indicates that residuals related to wooden structures are the lowest (given the great amount of data) and they progressively increase for masonry, steel, and RC structures. The latter is due to the fact that the number of observations for these three typologies of structures is smaller compared to the wooden buildings. It is therefore possible to conclude that considering the flow velocity in vulnerability modeling results in an improvement of conventional univariate models based on inundation depth. In addition, in this specific case, the model considering inundation depth, the interaction terms between inundation depth

and structural typologies, and flow velocity (**M4**), represents the best model for the vulnerability modeling.

To understand how the flow velocity improves the vulnerability modeling in terms of damage state and structural typology in detail, a disaggregation of deviance for each damage state and for each structural typology is carried out. **M4** is compared with **M3** which differs only in the inclusion of flow velocity. Table 2 lists the values of deviance for **M3** and **M4**. It can be observed that the deviance decreases more for damage states from DS4 to DS6&7 when the velocity is considered as additional explanatory variable, in comparison with less severe damage states. This fact can be explained by considering that the velocity affects more structural damage states than non-structural damage states.

Fig. 8 shows the fragility curves based on **M4**, for all damage states, for all structural materials, and for three velocity values obtained from the reference inundation model; the median value is represented by a solid line and the 16th and 84th percentiles are represented by dashed-dotted lines. Fig. 8 also includes fragility curves based on **M3** (dashed lines), neglecting flow velocity. Regression parameters according to Eqs. 11 and 12 are given in Tables 3 and 5 for **M3** and **M4**, respectively. Given the independence of **M3** from the velocity, the statistics in terms of inundation depth (i.e. median η and logarithmic standard deviation β) are also listed in Table 4, which are useful to build the lognormal approximation of **M3** according to the following equation:

$$P(DS \geq ds_i) = \Phi\left(\frac{\ln h - \ln \eta}{\beta}\right) \qquad (14)$$

It can be observed from Fig. 8 that for DS5 and DS6&7 fragilities corresponding to **M3** are contained in the interval between the fragility curves for **M4** corresponding to the 16th and 84th percentiles of flow velocity. The differences between the curves corresponding to the
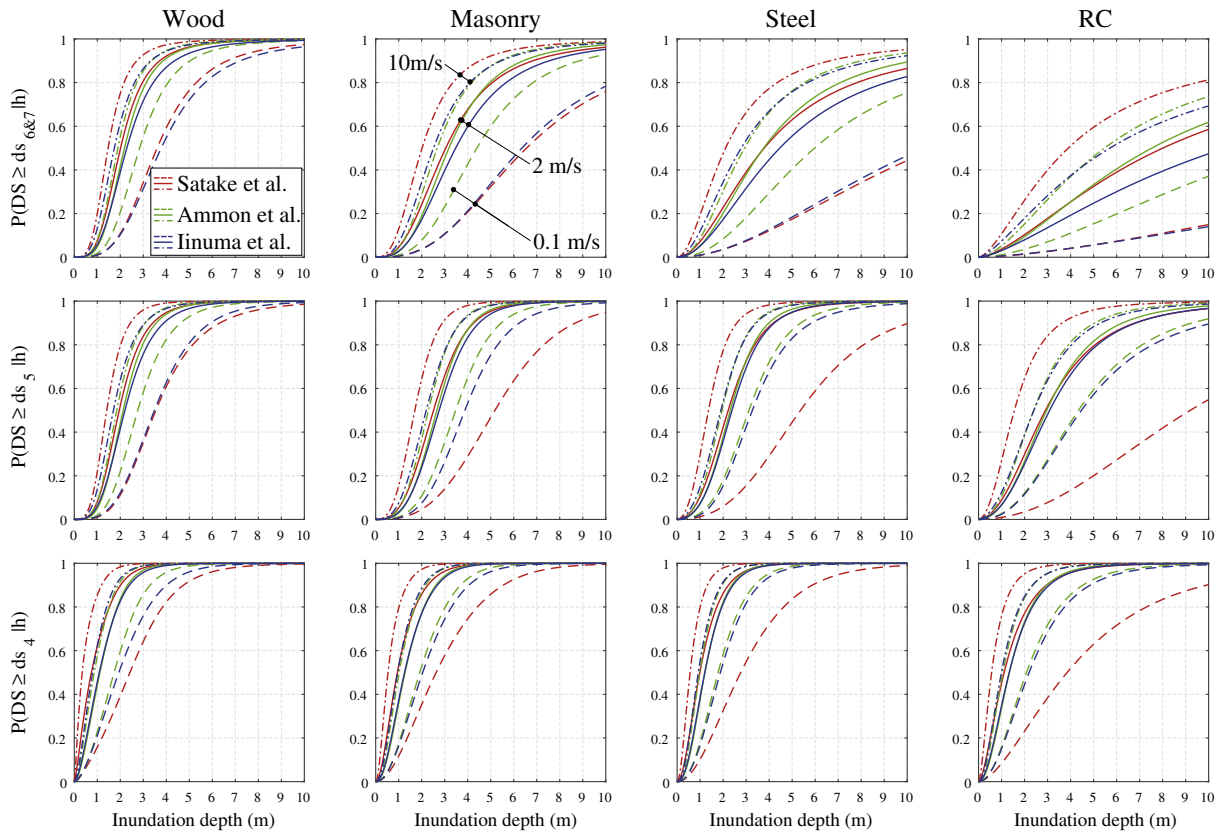
Fig. 11. Empirical fragility functions for four different structural typologies according to **M4** considering a velocity of 0.1 m/s (dashed line), 2 m/s (solid line), and 10 m/s (dashed-dotted line), and using data from simulation at 10-m resolution according to the Satake et al. model (red line), Ammon et al. model (green line), and the Iinuma et al. model (blue line).

16th and 84th percentiles of flow velocity are wider for steel and RC structures than wood and masonry structures. Thus the influence of the flow velocity becomes more significant to steel and RC structures. Those variations can affect significantly the risk analysis. The results obtained from **M4** also suggest that despite the consideration of a partially

ordered model, the fragility curves for each building class do not intersect one another.

