A multivariate generalized linear tsunami fragility model for Kesennuma City based on maximum flow depths, velocities and debris impact, with evaluation of predictive accuracy

I. Charvet^{1,3} · A. Suppasri² · H. Kimura² · D. Sugawara² ·
F. Imamura²

Received: 4 May 2014 / Accepted: 29 August 2015 / Published online: 15 September 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract The recent losses caused by the unprecedented 2011 Great East Japan Tsunami disaster have stimulated further research efforts, notably in the mechanisms and probabilistic determination of tsunami-induced damage, in order to provide the necessary information for future risk assessment and mitigation. The stochastic approach typically adopts fragility functions, which express the probability that a building will reach or exceed a predefined damage level usually for one, sometimes several measures of tsunami intensity. However, improvements in the derivation of fragility functions are still needed in order to yield reliable predictions of tsunami damage to buildings. In particular, extensive disaggregated databases, as well as measures of tsunami intensity beyond the commonly used tsunami flow depth should be used to potentially capture variations in the data which have not been explained by previous models. This study proposes to derive fragility functions with additional intensity measures for the city of Kesennuma, which was extensively damaged during the 2011 tsunami and for which a large and disaggregated dataset of building damage is available. In addition to the surveyed tsunami flow depth, the numerically estimated flow velocities as well as a binary indicator of debris impact are included in the model and used simultaneously to estimate building damage probabilities. Following the recently proposed methodology for fragility estimation based on generalized linear models, which overcomes the shortcomings of classic linear regression in fragility analyses, ordinal regression is applied and the reliability of the model estimates is assessed using a proposed penalized accuracy measure, more suitable than the traditional classification error rate for ordinal models. In order to assess the predictive power of the model, penalized accuracy is estimated through a repeated tenfold cross-validation scheme. For the first time, multivariate tsunami fragility functions are derived and represented in the

✉ I. Charvet
ingrid.charvet@rms.com

¹ Risk Management Solutions, 30 Monument Street, London EC3R 8NB, UK

² IRIDeS, Tohoku University, Aramaki, Aoba, Sendai 980-8579, Japan

³ EPICentre, Department of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, UK

form of fragility surfaces. The results show that the model is able to predict tsunami damage with satisfactory predictive accuracy and that debris impact is a crucial factor in the determination of building collapse probabilities.

Keywords Tsunami · Debris · Fragility functions · Accuracy · Cross-validation

1 Introduction

The widespread destruction caused by the 2011 Great East Japan tsunami has demonstrated that even state-of-the-art coastal protection measures and disaster mitigation strategies, in countries such as Japan, cannot prevent large tsunamis from causing severe large-scale building damage, thus significant economical losses and loss of life. In order to provide better tools to decision makers, engineers and the risk assessment community in exposed countries, it is necessary to further our understanding of structural failure mechanisms under tsunami loads and to improve the predictive ability of current models to estimate future potential tsunami damage. While from a deterministic standpoint some noticeable advancements are being made on the identification of wave inundation processes and actions of tsunami forces, from both large-scale hydraulic experiments (Arikawa 2009; Arnasson et al. 2009; Charvet et al. 2013a; Lloyd 2014), field surveys (Chock et al. 2013; EEFIT 2013), as well as simulation with particle–fluid mixture flows (Pudasaini 2014), from a probabilistic perspective the likelihood of future tsunami-induced building damage, or *fragility estimation*, still requires considerable improvements in order to reduce the uncertainty associated with the predictions (Suppasri et al. 2013a; Charvet et al. 2014a, b). Fragility estimations are obtained through the derivation of *fragility functions*, which are empirical stochastic functions giving the probability that a building will reach or exceed a given level of damage for a measure of tsunami intensity (IM). Damage levels are typically defined prior to a post-tsunami survey by engineering teams and describe the condition of the affected structure, from zero damage to complete failure, thus forming a damage scale. Such scale is then used in combination with tsunami flow depth measurements (Ruangrassamee et al. 2006; Mas et al. 2012; Suppasri et al. 2012a, 2013a, b) or results from numerical simulations (Koshimura et al. 2009; Suppasri et al. 2011, 2012b) in order to classify the surveyed buildings according to their damage state and a corresponding IM. The first column of Table 1 presents the damage scale defined by the Ministry of Land, Infrastructure, Transport and Tourism in Japan (MLIT) for the survey of the buildings affected by the Great East Japan tsunami that struck the country on March, 11, 2011.

Fragility functions are derived by applying regression analysis techniques to the classified observations, using damage state as response variable and the chosen IM (most commonly the tsunami flow depth—e.g., Suppasri et al. 2013a, b; in some cases numerically estimated flow velocities, or analytically estimated hydrodynamic forces—e.g., Suppasri et al. 2012b) as the explanatory variable. Typically, fragility functions are derived by using linear least squares regression, assuming that the response to be modeled follows a normal or lognormal distribution, and by grouping or re-grouping the data into bins of tsunami intensity (e.g., Suppasri et al. 2011, 2012a). Nevertheless, it was shown by Charvet et al. (2013b) and Charvet et al. (2014a, b) that considerable uncertainty was introduced in the damage predictions when such methods are used as a tool to analyze building damage data. The results from Charvet et al. (2014a) also indicate that highly aggregated databases

Table 1 Tsunami damage scale defined by the ministry of land, infrastructure and transport and tourism in Japan

MLIT DS	This study DS	Classification of damage	Description of observed damage	Condition of the structure
0	<i>DS0</i>	No damage	No damage	Promptly reusable without repair or floor cleaning
1	<i>DS1</i>	Minor	Flood only	Promptly reusable after floor cleaning
2	<i>DS2</i>	Moderate	No structural damage	Reusable after minor repair
3	<i>DS3</i>	Major	Partly damaged walls, but no damage in columns	May be reusable after great repair
4	<i>DS4</i>	Complete	Extensive damage in walls, some damage in columns.	May be reusable after great repair only in some cases
5	<i>DS5</i>	Collapse	More than half of walls were destroyed, most columns were bent or destroyed	Non-repairable
6		Washed away	Washed away, only foundation remained, overturned	

(i.e., when observations are grouped into a range of tsunami IM, and/or grouped over a wide range of geographical locations) result in a loss of information, potentially yielding variations which cannot be explained by the model. The building damage analysis following the 2009 Samoa tsunami by Reese et al. (2011) was the first study in the tsunami engineering field which derived fragility functions using a more adequate statistical approach, namely logistic regression, assuming the damage response to follow a binomial distribution (i.e., damaged/not damaged). However, logistic regression does not exploit fully the information available in the database, as it ignores the fact that damage is an ordinal categorical variable.