Fig. 9 shows the fragility functions already presented in Fig. 8 separately for the structural damage states and for each material typology; the curves obtained for the median velocity value are reported together
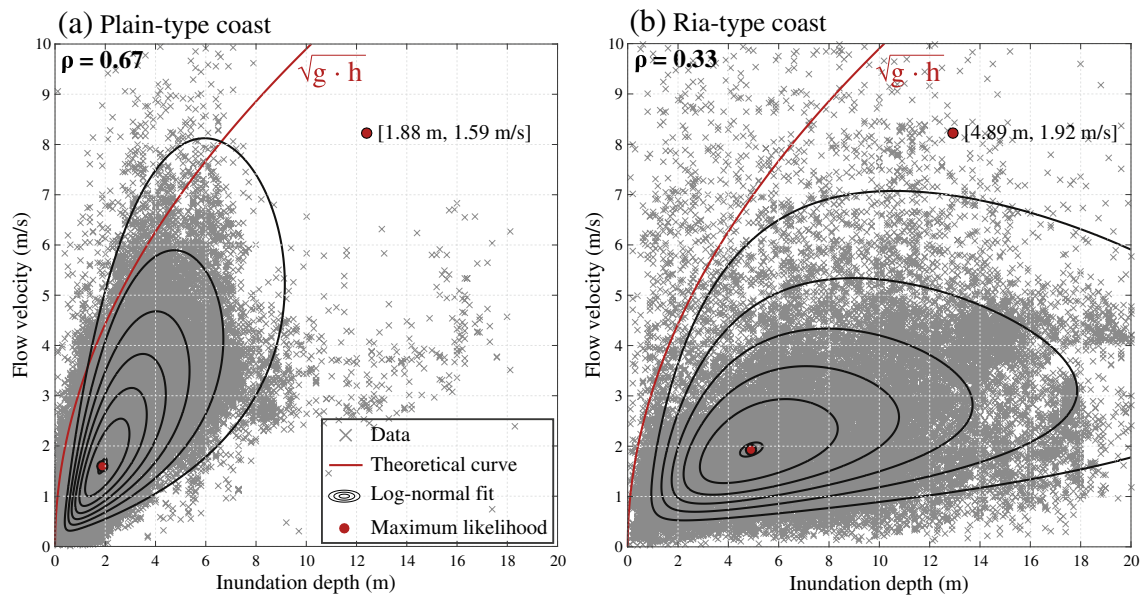

Fig. 12. Observed inundation depths versus calculated flow velocities for (a) plain-type coast and (b) ria-type coast. The red line represents the theoretical relation for shallow water between inundation depth (h) and flow velocity (v). g is the acceleration of gravity. The black contour is the log-normal fit of data. The red circle is the maximum likelihood of the distribution.
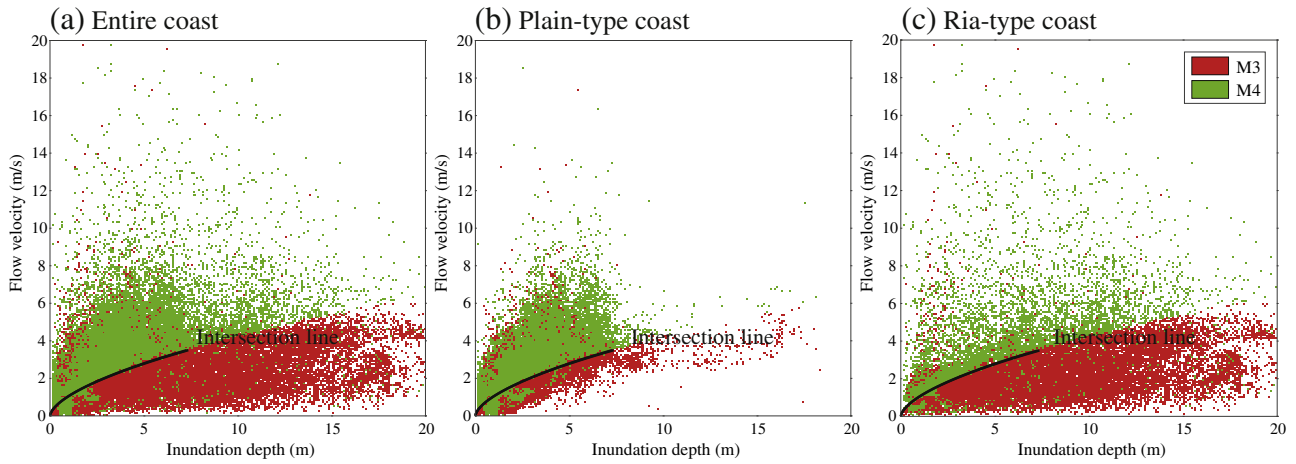
Fig. 13. Ratio between the SRSS of residuals for **M4** and **M3** related to (a) the entire coast, (b) plain-type coast only, and (c) ria-type coast only.

with the 90% asymptotic confidence intervals (dotted black lines). Since these intervals are strictly related to the number of observations and the available sample size is large, the intervals are very close to the central estimates. For this reason, the confidence intervals do not cross for different damage states. In addition, in the same figure, the red points are an alternative representation of the MLIT data; each point shows the frequency of reaching or exceeding a specific damage state within a bin having width equal to 0.1 m. These points can further help understand the performance of the developed models. It is observed that the points are well contained in the fragility intervals for **M4** corresponding
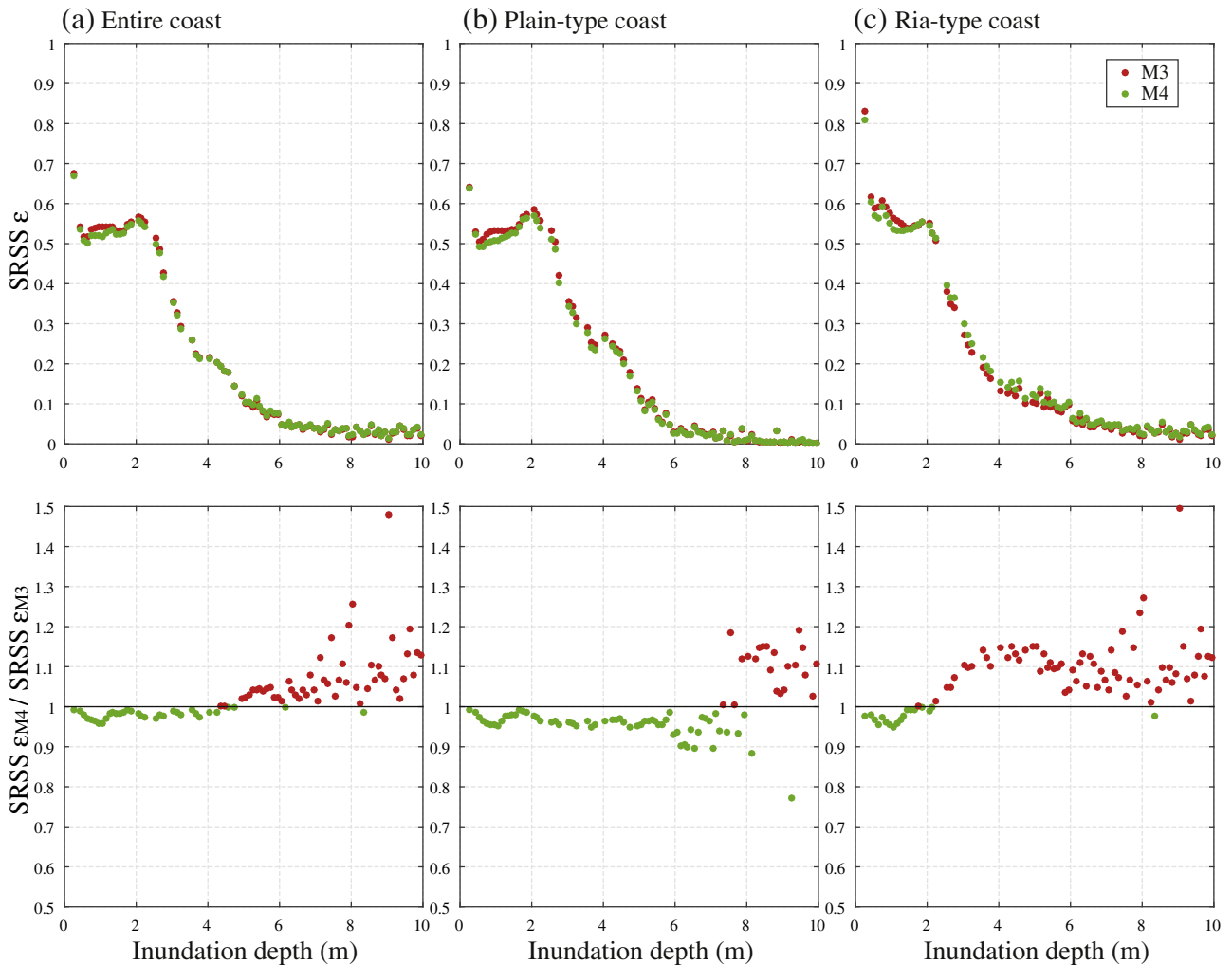

Fig. 14. SRSS of residuals and ratio between the SRSS of residuals for **M4** and **M3** as a function of inundation depth, related to (a) the entire coast, (b) plain-type coast only, and (c) ria-type coast only.
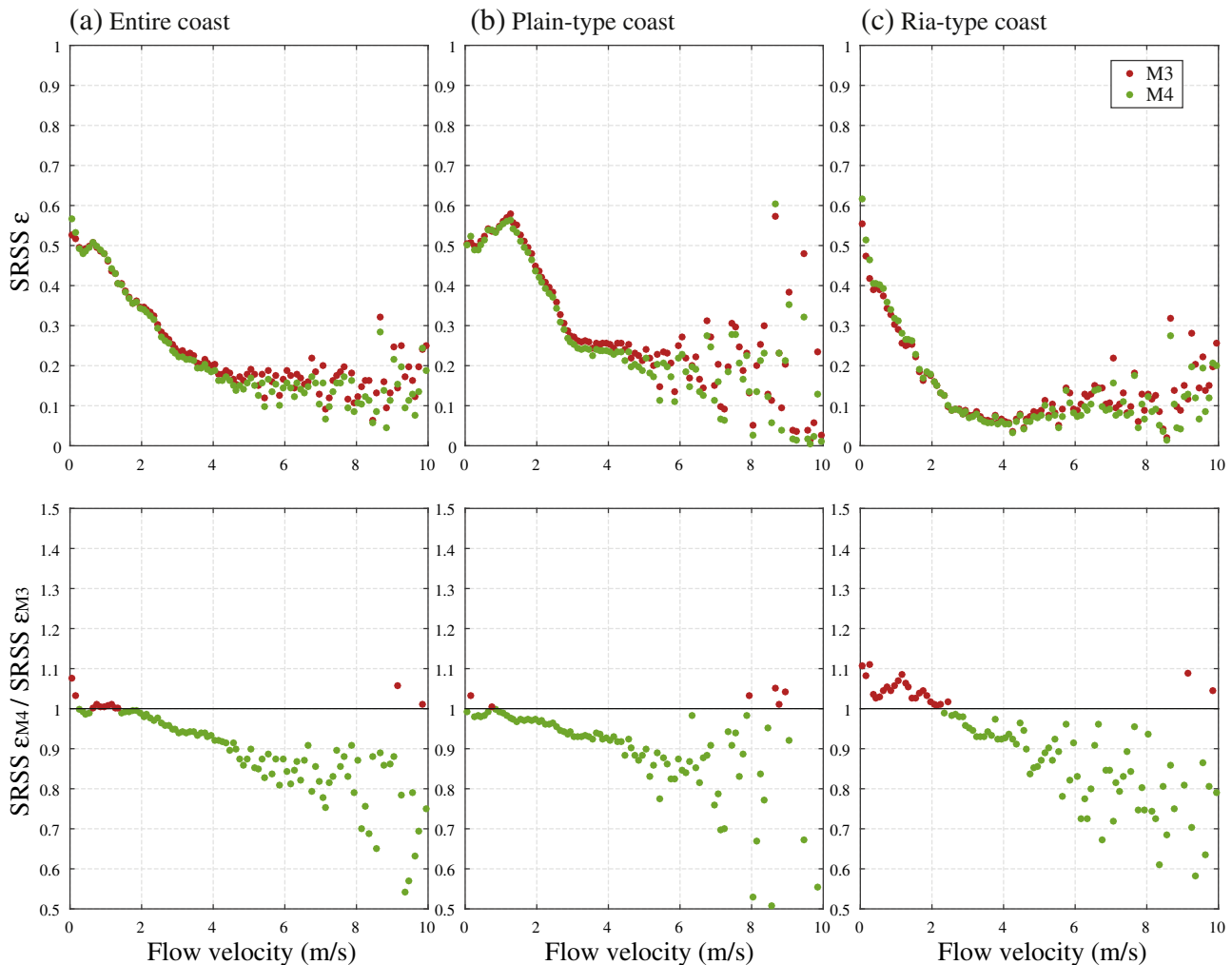
Fig. 15. SRSS of residuals and ratio between the SRSS of residuals for **M4** and **M3** as a function of flow velocity, related to (a) the entire coast, (b) plain-type coast only, and (c) ria-type coast only.