In order to address such issues, Charvet et al. (2014b) used ordinal regression analysis using the extensive disaggregated database (56950 individual buildings) from Ishinomaki city in Japan, one of the most severely affected cities during the 2011 Great East Japan tsunami. Ordinal regression is a statistical method based on the use of Generalized Linear Models, or GLM (Mc Cullagh and Nelder 1989). GLM relax many of the assumptions associated with linear least squares regression, thus addressing a number of issues associated with this method (more details can be found in Charvet et al. 2014a, b). In addition, they allow the response variable (i.e., damage state) to follow a number of statistical distributions, including discrete distributions (more adequate to describe the discrete nature of the damage states). However, despite the use of a detailed database and a more elaborate statistical estimation method, Charvet et al. (2014b) found that the model still did not capture all the systematic variability in the data, particularly for high damage states. In addition, for the strongest construction types such as RC (Reinforced Concrete) and Steel buildings, the data were heavily scattered, showing that the chosen IM alone (i.e., flow depth) is a poor predictor of damage for these structures. Therefore, it has become more and more apparent that at least a large proportion of the uncertainty associated with fragility estimations could be reduced by incorporating in the model explanatory variables beyond the commonly used tsunami flow depth. Indeed, there is much visual evidence to support the important contribution of variables such as debris impact and scour to severe

damage states, as well as the impact of high flow velocities which can be inferred from the occurrence of building overturn or from the residual deformation of columns (Fig. 1). A number of analytical expressions for tsunami forces exist and are consistent with the ones shown in Fig. 1—taken from FEMA (2008) and Chock et al. (2011). Nevertheless, such dynamical quantities (flow depth and velocities) can better be obtained by applying a real two-phase mass flow model and unified simulation techniques as proposed in Pudasaini (2012, 2014). In addition, variables such as the Froude number (for example, as used in the scour depth estimation in Fig. 1d) are crucial in designing coastal structures and estimating their strength. Therefore, they should be appropriately modeled. When the flow hits the obstacle, the flow is rapidly and largely sheared in the flow depth direction, then turns from shallow to full dimensional. In this situation, a more advanced generalized Froude number (Dominik and Pudasaini 2012) should be used that varies through the flow depth and takes

Fig. 1 Characteristic building damage triggered by the 2011 tsunami in Japan, with analytical expression of the tsunami force component(s) identified as the main cause of damage. **a** Building which has been overturned, due to the combined actions of hydrostatic (lateral and buoyancy forces) and hydrodynamic loads. Such forces are mainly determined by the tsunami flow depth h and the flow velocity u (other parameters are typically taken as constants—FEMA 2008). **b** Residual deformation of columns due to sustained lateral loads (EEFIT 2013), namely a combination of hydrostatic and hydrodynamic forces (respectively, F_h and F_d). These two force components always act on the building concurrently, as indicated by the double arrow (a)–(b). Other damaging mechanisms are also at play, such as debris impact (c), and scour (d). In this figure, h is the tsunami flow depth, u is the horizontal flow velocity, ρ_s is the fluid density (accounting for sediment density), g is the acceleration of gravity, b is the breadth of the structure, C_d is the drag coefficient, C_m is the virtual mass coefficient, and u_d is the velocity of the debris. u_d must be the relative velocity between the solid and the fluid with the virtual mass coefficient (Pudasaini 2012), k is the effective contact stiffness of the debris object and the impacted system as described in Chock et al. (2011), m is the mass of the debris, z_m is the scour depth, Fr is the Froude number (Hughes 1993), and K_a the approach angle

into account several important dynamical aspects of the flow including the gravity and pressure potential energies and the kinetic energy of the flow (Pudasaini and Dominik 2009). Moreover, the Froude number should also include the phase-Froude numbers for solid and fluid as proposed in Pudasaini (2012), which takes into account interactions between the solid and fluid components.

While flow depth has been found to be a good predictor of tsunami damage for low damage states (Charvet et al. 2014a, b), it is clear that the contribution of such variables will be increasingly important in explaining higher damage states and should be included in the model in order to yield better results (Fig. 2).

The choice of flow depth as a unique demand parameter, or explanatory variable, in the construction of fragility curves stems from the fact this measure can be retrieved directly from field surveys (for example, by measuring the height of the water mark on an affected building). Flow velocities, on the other hand, can be used as an IM but require detailed numerical simulations (Mader 2004). Recent advances in phase flow modeling also demonstrate flow height and velocity could better be obtained by applying sophisticated two-phase models such as those proposed by Pudasaini (2012, 2014). Scour is a complex process which depends on factors such as flow approach angle, number of waves, soil conditions (Chock et al. 2011), which are typically not reported or straightforward to identify. For obvious reasons, debris are always removed by the authorities as quickly as possible after a tsunami attack, thus information on their contribution to damage is very difficult to quantify from post-tsunami surveys.

Challenges associated with the inclusion of several variables include the representation of fragility functions in a different form (e.g., more than two dimensions), and a meaningful quantitative assessment of the model's reliability. In the case of ordinal regression, the latter can be assessed qualitatively through a graphical representation of the expected (model) versus observed (data) probabilities or counts (Charvet et al. 2014a, b). Nevertheless, this

Fig. 2 Schematic representation of the contributions of different tsunami intensity measures to the severity of observed damage. In general, the number of influential tsunami IMs increases as well as the contribution of each IM along with the damage level. More specifically, the contribution of flow depth will be more prominent for lower damage states (i.e., DS1: “flooding”), while flow velocity and debris impact will be increasingly influential from non-structural to the heaviest structural damage, and scour corresponding mainly to heavy structural damage states

graphical tool becomes insufficient as the number of variables increases—indeed, in order for such diagnostics plot to be meaningful each point has to represent a unique combination of values for all explanatory variables. This leads to very low counts of instances (e.g., less than 10) when an increasing number of explanatory variables is included. Furthermore, it would be desirable to provide a quantitative and straightforward measure of predictive ability for ordinal models, instead of a quantitative assessment.

In the light of these findings and considerations, this study aims to perform the building fragility analysis for the city of Kesennuma in Japan, heavily affected by the 2011 tsunami, for which a fully disaggregated damage database, surveyed flow depth measurements and velocity simulation results are available. Due to its adequacy in representing the damage response scale, the statistical method used is ordinal regression, although the adequacy of a more complex, partially ordered model will also be tested. One main objective of this work is to include not only flow depth and the simulated velocities as explanatory variables, but also an indicator of debris impact, based on the geographical location of sources of harmful debris (i.e., washed away structures). The model is built step by step, by considering notably the relative influence of the variables to be included on the response. The predictive power of the model is assessed using a proposed penalized accuracy measure (accounting for the magnitude of the error), and a repeated tenfold cross-validation scheme. Finally, the multivariate fragility functions are represented in the form of fragility surfaces, allowing for an assessment of the influence of debris impact on high damage probabilities and a comparison of the relative fragility of different building classes.

2 Presentation of the data

2.1 Building damage and tsunami flow depth

The building damage dataset was provided by the city of Kesennuma, it contains 19815 individual buildings, and the available information includes the tsunami flow depth retrieved from the post-disaster survey (i.e., maximum tsunami flow depth), for 14760 buildings the main construction material (RC, Steel, Wood or Masonry), and for 9415 buildings the number of stories. Detailed construction practices for each of these main construction classes for Japan can be found in Architectural Institute of Japan (2012). The damage scale that was used to classify the observation is the scale presented in Table 1. The examination of the damage scale designed by the MLIT reveals that DS5 and DS6 do not represent mutually exclusive damage states, nor do they necessarily represent an increase in tsunami intensity. Rather, they represent different failure modes of the structure. In order to apply GLM analysis, Mc Cullagh and Nelder (1989) advise for such requirements to be met; therefore, in this study these two levels will be aggregated (i.e., DS5 will correspond to the buildings that have collapsed, and the ones that have been washed away). This transforms the given seven-state (DS0 to DS6) into a six-state damage scale (DS0 to DS5).

2.2 Flow velocities

The flow velocity estimations for Kesennuma are obtained from a nonlinear shallow water model, namely the TUNAMI model. A set of nonlinear shallow water equations is discretized using the Staggered Leap-frog finite difference scheme (Imamura 1996). Bottom

friction for urban areas—where the concentration of buildings is significant—can be represented using high Manning’s roughness coefficients values (i.e., 0.04 or higher). In reality, when a building is washed away (a frequent occurrence in this particular event), the friction is reduced dramatically as the site becomes effectively concrete or bare ground. This increases the actual flow velocity over that area, effect which would be missed in the modeling. Therefore, in the absence of land cover models for Kesennuma, a constant Manning’s roughness coefficient of 0.025 was chosen for the present simulation, an avenue for improvement would be to consider friction as a dynamically varying parameter conditional on the washing away of structures (or other significant changes in land cover due to the tsunami flow). The requires the availability of detailed land use land cover data, concurrently with the determination of a washed away state conditional on a set of depth-velocity values.

Five computational domains are used as a nested grid system: 405 m (Region 1), 135 m (Region 2), 45 m (Region 3), 15 m (Region 4) and 5 m (Region 5), as shown in Fig. 3. The bathymetry data for Region 1 were derived from a combination of the original data from the Japan Coast Guard and the Cabinet Office, Government of Japan, Regions 2 to 5 are from the Cabinet Office, Government of Japan. The topography data for Regions 1 to 3 were derived from the Cabinet Office, Government of Japan and from the Geospatial Information Authority of Japan for Regions 4 to 5. All bathymetry and topography data used in this study were collected before the occurrence of the 2011 tsunami. In order to carry out fragility estimations and limit the uncertainty associated with the velocity estimations from coarse grid sizes, this study only uses the buildings located in Region 5 (10812 buildings), for which a 5-m grid size has been used.

Fig. 3 Nested grid system used for the tsunami numerical modeling in the region of Kesennuma

The flow velocity values retrieved from the simulation are maximum estimated flow velocities. Suppasri et al. (2014) used a similar simulation setup (source deformation, wave propagation model), with the exception of a coarser resolution of 135 m for the region of interest (region 5). They found a reasonable agreement between the modeled tsunami flow velocity ranges, and local velocities measured from available survivor videos in Sendai plain (Hayashi and Koshimura 2013) and Kesennuma City (Fritz et al. 2012). However, better simulation of the tsunami velocities can be obtained by utilizing a real two-phase tsunami run-up model and unified simulations (Pudasaini 2014) as the flow is mainly two-phased in nature (processes such as runup and impact on structures).

While it is worth noting that the time of occurrence of maximum depth and maximum flow velocity is not necessarily the same at a given location, it is expected that the final observed damage is a result of successive failures triggered by the overall maximum demands.

2.3 Debris impact

A number of tsunami disasters have shown that debris impact can severely damage a structure. Following the 2006 Java tsunami survey, Reese et al. (2007) found robust buildings could act as a shield for weaker buildings, which would have typically suffered heavy damage from impacting debris. In the case of Kesennuma, it was not possible to

Fig. 4 Photographic evidence from the immediate tsunami aftermath in Kesennuma, showing the nature of debris generated by the tsunami (a lot of remains from washed away houses)

Fig. 5 Principle of debris impact variable assignment in Kesennuma. If a building is washed away, it is assumed the generated debris can travel a distance of 50 m from this building (for instance), then all buildings located within 50 m of this washed away structure are assigned a value of '1' for debris impact (the other ones are assigned a value of '0')

directly quantify debris impact or the effects of sheltering from post-tsunami surveys, not only because of early debris removal by the authorities, but also because in the case of collapsed or washed away structures clear evidence of debris impact was not available. However, it can be seen (Fig. 4) that the bulk of debris material generated by the 2011 event seems to be composed mainly of the remains of buildings which have been washed away. Thus, in this study we assume that impact will occur if the building under consideration is in the surroundings of a structure which has been washed away (as defined by the damage scale in Table 1) and define such impact as a binary occurrence, i.e., either a building is impacted (*debris impact* = 1) or not (*debris impact* = 0). Because the distance travelled by the projectile from the point of debris generation (i.e., initial location of the washed away structure) to the affected building is unknown, a range of debris action radiuses (from 10 m to 150 m) will be tested (Fig. 5), using cross-validation and the accuracy measure presented in the following section.

3 Fragility estimation methodology

3.1 Ordinal regression model

In this study, fragility functions are estimated using GLM, more specifically ordinal regression. The generalized linear models methodology implies firstly, the selection of an appropriate statistical distribution for the response variable (i.e., the damage state), which will be referred to as the *random component* of the model; secondly, the expression of the damage probability exceedance (or fragility function μ) through a chosen linear predictor η

and link function g (this will be referred to as the *systematic component*). In ordinal regression analysis, the random component chosen is the multinomial distribution (Forbes et al. 2011), a multicategorical discrete distribution, expressed as:

$$Y_{k,i} \sim \prod_{k=0}^K \frac{N_i!}{Y_{k,i}!} P(DS = ds_k | x_i)^{Y_{k,i}}. \quad (1)$$

In Eq. (1), $Y_{k,i}$ corresponds to the counts of buildings being at damage level ds_k ($k \in \mathbb{N}; 0 \leq k \leq K$) for each value of the tsunami intensity measure x_i ; and N_i is the total number of buildings. This probability can be transformed into the probability of reaching or exceeding ds_k given x_i , $P(DS \geq ds_k | x_i)$, essentially expressing the required fragility function, as:

$$P(DS = ds_k | x_i) = \begin{cases} 1 - P(DS \geq ds_k | x_i) & k = 0 \\ P(DS \geq ds_k | x_i) - P(DS \geq ds_{k+1} | x_i) & 0 < k < K \\ P(DS \geq ds_k | x_i) & k = K \end{cases} \quad (2)$$

When using GLM, the linear predictor is an additive function of all explanatory variables included. Posing $P(DS \geq ds_k) = \mu_k$, the linear predictor defines the systematic component of the model as follows:

$$(\eta_k | 1 \leq k \leq K) = g(\mu_k) = \theta_{0,k} + \sum_{j=1}^J \theta_{k,j} X_j. \quad (3)$$

In Eq. (3), X_j are the J explanatory variables that can be used for the regression analysis, and $\{\theta_{0,k}, \dots, \theta_{k,j}\}$ are the parameters of the model to be estimated. In the case of Kesennuma city, we have $J = 5$ explanatory variables: the surveyed tsunami flow depth, the simulated flow velocities, the debris impact (dummy coded as 0 or 1), the building class (Wood, Masonry, RC, or Steel, also dummy coded), and the number of stories (1 to 6, treated as a continuous variable). $e^{\theta_{k,j} \neq 0}$ represents the “odds ratio”. For example, assuming the usual logit function $\left(\log\left(\frac{p}{1-p}\right)\right)$ as link for the model, for a given damage class k and explanatory variable X_j the odds that the damage state of a building is above level k are multiplied by $e^{\theta_{k,j} \neq 0}$ for each unit increase in X_j .