to the 16th and 84th percentiles of flow velocity, respectively (the blue dashed lines), especially for wood and masonry structures.

4.2.2. Effects of DEM resolution and source models

In this subsection, the effects due to the simulation accuracy are analyzed considering the three aforementioned tsunami simulation cases (Section 2.2.3). First, the model selection analysis for the three additional cases indicates that models considering inundation velocity are preferable with respect to the models considering inundation depth only. More specifically, **M4** is still the preferred one among the five functional forms mentioned in Section 4.1. The regression parameters obtained for the three additional cases are given in the supplementary electronic data associated with the paper (Appendix A).

Figs. 10 and 11 present the comparisons between fragilities based on the three additional variations and the base model. Fig. 10 compares fragilities obtained considering simulation based on the Satake et al. inversion model, but with different DEM resolutions. Fig. 11 shows the comparisons among the three considered inversion models, using the same DEM resolution. Since **M4** is a function of inundation depth and flow velocity, to facilitate the comparison, only the marginal fragility curves, in terms of inundation depth, and for three values of flow velocity are shown, i.e. 0.1 m/s, 2 m/s, and 10 m/s, representative of low, moderate, and high velocity values, respectively. With respect to the base model, a less refined inundation model (both in terms of DEM and source model) leads to lower median capacity for low values of velocity, and greater median capacity for high values of velocity.

Moreover, the model dispersion for structural damage states tends to decrease, while that for non-structural damage states tends to increase slightly. According to these observations, these variations in terms of fragility curves with respect to the base case lead to a potential risk underestimation for low values of velocity, and to a risk overestimation for high values of velocity. In relative terms, it is possible to conclude that the effects of the Iinuma et al. model on fragility curves are comparable with the effects due to the Satake et al. model with 50-m resolution DEM. All the previous results are still valid when data points having the supercritical flow velocity are removed in the fragility modeling.

The features of fragility functions obtained with the three alternative simulations are the direct consequences of the differences in maximum flow velocity. In fact, as observed in Fig. 6, for all the less refined cases, there is an average increase of velocity values and an increase of dispersions. Such data imply large values of damage probability for the small values of flow velocity, and then the large values of damage probability are reached for very high values of velocity. On the other hand, the model dispersion varies rapidly along the velocity axis for the refined model, while it varies smoothly (over a wider range of flow velocity) for the less refined models.