Ordinal regression assumes that the response categories are ordered, this is taken into account by fixing the free parameters in the systematic component (3) across all damage states, allowing only the constant $\theta_{0,k}$ to vary, thus (3) becomes:

$$(\eta_k | 1 \leq k \leq K) = g(\mu_k) = \theta_{0,k} + \sum_{j=1}^J \theta_j X_j \quad (4)$$

The parameters in (4), the ordered model, are estimated using maximum likelihood estimation or MLE (Mc Cullagh and Nelder 1989; Myung 2003). MLE consists in retrieving the set of parameters which maximizes the probability of the observed data. Note that the ordered model now assumes that the rate of increase in the odds for a unit increase in X_j is constant across the classes (i.e., $\theta_{k,j} \neq 0$ becomes $\theta_j \neq 0$). This parameter estimation method assumes that observed values of the response and explanatory variables were measured without error (i.e., there is no difference between the estimated and true values of the model parameters). In reality, the damage classification of a building (response) is by nature subjective, and the measurement of explanatory variables such as flow depth (in the field) or velocity (numerical modeling approximations) is also subject to uncertainty. In

order to address such uncertainty, Bayesian estimation methods can be used, as they allow for measurement errors to be incorporated via appropriate statistical distributions. However, due to the lack of prior information on measurement errors and resulting parameter uncertainty, the use of these more complex methods is unnecessary. Indeed, the use of uninformative priors in Bayesian parameter estimation converges to the MLE result.

Generally, for regression based on binary or multinomial outcomes, the appropriate link functions g are the logit, probit, and in some cases the complementary loglog functions, as described in Rossetto et al. (2014). Note that the complementary loglog link is not considered here, as this link provides estimations close to the logit link except for a heavier right tail; and cannot be differentiated when the linear predictor takes the value of zero.

3.2 Diagnostics

3.2.1 Relative goodness-of-fit: deviance and AIC

Measures of relative goodness-of-fit consist in comparing the performance of several possible models, based on the same statistical distribution, but with a different number of parameters or different link functions. When the two models compared use the same link function but have a different number of parameters (i.e., *nested* models), the likelihood ratio test is appropriate (Fahrmeir and Tutz 2001). It reveals whether the model which has more parameters gives a significantly better fit to the data than its simpler alternative. Thus it is useful to use the likelihood ratio test in order to determine:

- Which IMs have a significant influence on the response. The IMs are added step by step, after systematically checking their usefulness as explanatory variables.
- If interaction terms should be included in the model. When at least 2 IMs are included, there is a possibility that the influence of one IM on the response depends on the value of the other. The interaction term for each combination of IMs is a multiplicative function of those IMs.
- If the assumption of an ordered response does not hold, i.e., if the multinomial model [Eq. (3)], which has more parameters, provides a better fit to the data than its simple, ordinal alternative.

When two models have the same number of parameters and are based on the same statistical distribution of the response, thus differ only by their link function, the Akaike information criterion, or AIC (Akaike 1974) should be used:

$$\text{AIC} = 2p - 2\log(L), \quad (5)$$

In (5), p is the number of parameters in the model, $-2\log(L)$ is the model's deviance, a measure of error and a function of the maximum likelihood L . The best fitting model is the model with the smallest AIC.

3.2.2 Absolute goodness-of-fit: penalized accuracy and cross-validation

Contrary to models fitted using least squares regression and applied to continuous responses which typically use the coefficient of determination (R^2) to assess goodness-of-fit; a common measure of absolute goodness-of-fit of GLM applied to discrete outcomes is the model accuracy (Baldi et al. 2000; Leelawat et al. 2014). Indeed, if some “equivalent” R^2 values can theoretically be calculated for generalized linear models, their meaning is significantly

different from the classic coefficient of determination used in linear least squares regression (Hu et al. 2006) and the use of such measure in this context is often discouraged (Zheng and Agresti 2000). Accuracy is typically calculated by deducing the classification error rate (i.e., the number of incorrectly classified outcomes) from a perfect model accuracy of 1, or directly (and equivalently) defined as the proportion of well-classified outcomes. In this case, the accuracy acc_i of the model for each individual data point is:

$$acc_i = \begin{cases} 1 & DS_i = \widehat{DS}_i \\ 0 & DS_i \neq \widehat{DS}_i \end{cases} \quad (6)$$

In (6), DS_i is the actual damage state and $\widehat{DS}_i$ the damage state estimated from the model output for the i th observation (or expected damage). The overall classifier's accuracy Acc is typically defined as a percentage, representing the ratio of correctly classified outcomes to the total sample size n :

$$Acc = \frac{1}{n} \sum_{i=1}^n acc_i \quad (7)$$

This measure has the advantage to be easy to calculate and interpret, particularly for binomial models. However, a major shortcoming of accuracy for ordinal models is the fact it does not take into account the ordered nature of the response categories. For example, if a building has been classified as DS4 (field observation) and the model misclassifies this observation as DS3, the error is not as large as if the model had misclassified it as DS1. Unfortunately, typical accuracy calculations only count the number of correctly classified outcomes, so information on the discrepancy between the observed and predicted measures for unsuccessful model estimations is not exploited. It is therefore necessary to include a correction term in the accuracy calculation to obtain more meaningful estimates. The proposed amendment to this approach is to take into account the distance between the observed and predicted classes for each instance in (6). Assuming the classes are equidistant, we adopt a similar approach to that proposed by Rennie and Srebro (2005), by incorporating an immediate-threshold loss function which penalizes the accuracy estimation proportionally to the distance between the observed and predicted outcomes:

$$acc_i = 1 - \frac{|DS_i - \widehat{DS}_i|}{K - 1} \quad (8)$$

The penalized accuracy estimations (8) are then substituted in (7) to obtain the overall penalized accuracy of the ordinal model. Note that because ordinal regression analysis estimates the response probabilities directly, the expected damage state is calculated a posteriori by assuming that $\widehat{DS}_i$ is reached or exceeded from a chosen probability threshold (here this threshold is set at 0.5). Thus the accuracy results may vary slightly depending on the threshold chosen.

Finally, in order to avoid the issue of model overfitting (i.e., over-confidence in the results) and give a realistic estimation of future model performance, it is necessary to test the ability of the model to classify correctly unseen data. This can be done by using a repeated cross-validation algorithm, which, in comparison with other estimators such as repeated hold-out and bootstrap, was shown to be an adequate estimator of accuracy (Kohavi 1995) or related classification error rates (Kim 2009). The penalized accuracy estimations presented in the following sections correspond to the cross-validated estimates, derived using repeated tenfold cross-validation.

4 Results and discussion

4.1 Model selection

Following the methodology described in 3.2.1, explanatory variables are added one by one to the ordered systematic component of the GLM [Eq. (4)], as well as possible interaction terms between these variables. At each step, a likelihood ratio test is used to determine whether the added term results in a significantly better fit of the most complex model compared with the simple model. The same test is used to compare the ordered model with an equivalent but more complex, multinomial model [Eq. (3)], which relaxes the assumption that the response categories (i.e., damage states) are ordered.

Two different link functions will be considered: the probit and logit link functions. Because models differing by their link functions are non-nested models, we use the AIC [Eq. (5)] in order to determine which link function provides a better fit to the data.

In this stage of the analysis, the binary debris impact variable has been defined for a radius of projectile action of 50 m. Indeed, this distance is thought to be significant enough for the linear predictor to also contain meaningful variations in measured flow depths and simulated flow velocities, while not being too large, which could increase the uncertainty of whether a building has in reality been impacted or not. Each possible radius is considered individually in the next stage (Sect. 4.2). The variable ‘number of stories’ was treated as a continuous variable.

The results from Tables 2 and 3 indicate that, for both link functions the more complex multinomial model generally does not yield a better fit compared to the ordered model. Thus ordinal regression is an adequate choice for the analysis. A comparison between the AIC values of the best fitting probit model and the best fitting logit model (i.e., 5862 and 5766, respectively) indicates that the logit function leads to a slightly better fit.