4.3. Models accounting for coastal topographical effects

The MLIT database offers a great opportunity to analyze the effect of different coastal topographies on the vulnerability models, since damaged buildings in Miyagi prefecture are distributed along plain-type
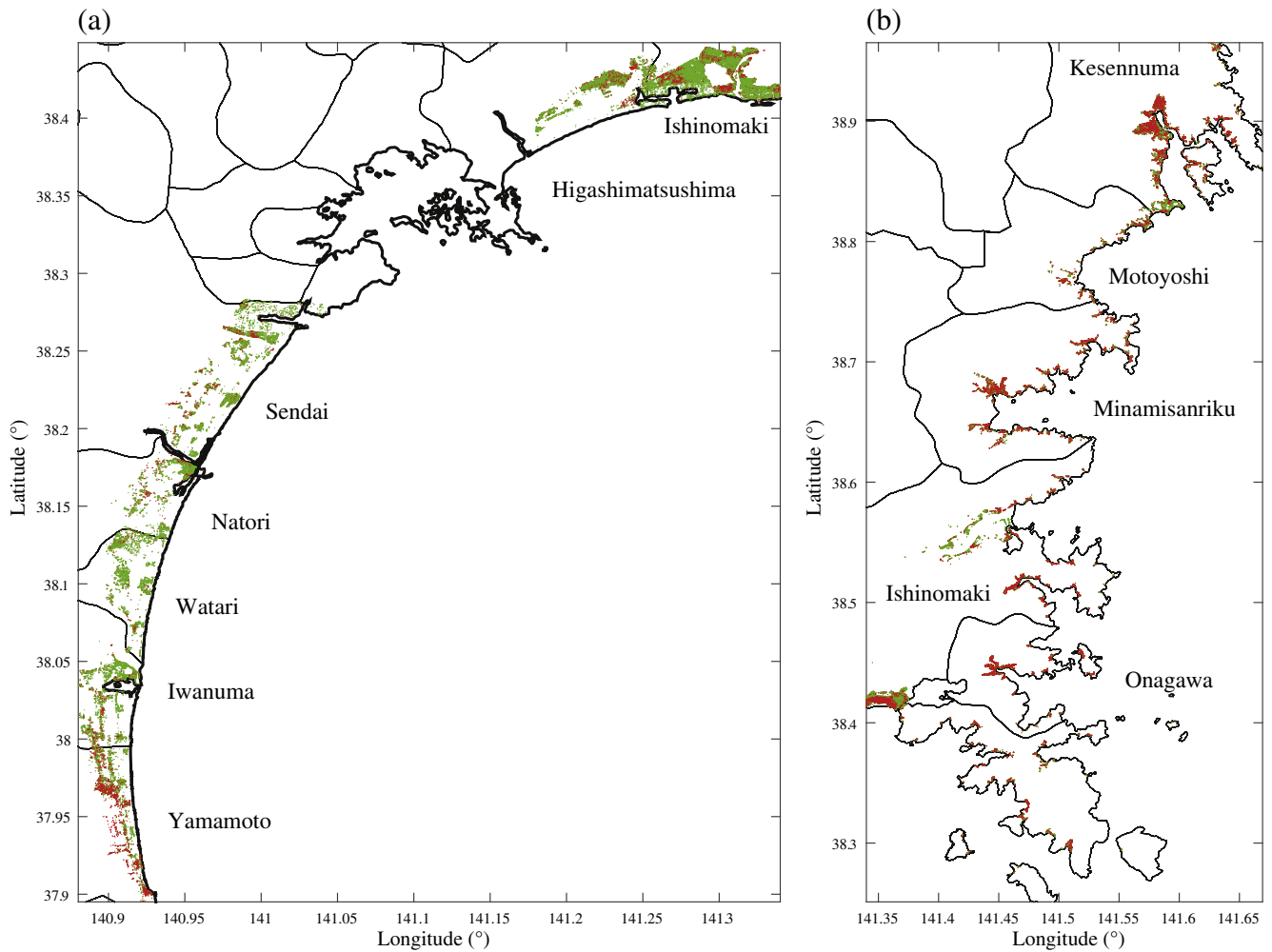
Fig. 16. Spatial representation of the importance of flow velocity in the vulnerability model for (a) plain-type coast and (b) ria-type coast.

and ria-type coasts (Fig. 2). As in Section 4.2.1, the reference inundation model (i.e. Satake et al. source model and 10-m resolution DEM) is considered herein for the further analysis. Fig. 12 shows the differences between the two coastal typologies. Fig. 12(a) shows that measured inundation depths and calculated maximum flow velocities are moderately correlated in plain-type coast. On the other hand, Fig. 12(b) shows that in ria-type coast, the correlation is low and the data are spread along the inundation depth axis. In Fig. 12, the curve representing the flow velocity for shallow water waves; it can be observed that some points exceed the theoretical value. Moreover, the median values of inundation depth and flow velocity are slightly greater for ria-type coast with respect plain-type. Finally, the number of structures that experienced large values of inundation depth in the ria-type coast is higher than the plain-type coast. The latter aspect leads to a higher number of structures experiencing structural damage, and has a direct influence on the shape of fragility functions. Therefore, two distinct vulnerability models accounting explicitly for the coastal topography could be more detailed with respect to the base model presented in Section 4.2.1. To demonstrate this point, residual analysis between observations and the reference vulnerability model is carried out.

For each observation, the SRSS of raw residuals between the base fragility model and observations is focused on. To be consistent with Section 4.2.1, **M4** is compared with **M3**. Fig. 13 shows the ratio between the SRSS of raw residuals obtained with **M4** and the SRSS of residuals obtained with **M3**, for different bins of inundation depth and flow velocity, by considering the entire coast (Fig. 13(a)), the plain-type coast only (Fig. 13(b)), and the ria-type coast only (Fig. 13(c)). The green color

indicates that the ratio is less than one, i.e. **M4** is more suitable than **M3**, whereas the red colors indicate the opposite. The results essentially suggest that inclusion of flow velocity is useful when the inundation depth is not extremely high. From a physical point of view, this can be interpreted that the flow velocity has less influence on damage when the inundation depth is high, especially on structures with less tsunami resistance, such as wooden and masonry structures. In the same figures, continuous black lines represent the boundary of the intersection points between the fragilities obtained with the two models, which are applicable to all material types and all structural damage states. Those lines automatically identify the separation between green and red regions. In the red region, the fragility curves for **M3** are under the fragility curves for **M4**; whereas in the green region the opposite is true.