According to the results in Table 3, the explanatory variables to be included in the model [i.e., Eq. (4)] are: the available tsunami IMs (flow depth, flow velocity, debris impact), the building’s construction material, the number of stories; and the interaction terms to be included are interactions between flow depth and velocity (h.v), flow depth and debris impact (h.I), flow velocity and debris impact (v.I), flow depth and building material (h.M), flow velocity and building material (v.M), material and debris impact (M.I), number of stories and flow depth (S.h), number of stories and debris impact (S.I), as well as number of stories and building material (S.M). These findings are consistent with the physical processes at play during tsunami flow-structure interaction. Indeed, although the extent to which maximum flow velocity interacts with maximum flow depth is unknown, it is expected an increase in maximum flow velocity would have an effect on the rate of increase in damage probability which depends on the value of flow depth. Similarly, it is expected that the effects of hydrostatic and hydrodynamic loads (i.e., flow depth and velocity) on increasing damage probabilities will depend on whether the building is also suffering debris impact loads. The same reasoning can be applied for other interaction terms.

Finally, it is interesting to note for both link functions, and aside of significance testing, the largest reductions in model deviance (i.e., more than 1000) result from the addition of building material and number of stories to the initial flow depth as explanatory variables. This result is consistent with previous findings (Suppasri et al. 2013a, b; Leelawat et al. 2014) and indicates that aside from tsunami IMs, the construction characteristics of a building strongly determine the severity of damage. Comparatively, the difference in

Table 2 Deviance and AIC values of the ordered and multinomial probit models. The values in bold represent minimum deviance and AIC values (i.e. the best fitting model)

Link: PROBIT $g(\mu) = \Phi^{-1}(\mu)$	Ordered		Multinomial	
	Deviance	AIC	Deviance	AIC
Flow depth	11,423	11,435	12,772	12,782
Flow depth, velocity	11,417	11,431	12,690	12,705
Flow depth, debris impact (50 m)	10,460 ^a	10,474	10,969	10,989
Flow depth, debris impact (50 m), interaction (h.I)	10,447 ^a	10,463	10,206 ^b	10,226
Flow depth, debris impact (50 m), interaction (h.I), material	8253 ^a	8275	10,126	10,161
Flow depth, debris impact (50 m), interaction (h.I), material, interaction (h.M)	7414 ^a	7442	12,356	12,406
Flow depth, debris impact (50 m), interaction (h.I), material, interaction (h.M), interaction (I.M)	7402	7436	13,334	13,399
Flow depth, debris impact (50 m), interaction (h.I), material, interaction (h.M), number of stories	5911 ^a	5941	10,091	10,146
Flow depth, debris impact (50 m), interaction (h.I), material, interaction (h.M), number of stories, interaction (S.h)	5908	5940	9764	9824
Flow depth, debris impact (50 m), interaction (h.I), material, interaction (h.M), number of stories, interaction (S.I)	5868 ^a	5900	9997	10,057
Flow depth, debris impact (50 m), interaction (h.I), material, interaction (h.M), number of stories, interaction (S.I), interaction (S.M)	5824^a	5862	9,476	9,551

Φ represents the standard normal distribution function in the expression of the probit link function. In the interaction terms, *h* refers to flow depth, *I* to debris impact, *M* to building material, and *S* to number of stories

^a The result of the likelihood ratio test between two ordered models with a different number of explanatory variables is significant ($p < 0.05$)

^b The result of the likelihood ratio test between the ordered and multinomial models is significant ($p < 0.05$)

deviance associated with the inclusion of debris impact and flow velocity is smaller. This can be explained by the uncertainty associated with the determination of both variables, as flow velocity inland was calculated with a constant roughness coefficient and using non-linear shallow water equations, which may not be perfectly accurate onshore (Charvet 2012) and the debris impact variable is relatively coarse, based on the assumption that debris impact is due to washed away buildings rather than actual individual observations. Nevertheless, these variables still yield significant model improvement (which is consistent with field survey evidence, showing the unquestionable contribution of these IMs to tsunami-induced damage—see Sect. 1, as well as Figs. 1 and 4). It is expected that more dramatic reductions in deviance and AIC will result if refined estimations for these two IMs are obtained in the future.

4.2 Penalized accuracy and debris impact radius

In the aftermath of a tsunami attack, and due to the different responses of structures to tsunami loads depending on the building's construction type, fragility is often represented separately for different building classes. Thus, following the method presented in 3.2.2, the overall penalized accuracy of the model is calculated for each building class and for each

Table 3 Deviance and AIC values of the ordered and multinomial logit models. The values in bold represent minimum deviance and AIC values (i.e. the best fitting model)

Link: LOGIT $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$	Ordered		Multinomial	
	Deviance	AIC	Deviance	AIC
Flow depth	10,961	10,973	12,597	12,617
Flow depth, velocity	10,933 ^a	10,947	12,515	12,545
Flow depth, velocity, interaction (h.v)	10,804 ^a	10,820	12,453	12,493
Flow depth, velocity, interaction (h.v), debris impact (50 m)	10,157 ^a	10,175	12,916	12,966
Flow depth, velocity, interaction (h.v), debris Impact (50 m), interaction (h.I)	10,136 ^a	10,156	12,900	12,960
Flow depth, velocity, interaction (h.v), debris impact (50 m), interaction (h.I), interaction (v.I)	10,132 ^a ($p = 0.044$)	10,154	12,898	12,968
Flow depth, velocity, interaction (h.v), debris Impact (50 m), interaction (h.I), interaction (v.I), material	7460 ^a	7488	8875	8975
Flow depth, velocity, interaction (h.v), debris Impact (50 m), interaction (h.I), interaction (v.I), material, interaction (h.M)	7276 ^a	7310	12,750	12,880
Flow depth, velocity, interaction (h.v), debris Impact (50 m), interaction (h.I), interaction (v.I), material, interaction (h.M), interaction (v.M)	7269 ^a ($p = 0.048$)	7309	12,223	12,383
Flow depth, velocity, interaction (h.v), debris Impact (50 m), interaction (h.I), interaction (v.I), material, interaction (h.M), interaction (v.M), interaction (M.I)	7180 ^a	7226	12,153	12,343
Flow depth, velocity, interaction (h.v), debris Impact (50 m), interaction (h.I), interaction (v.I), material, interaction (h.M), interaction (v.M), interaction (M.I), number of stories	5751 ^a	5799	10,081	10,281
Flow depth, velocity, interaction (h.v), debris Impact (50 m), interaction (h.I), interaction (v.I), material, interaction (h.M), interaction (v.M), interaction (M.I), number of stories, interaction (S.h)	5705 ^a	5755	9742	9952
Flow depth, velocity, interaction (h.v), debris Impact (50 m), interaction (h.I), interaction (v.I), material, interaction (h.M), interaction (v.M), interaction (M.I), number of stories, interaction (S.h), interaction (S.v)	5705	5757	9727	9917
Flow depth, velocity, interaction (h.v), debris impact (50 m), interaction (h.I), interaction (v.I), material, interaction (h.M), interaction (v.M), interaction (M.I), number of stories, interaction (S.h), interaction (S.I)	5681 ^a	5733	9932	10,122
Flow depth, velocity, interaction (h.v), debris impact (50 m), interaction (h.I), interaction (v.I), material, interaction (h.M), interaction (v.M), interaction (M.I), number of stories, interaction (S.h), interaction (S.I), interaction (S.M)	5656^a	5766	9411	9631

^a The result of the likelihood ratio test between two ordered models with a different number of explanatory variables is significant ($p < 0.05$). When results are close to non-significance, the p values are provided. In the interaction terms, h refers to flow depth, v to flow velocity, I to debris impact, M to building Material, and S to number of stories

debris impact radius (10 m–150 m). The required number of iterations of the repeated cross-validation scheme is determined through the examination of the results' stability, for a debris impact radius of 50 m. We consider the model is stable when the variability of the

penalized accuracy results is in the order of 0.01 % (so no additional iterations are required). Table 4 presents the overall penalized accuracy for each building class and debris impact radius, as well as the number of iterations used. Note that for Steel buildings, the number of stories variable and associated interactions could not be included in the model. Indeed, this information was only available for 12 points in the Steel database, which is not sufficient for the present regression analysis (Green 1991).

The results in Table 4 show that the model yields a high predictive accuracy (>90 %) for all building classes. This result partly reflects the commonality of heavy damage states (Fig. 6), particularly for Wood and Masonry buildings, for which 85 % and 65 % of the data, respectively, lie in the “collapse” category, making the overall possible loss trivial for potentially misclassified data. However, the model performs reasonably well also with regard to accuracy for Steel and RC structures, for which the damage distributions, while skewed to the left, are more even. It is therefore considered from such results that the predictors and model chosen are adequate to predict building damage probabilities for this event.

The highest accuracy for RC buildings is obtained for a debris impact radius of 90 m, for Steel buildings 30 m, for Wood buildings 130 m to 140 m, and for Masonry buildings 120 m. These results appear consistent with the mechanisms of debris action; as the kinetic energy of the flow (as well as the moving debris) will decrease progressively as it travels: Wood and Masonry buildings being the weakest construction types, it is expected that the effects of projectiles impacting such buildings will be influential over longer distances

Table 4 Overall penalized accuracy of the ordinal model calculated using N repetitions of a tenfold cross-validation algorithm. The values in bold represent the maximum accuracy obtained for each building type

Debris impact radius (m)	Cross-validated penalized accuracy (RC) (%) $N = 27$	Cross-validated penalized accuracy (Steel) ^a (%) $N = 81$	Cross-validated penalized accuracy (Wood) (%) $N = 1^b$	Cross-validated penalized accuracy (Masonry) (%) $N = 1^b$
10	90.79	91.64	96.30	93.06
20	91.25	92.26	96.41	93.40
30	91.07	92.84	96.48	93.43
40	91.34	92.71	96.59	93.14
50	91.05	92.50	96.62	93.20
60	90.99	92.34	96.63	93.40
70	91.06	92.22	96.74	93.55
80	91.38	91.85	96.75	93.57
90	91.82	91.81	96.79	93.60
100	91.04	91.82	96.85	93.64
110	91.48	91.63	96.97	93.79
120	91.43	91.74	97.01	93.86
130	91.42	91.63	97.12	93.82
140	91.39	91.72	97.12	93.79
150	91.41	91.55	97.10	93.83

^a Steel buildings, the number of stories and associated interactions were excluded from the model

^b Wood and Masonry buildings, the model was found to be perfectly stable thus a 10-fold cross-validation procedure without repetitions is sufficient

Fig. 6 Density histograms representing the distribution of each building material according to the observed damage, **a** Wood, **b** Masonry, **c** RC, **d** Steel

from the point of debris generation in comparison with strong structural materials such as RC and Steel. We can also observe that the variations in model accuracy across the different radiuses considered for debris impact are relatively small: a maximum of 1.03 % (± 2 buildings) for RC, 1.29 % (± 1 building) for Steel, 0.82 % (± 44 buildings) for Wood and 0.80 % (± 10 buildings) for Masonry. This means that between 10 m and 150 m only from a washed away house, the change in the debris impact variable (i.e., the consideration of buildings impacted further from the debris generation point) does not yield large differences in model performance. This possibly indicates that the effects of debris impact should be considered on a much larger scale and/or that the debris impact variable should be refined (for example, by using visual evidence of impact for individual buildings, as done by Reese et al. (2011) in the case of the 2009 Samoa tsunami). Because such evidence is not available, for derivation of the final fragility functions in this study we will use the debris impact radius resulting in the highest predictive accuracy, as shown in Table 4.

4.3 Fragility functions

The multivariate fragility functions—Eq. (4)—corresponding to the highest damage state (DS5, defined in this study as the aggregation of damage states 5 and 6 defined by the MLIT, see Table 1) are represented for each building class in Figs. 7, 8, 9 and 10 by fragility surfaces. The range of observed explanatory variable values for the expected

probabilities (circles) plotted on Figs. 7, 8, 9 and 10 correspond to the values of maximum flow depths and velocities from the observed data, illustrating the extent of available data coverage. The smooth surfaces are constructed not through spatial interpolation, but by approximating the discrete expected probabilities on a two-dimensional grid (D'Errico 2010). Because the debris impact has been defined as a binary variable, it is possible to represent separately the probability outcomes with and without debris impact.

The results presented for Wood buildings (Fig. 7) show very clearly that buildings which have been impacted will rapidly reach high probabilities of collapse (i.e., >0.9), from flow depths and velocities as low as 2.5 m and 5.4 m/s, respectively. On the other hand, buildings which have not been impacted will experience a milder increase in collapse probability with flow depth and particularly velocities, ultimately reaching the higher probability range for a combination of very high depths and velocities.

The evolution of collapse probabilities for Masonry buildings (Fig. 8) behaves similarly to the one of Wood buildings. Again the occurrence of debris impact seems to trigger a much steeper rise in damage probability in comparison with the no impact case. However, compared to Wood buildings, high probabilities of collapse ($>90\%$) when the structure has been impacted seem to occur after flow depth and velocity have reached higher values (7 m and 9.5 m/s, respectively). According to the present results, buildings which have not been impacted would reach a 90 % collapse probability for a combination for moderately

Fig. 7 Fragility surfaces showing the probability of collapse (i.e., DS5) for Wood buildings, depending on the occurrence of debris impact (radius: 130 m). The expected probabilities are displayed, which color and circle size vary to represent the number of stories (*red* 1 story, *orange* 2 stories, *cyan* 3 stories, *black* 5 stories or more. There are no buildings in the Wood database with 4 stories)

Fig. 8 Fragility surfaces showing the probability of collapse (i.e., DS5) for Masonry buildings, depending on the occurrence of debris impact (radius: 120 m). The expected probabilities are displayed, which color and circle size vary to represent the number of stories (red 1 story, orange 2 stories, cyan 3 stories, green 4 stories, black 5 stories or more)

severe flow conditions (a flow depth of 4 m associated with a flow velocity of 6 m/s). This can be explained by the fact that Masonry structures are relatively more resistant than Wood buildings (Leone et al. 2011; Suppasri et al. 2013a, b). In addition, it can be noticed from Fig. 8 that the range of flow depths and velocities available for buildings which have not been impacted is very reduced in comparison with the impact case. A meaningful extrapolation and interpretation of the fragility surface beyond the range of the available data would require additional statistical prediction methods to be discussed and applied; however, it is expected for larger values of tsunami IMs the shape of the fragility surface in the no impact case would resemble the one for Wood buildings.