To better understand the goodness of the base model for plain-type and ria-type cases, the residuals calculated individually with respect to inundation depth and flow velocity are presented in Figs. 14 and 15, respectively. These figures show that for the plain-type coast it is important to consider the velocity as additional explanatory variable for the greatest part of the inundation depth and velocity intervals (i.e. inundation depth between 0 m and 8 m and flow velocity between 0 m/s and 8 m/s). On the other hand, for the ria-type coast there are improvements for inundation depth between 0 m and 2 m and for flow velocity greater than 2 m/s. The previous results can be also presented according to the geographical distribution of the observed/calculated intensity measures for each building location. Fig. 16 shows the spatial distribution of observations indicating the importance (green dots) or unimportance (red dots) to consider the velocity in the fragility
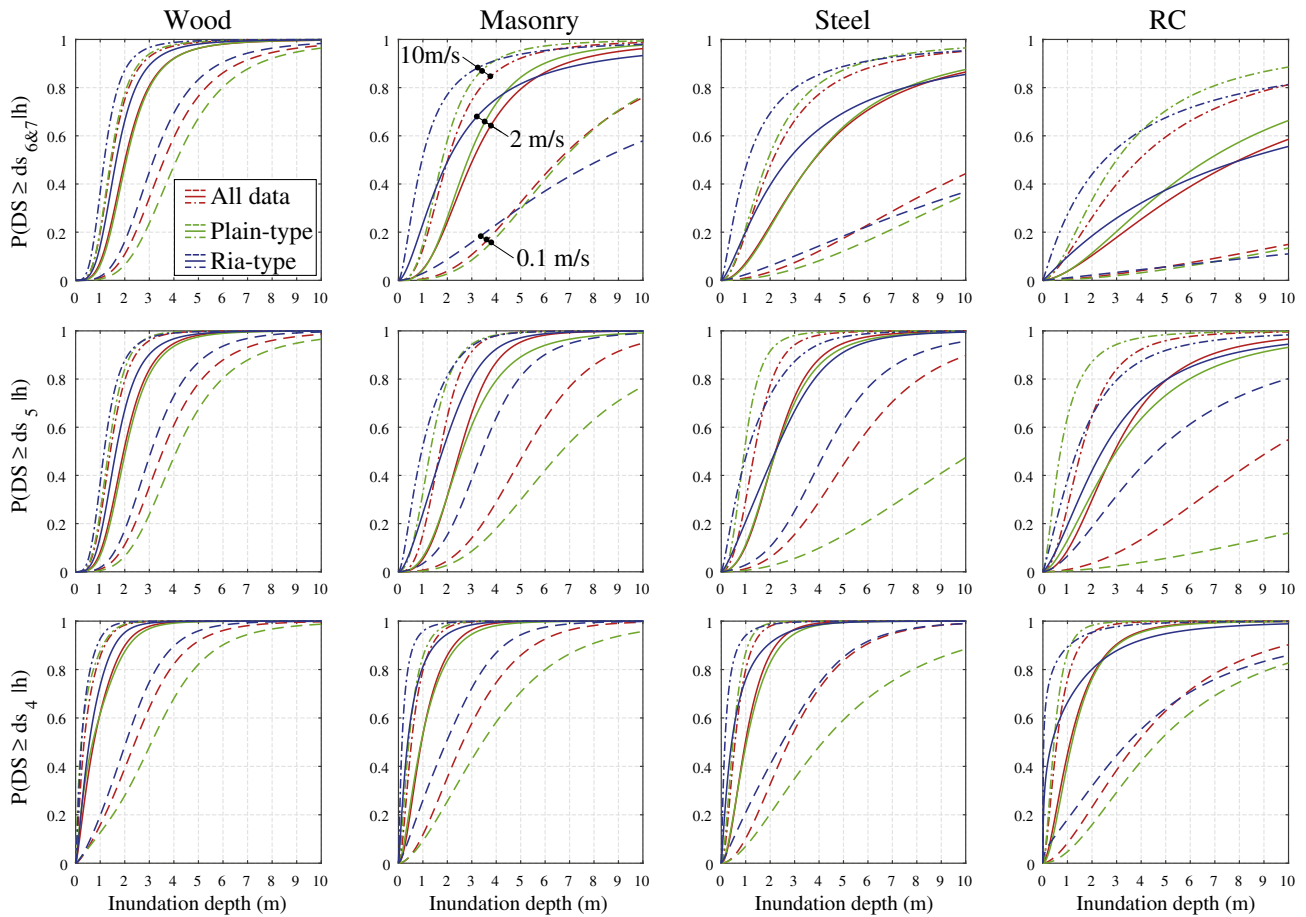
Fig. 17. Empirical fragility functions for four different structural typologies according to **M4** considering a velocity of 0.1 m/s (dashed line), 2 m/s (solid line), and 10 m/s (dashed-dotted line), and using data for the entire Miyagi prefecture (red line), data for plain-type coast only (green line), and data for ria-type coast only (blue line).

modeling. In plain-type coast, there are many locations where the consideration of flow velocity is important. In ria-type coast, buildings located very close to the shoreline (e.g. in Kesennuma) and several buildings far from the shoreline (e.g. Minamisanriku and Ishinomaki) only present residuals in favor of **M4**. Therefore, it is more appropriate to develop vulnerability models related to a specific topographic

context (i.e. ria-type or plain-type). In the following, two fragility models are developed for the two different coastal typologies.

The model selection analysis shows that, for both subsets of data, bivariate models considering inundation depth and flow velocity (i.e. **M4**) are preferable to those neglecting flow velocity. The regression parameters according to Eq. 12 are given in Tables 6 and 7, for plain-type
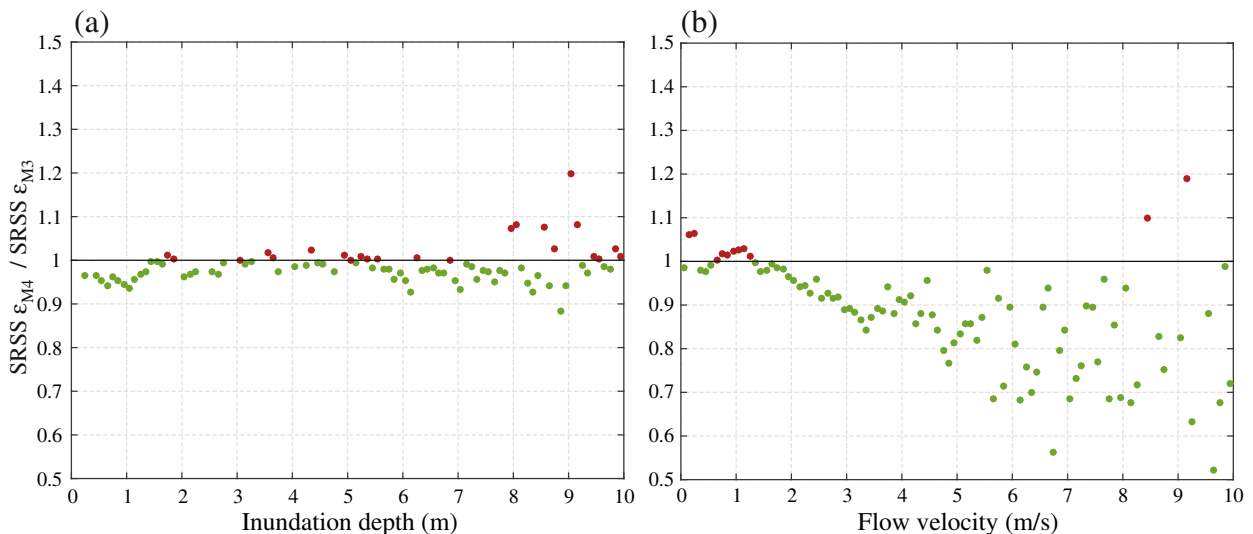