The fragility surfaces for RC buildings (Fig. 9) also show an important increase in collapse probability in the case of debris impact, with very rare occurrences of collapse when the building is not impacted. Similarly to the dataset for Masonry buildings, the range of flow depths and velocities in the no impact case is much smaller than in the impact case, with most expected probabilities of collapse being very small (i.e., <0.15). The shape of the corresponding fragility surface for larger values of these IMs is thus difficult to infer. In line with previous studies (Rossetto et al. 2007; Mas et al. 2012; Suppasri et al. 2013a, b), the much steeper fragility surfaces for Wood and Masonry structures show that RC buildings are more resistant to tsunami loads. An interesting feature of the fragility surface corresponding to impacted buildings is the visibly stronger influence of flow velocity on

Fig. 9 Fragility surfaces showing the probability of collapse (i.e., DS5) for RC buildings, depending on the occurrence of debris impact (radius: 90 m). The expected probabilities are displayed, which color and circle size vary to represent the number of stories (red: 1 story, orange: 2 stories, cyan: 3 stories, green: 4 stories, black: 5 stories or more)

the rate of increase in damage probability, compared to flow depth. This could be expected as the dynamic impact force is proportional to the square of the flow velocity, although such relativity between the effects of the two variables seems less prominent in the case of weaker structural types. This indicates that hydrostatic loads alone may not significantly determine RC building collapse, and is consistent with previous findings demonstrating flow depth alone to be a poor predictor of tsunami damage for RC structures (Charvet et al. 2014b).

In the case of Steel buildings (Fig. 10), the difference between the impact and no impact cases is less striking than for other building types. The rate of increase in collapse probability appears similar in both cases, with regard to flow velocity. The increase in damage probability with flow depth appears very minor for buildings which have not suffered debris impact, and in line with the result from Charvet et al. (2014b) for Steel buildings this also indicates the hydrostatic loads may not have a significant influence on damage for this construction material. Nevertheless, it can be seen that the scarcity of data for this construction type leads to results which are inconsistent with the other building classes.

The construction of the fragility surface for the non-impacted buildings in the high probability region is mainly driven by one high probability point, with most of the data lying in a range of very low probabilities (i.e., <0.10). Similarly, the collapse probabilities for impacted buildings are mostly <0.50 , so half of the probability domain cannot be

Fig. 10 Fragility surfaces showing the probability of collapse (i.e., DS5) for Steel buildings, depending on the occurrence of debris impact (radius: 30 m). The expected probabilities are displayed; however, due to the lack of information on the number of stories for Steel buildings, no graphical distinction could be made (all circles represent “unknown” number of stories)

reliably covered by the fragility function. Thus, the probability values shown by the fragility surfaces for Steel buildings should be considered of limited use, until such functions can be refined using additional data. Information on building height is also crucial to update model estimations consistently with the other construction classes considered.

Given the nature of the new fragility functions estimated in this study, which are multivariate, derived using different statistical assumptions, methods, and model diagnostics compared to previously published research, the fragility results cannot be directly and qualitatively compared. Nevertheless, and regardless of the statistical estimation methodology adopted, all previous studies without exception derived two-dimensional fragility functions (i.e., fragility curves), mostly relating the maximum tsunami flow depth to damage probabilities. Therefore, as a point of reference fragility curves are also derived here, based on the measured flow depth only as explanatory variable (Fig. 11), and consistently with the ordinal regression method adopted throughout the paper. Table 5 shows example combinations of maximum flow depths and velocities yielding specific collapse probability thresholds (i.e., 0.25, 0.50, 0.75, 0.95) from the fragility results presented in this section (Figs. 7, 8, 9, 10, 11). This allows for the comparison of outcomes between the multivariate fragility functions (in cases of both impact and no impact), and simple fragility curves. The examples values chosen indicate fragility curves based on flow depth

Fig. 11 Two-dimensional fragility functions derived using ordinal regression, and flow depth as the only explanatory variable

only may underestimate collapse probabilities, particularly in cases where debris impact has a dominant effect.

5 Conclusions

The most commonly used tsunami intensity measure in the derivation of fragility functions is the flow depth measured from post-tsunami surveys; in a few cases the tsunami velocity, retrieved from numerical simulations. To date, fragility functions have all been empirical (based on field damage data) and typically derived using linear least squares regression, assuming a normal or lognormal distribution of the response (i.e., damage state). Latest research has found this method to have serious limitations and be in principle inadequate in the estimation of tsunami-induced damage probability. Therefore, improved statistical modeling techniques, based on generalized linear models, have been recently proposed and applied to case studies of building damage caused by the 2011 Great East Japan tsunami. The uncertainty still visible from recent model estimations was attributed to a lack of explanatory variables, as flow depth alone may not be able to explain the full extent of observed building damage. Indeed, flow velocity, debris impact and scour concurrently act on the affected structure during a tsunami attack.

Table 5 Combinations of flow depth and velocity values for corresponding collapse probabilities, using fragility surfaces (impact/no impact) and fragility curves (one explanatory variable only)

Collapse probability	Impact (all variables)	No impact (all variables)	Flow depth only (m)
Wood			
0.25	0.50 m and 0.50 m/s	2.30 m and 3.80 m/s	1.1
0.50	1.00 m and 0.50 m/s	3.50 m and 4.50 m/s	1.6
0.75	1.60 m and 1.80 m/s	5.50 m and 2.50 m/s	2.4
0.95	4.00 m and 6.50 m/s	7.00 m and 9.00 m/s	4.5
Masonry			
0.25	1.00 m and 1.50 m/s	2.20 m and 3.00 m/s	2.1
0.50	2.50 m and 3.50 m/s	3.50 m and 4.50 m/s	3.2
0.75	3.50 m and 5.50 m/s	4.00 m and 5.00 m/s	4.8
0.95	5.50 m and 5.80 m/s	5.00 m and 5.00 m/s	9.5
RC			
0.25	3.10 m and 4.00 m/s	2.00 m and 2.00 m/s	4.2
0.50	5.00 m and 6.00 m/s	2.50 m and 2.00 m/s	6.1
0.75	6.40 m and 8.00 m/s	3.00 m and 2.00 m/s	8.8
0.95	8.50 m and 12.5 m/s	3.00 and 2.50 m/s	N/A
Steel			
0.25	4.00 m and 5.00 m/s	4.00 m and 5.50 m/s	4.8
0.50	6.00 m and 7.00 m/s	5.50 m and 6.00 m/s	6.5
0.75	2.00 m and 16.5 m/s	6.00 m and 7.50 m/s	8.9
	Or		
	11.5 m and 2.5 m/s		
0.95	1.00 m and 19 m/s	8.00 m and 8.00 m/s	N/A
	Or		
	14.5 m and 2.5 m/s		

In the case of fragility surfaces, more than a single combination of flow depth and velocity may yield the same damage probability (the values given here are only examples). N/A indicates the fragility curve does not extend to the target probability range

Therefore, the focus of the present analysis was the application of generalized linear models (more specifically ordinal regression) to building damage data from the 2011 Great East Japan tsunami event, simultaneously taking into account the maximum tsunami flow depth, velocity, and debris impact as explanatory variables. Ordinal regression is a specific case of GLM application where the response variable (i.e., damage state) is assumed to be discrete and ordered, which was deemed suitable to represent the damage classification. Kesennuma city in Japan was one of the most severely affected cities during the event, and a large disaggregated database of building damage and measured flow depths was provided by the city thus making this area the chosen case study. Tsunami inundation simulations were carried out using nonlinear shallow water equations and a nested grid system. The maximum numerical flow velocities were retrieved from the simulation results. The debris impact variable was defined as a binary indicator, determined by the vicinity of the target building to a structure which had been washed away (i.e., assuming the remains of this structure would impact the surrounding buildings). As such, the debris impact variable was assigned a value of ‘1’ if the target building was considered impacted by debris, i.e., close

to at least one washed away structure, ‘0’ otherwise. Because the distance of significant debris impact is unknown, a range of debris action radiuses (10 m–150 m) was considered.