Fig. 18. Ratio between the SRSS of residuals for M4 and M3 as a function of the (a) inundation depth of (b) flow velocity for the models calibrated based on data for ria-type coast only.

and ria-type coasts, respectively. Fig. 17 presents the fragility curves for all structural typologies, considering the entire database (red lines), the plain-type coast data only (green lines), and the ria-type data only (blue lines). As in the previous fragility representations, fragility curves are represented as function of inundation depth and are given for three values of flow velocity: 0.1 m/s (dashed lines), 2 m/s (continuous lines), and 10 m/s (dashed-dotted lines). The fragility curves obtained considering the entire database lie between the curves obtained considering the two sub-sets of data separately, and are closer to the fragility curves for plain-type coast. Fragility curves related to ria-type coast present high reduction of median values (up to about 400%) and a reduction of the dispersions; the steeper fragility curves for ria-type coast can be justified by observing that for the lower values of inundation depth, the higher values of velocity lead to a faster attainment of damage. The latter observation implies a greater damage probability in ria-type coast than plain-type coast for the same value of inundation depth.

The proposed separation of coastal topography improves accuracy of vulnerability models making the velocity more crucial in vulnerability modeling. For example, Fig. 18 represents the ratio between residuals for **M4** and **M3** calibrated using ria-type observations, versus inundation depth and velocity only. Incorporating flow velocity improves the model performance for almost all the values of inundation depth and flow velocity (in comparison with Figs. 14(c) and 15(c)).

5. Summary and conclusions

The importance to consider flow velocity in multivariate empirical fragility models was investigated through rigorous statistical analysis based on the tsunami damage data available from the MLIT database for the 11th March 2011 Tohoku earthquake. The analyses focused on Miyagi prefecture, since a refined DEM (10-m resolution) was available for this region. Given the lack of actual observations of flow velocity, tsunami simulation using calibrated source models was carried out to obtain velocity data at locations of the surveyed buildings. The effects of different inundation models on vulnerability modeling were investigated by considering three source models and two DEM datasets having different spatial resolutions. Results showed that the inundation model based on the tsunami source inversion (Satake et al. model) and 10-m resolution was more reliable and accurate, with respect to recorded inundation depths data in the MLIT database, than other inundation models. Subsequently, this model was adopted as the reference.

Various fragility functions were developed through multinomial logistic regression using un-binned data by considering five nested models: **M1** considering only flow depth and structural typology; **M2** considering flow velocity; **M3** as extension of **M1** considering also the interaction terms between flow depth and structural typology; **M4** as extension of **M2** considering the interaction terms between flow depth and structural typology; and finally **M5** as extension of **M4** considering the interaction terms between flow velocity and structural typology. The diagnostic analysis indicated that **M4** performed best among the tested models; therefore, inclusion of flow velocity as additional tsunami hazard parameter improves the predictability of vulnerability models. The conclusion was applicable to other cases where different inundation models were used to derive the flow velocity data. Moreover, looking at disaggregated results, flow velocity was particularly useful for predicting the occurrence of structural damage states rather than non-structural damage states. It was also found that using coarse DEM or less refined source models leads to significant variations in developed fragility functions (because of the overestimation of the velocity data with respect to the more accurate reference case). The conclusion can be important in assessing tsunami risks (e.g. risk underestimation for low values of velocity and overestimation for high values of velocity).

The reference vulnerability model (i.e. **M4**) based on the reference scenario was further scrutinized to take into account the influence of the coastal topographic configuration via detailed residual analysis. Results showed that it was necessary to build fragility curves that reflect the regional characteristics of coastal topography. Therefore, two more refined vulnerability models were built for plain-type and ria-type coasts, respectively, in addition to the reference model that does not distinguish coastal topography.

Finally, it is worth noting that all regression parameters for all vulnerability models investigated in this work were provided in the tables of this paper or in the Supplementary electronic material. This will facilitate the implementation of the developed models in future tsunami vulnerability and risk analyses.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.earscirev.2016.12.015.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723.

Ammon, C.J., Lay, T., Kanamori, H., Cleveland, M., 2011. A rupture model of the 2011 off the Pacific coast of Tohoku Earthquake. Earth Planets Space 63 (7), 693–696.

Charvet, I., Ioannou, I., Rossetto, T., Suppasri, A., Imamura, F., 2014. Empirical fragility assessment of buildings affected by the 2011 Great East Japan tsunami using improved statistical models. Nat. Hazards 73 (2), 951–973.

Charvet, I., Suppasri, A., Kimura, H., Sugawara, D., Imamura, F., 2015. A multivariate generalized linear tsunami fragility model for Kesennuma City based on maximum flow depths, velocities and debris impact, with evaluation of predictive accuracy. Nat. Hazards:1–27 http://dx.doi.org/10.1007/s11069-015-1947-8.

De Risi, R., Goda, K., Mori, N., Yasuda, T., 2016. Bayesian tsunami fragility modeling considering input data uncertainty. Stoch. Env. Res. Risk A. http://dx.doi.org/10.1007/s00477-016-1230-x.

Dutykh, D., Poncet, R., Dias, F., 2011. The VOLNA code for the numerical modeling of tsunami waves: generation, propagation and inundation. Eur. J. Mech. B. Fluids 30 (6), 598–615.

Etienne, S., Buckley, M., Paris, R., Nandasena, A.K., Clark, K., Strotz, L., Chagué-Goff, C., Goff, J., Richmond, B., 2011. The use of boulders for characterising past tsunamis: lessons from the 2004 Indian Ocean and 2009 South Pacific tsunamis. Earth-Sci. Rev. 107 (1), 76–90.

Fritz, H.M., Phillips, D.A., Okayasu, A., Shimozono, T., Liu, H., Mohammed, F., Skanavis, V., Synolakis, C.E., Takahashi, T., 2012. The 2011 Japan tsunami current velocity measurements from survivor videos at Kesennuma Bay using LiDAR. Geophys. Res. Lett. 39 (7), 1–6, L00G23.

Ghobarah, A., Saatcioglu, M., Nistor, I., 2006. The impact of the 26 December 2004 earthquake and tsunami on structures and infrastructure. Eng. Struct. 28 (2), 312–326.

Goda, K., Mai, P.M., Yasuda, T., Mori, N., 2014. Sensitivity of tsunami wave profiles and inundation simulations to earthquake slip and fault geometry for the 2011 Tohoku earthquake. Earth Planets Space 66 (1), 1–20.

Goto, C., Ogawa, Y., Shuto, N., Imamura, F., 1997. Numerical method of tsunami simulation with the leap-frog scheme (IUGG/IOC Time project), IOC Manual, UNESCO, No. 35. United Nations Educational, Scientific and Cultural Organization, Paris, France.

Hayashi, S., Koshimura, S., 2013. The 2011 Tohoku Tsunami flow velocity estimation by the aerial video analysis and numerical modeling. J. Disaster Res. 8 (4), 561–572.

Hosmer Jr., D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied logistic regression. John Wiley & Sons, Hoboken, New Jersey, United States of America (528 p.).

Iinuma, T., Hino, R., Kido, M., Inazu, D., Osada, Y., Ito, Y., Ohzono, M., Tsushima, H., Suzuki, S., Fujimoto, H., Miura, S., 2012. Coseismic slip distribution of the 2011 off the Pacific Coast of Tohoku Earthquake (M9.0) refined by means of seafloor geodetic data. J. Geophys. Res. 117 (B7). http://dx.doi.org/10.1029/2012JB009186.

Jaffe, B., Buckley, M., Richmond, B., Strotz, L., Etienne, S., Clark, K., Watt, S., Gelfenbaum, G., Goff, J., 2011. Flow speed estimated by inverse modeling of sandy sediment deposited by the 29 September 2009 tsunami near Satitoa, east Upolu, Samoa. Earth-Sci. Rev. 107 (1), 23–37.

Kaiser, G., Scheele, L., Kortenhaus, A., Løvholt, F., Römer, H., Leschka, S., 2011. The influence of land cover roughness on the results of high resolution tsunami inundation modeling. Nat. Hazards Earth Syst. Sci. 11, 2521–2540.

Koshimura, S., Kayaba, S., 2011. Tsunami fragility inferred from the 1993 Hokkaido Nansei-oki earthquake tsunami disaster. J. Jpn. Assoc. Earthq. Eng. 10 (3), 87–101.

Koshimura, S., Oie, T., Yanagisawa, H., Imamura, F., 2009. Developing fragility functions for tsunami damage estimation using numerical model and post-tsunami data from Banda Aceh, Indonesia. Coast. Eng. J. 51, 243–273.

Kreibich, H., Piroth, K., Seifert, I., Maiwald, H., Kunert, U., Schwarz, J., Merz, B., Thieken, A.H., 2009. Is flow velocity a significant parameter in flood damage modelling? Nat. Hazards Earth Syst. Sci. 9 (5), 1679–1692.

Lipa, B., Isaacson, J., Nyden, B., Barrick, D., 2012. Tsunami arrival detection with high frequency (HF) radar. Remote Sens. 4 (5), 1448–1461.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models. CRC press, London, UK (514 p.).

Ministry of Land, Infrastructure, and Transportation (MLIT) (2014). Survey of Tsunami Damage Condition. (Accessible at: http://www.mlit.go.jp/toshi/toshi-hukkou-arkaibu.html. Last access: 1 July 2014).

Okada, Y., 1985. Surface deformation due to shear and tensile faults in a half-space. Bull. Seismol. Soc. Am. 75 (4), 1135–1154.

Porter, K., Kennedy, R., Bachman, R., 2007. Creating fragility functions for performance-based earthquake engineering. Earthquake Spectra 23 (2), 471–489.

Reese, S., Bradley, B.A., Bind, J., Smart, G., Power, W., Sturman, J., 2011. Empirical building fragilities from observed damage in the 2009 South Pacific tsunami. Earth-Sci. Rev. 107 (1), 156–173.

Satake, K., Fujii, Y., Harada, T., Namegaya, Y., 2013. Time and space distribution of coseismic slip of the 2011 Tohoku earthquake as inferred from tsunami waveform data. Bull. Seismol. Soc. Am. 103 (2B), 1473–1492.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.

Song, Y.T., Fukumori, I., Shum, C.K., Yi, Y., 2012. Merging tsunamis of the 2011 Tohoku-Oki earthquake detected over the open ocean. Geophys. Res. Lett. 39 (5), 1–6, L05606.

Suppasri, A., Koshimura, S., Imamura, F., 2011. Developing tsunami fragility curves based on the satellite remote sensing and the numericalmodeling of the 2004 Indian Ocean tsunami in Thailand. Nat. Hazards Earth Syst. Sci. 11, 173–189.

Suppasri, A., Mas, E., Charvet, I., Gunasekera, R., Imai, K., Fukutani, Y., Imamura, F., 2013. Building damage characteristics based on surveyed data and fragility curves of the 2011 Great East Japan tsunami. Nat. Hazards 66 (2), 319–341.

Tanioka, Y., Satake, K., 1996. Tsunami generation by horizontal displacement of ocean bottom. Geophys. Res. Lett. 23 (8), 861–864.

Tarbotton, C., Dall'Osso, F., Dominey-Howes, D., Goff, J., 2015. The use of empirical vulnerability functions to assess the response of buildings to tsunami impact: Comparative review and summary of best practice. Earth-Sci. Rev. 142, 120–134.