The first stage of the analysis consisted in selecting the model components through an assessment of the relative goodness-of-fit between several candidate models. Firstly, the assumption that an ordered model was preferable to a multinomial model was verified using a likelihood ratio test. The same procedure allowed for checking the significance of all explanatory variables and interactions to be added to the simplest model, based on maximum flow depth only. The additional variables considered were maximum flow velocity, debris impact, building material, and number of stories. The AIC was used to choose the most appropriate link function. The final GLM was chosen to be ordered, with a logit link function, and including flow depth, flow velocity, debris impact, building material, number of stories and most associated interactions as explanatory variables. In order to quantify the absolute goodness-of-fit, a penalized accuracy measure was proposed. This measure is considered a more suitable alternative for ordinal models compared to the traditional accuracy (or classification error rate), better suited to binary classifications.

The second stage of the analysis consisted in using this measure, estimated through a repeated tenfold cross-validation scheme, to assess the predictive accuracy (i.e., performance of the proposed model with unseen data) for each building material and each debris impact radius. It was found the model chosen was able to predict the observed damage states in Kesennuma with high accuracy (>90 %) for all building classes. It is possible, however, that these values of accuracy also partly reflect the extreme commonality of high damage states in the database, and alternative measures of goodness-of-fit, such as meaningful correlation measures between observed and expected probabilities (Zheng and Agresti 2000) may need to be developed to complete our understanding of model performance.

The final stage of the analysis focused on the representation of the building fragility results in Kesennuma in the form of fragility surfaces for each building material. Each surface showed the change in building collapse probability with simultaneous changes in maximum flow depth and velocity, considering also the number of stories of the structure. The fragility trends were found to be consistent with previous studies, with Wood buildings found to be the least resilient to tsunami loads, followed by Masonry buildings, RC and Steel buildings found to be the strongest structural types. For RC and particularly for Steel buildings in the no impact case, it was also found that the change in damage probability was only weakly explained by flow depth, confirming previous findings on the poor performance of flow depth alone as a predictor of tsunami damage for these structures. The results obtained with multidimensional fragility functions were compared with classic two-dimensional fragility curves for the city of Kesennuma, and the results indicate that fragility curves, based on flow depth only, may significantly underestimate building collapse probabilities. Except for the case of Steel buildings, for which additional data would be desirable to obtain reliable fragility results, fragility surfaces for all structural materials show the influence of debris impact to be paramount in the determination of high collapse probabilities. In order to provide high-quality fragility assessment tools, interpretable and usable for the tsunami risk assessment community, future studies on tsunami-induced damage probability should aim to:

- Refine the debris impact variables to be used for fragility assessment (for example by considering variability in debris size, mass, sheltering, azimuth of impact),
- Improve numerical velocity calculations by using more sophisticated fluid dynamics models, dynamically varying friction conditions, and detailed topography information,

- Investigate the effects of combined, time-dependent flow depths and velocity to capture more accurately the demand imposed by hydrodynamic loads,
- Quantify the effects of scour on high structural damage probabilities,
- Provide meaningful representations of multivariate fragility functions. Address uncertainties in parameter estimation (measurement error in the variables) by using Bayesian estimation techniques as alternatives to maximum likelihood.

Acknowledgments This work has been supported by Grant-in-Aid for Scientific Research (A) NO.26242033 of the Japan Society for the Promotion of Science (JSPS), the work of Dr Ingrid Charvet has been funded by JSPS. Detailed building damage data were officially provided by Kesennuma city governmental office. The contribution of Dr. Anawat Suppasri and Pr Fumihiko Imamura has also been supported by the Tokio Marine Nichido via Tohoku University.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Akaike, H (1974) A new look at the statistical model identification. *IEEE Trans Automat Control* 19(6):716–723
- Architectural Institute of Japan (2012) Preliminary reconnaissance report of the 2011 Tohoku-Chiho Taiheiyō-Oki Earthquake. Series: geotechnical, geological and earthquake engineering, vol 23, 1st edn. Springer, Japan
- Arikawa T (2009) Structural behavior under impulsive tsunami loading. *J Disaster Res* 4(6):377–381
- Arnasson H, Petroff C, Yeh H (2009) Tsunami Bore impingement onto a vertical column. *J Disaster Res* 4(6):391–403
- Baldi P, Brunak S, Chauvin Y, Andersen CAF, Nielsen H (2000) Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 16(5):412–424
- Charvet I (2012) Experimental modelling of long elevated and depressed waves using a new pneumatic wave generator. PhD thesis, University College London
- Charvet I, Eames I, Rossetto T (2013a) New tsunami runup relationships based on long wave experiments. *Ocean Model* 69:79–92
- Charvet I, Suppasri A, Imamura F, Rossetto T (2013b) Comparison between linear least squares and GLM regression for fragility functions: example of the 2011 Japan tsunami. In: Proceedings of international sessions in coastal engineering, JSCE, vol 4, November 13–15, 2013, Fukuoka, Japan
- Charvet I, Ioannou I, Rossetto T, Suppasri A, Imamura F (2014a) Empirical vulnerability assessment of buildings affected by the 2011 Great East Japan tsunami using improved statistical models. *Nat Hazards*. doi:10.1007/s11069-014-1118-3
- Charvet I, Suppasri A, Imamura F (2014b) Empirical fragility analysis of building damage caused by the 2011 Great East Japan Tsunami in Ishinomaki City using ordinal regression, and influence of key geographical features. *Stoch Environ Res Risk Assess*. doi:10.1007/s00477-014-0850-2
- Chock GYK, Robertson I, Riggs HR (2011) Tsunami structural design provisions for a new update of building codes and performance-based engineering. In: Proceedings of solutions to coastal disasters 2011, June 26–29, 2011, Anchorage, Alaska
- Chock F, Carden L, Robertson I, Olsen N, Yu G (2013) Tohoku Tsunami-induced building failure analysis with implications for U.S. Tsunami and seismic design codes. *Earthq Spectra* 29(S1):S99–S126
- D'Errico J (2010) Surface fitting using gridfit. MATLAB Central File Exchange. http://uk.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit/content/gridfitdir/demo/html/gridfit_demo.html
- Dominik B, Pudasaini SP (2012) Full two-dimensional rapid chute flows of simple viscoplastic granular materials with a pressure-dependent dynamic slip-velocity and their numerical simulations. *J Non-newton Fluid Mech* 173–174:72–86

- EEFIT (2013) Recovery after the 2011 Tohoku Earthquake and Tsunami: a return mission report by EEFIT. Institution of Structural Engineers (online), www.istructe.org/resources-centre/technical-topic-areas/eeffit/eeffit-reports
- Fahrmeir L, Tutz G (2001) Multivariate statistical modelling based on generalized linear models. Springer series in statistics, 2nd edn. Springer, New York
- Federal Emergency Management (FEMA) (2008) Guidelines for design of structures for vertical evacuation from tsunamis. FEMA report 646. <http://www.fema.gov/media-library/assets/documents/14708?id=3463>
- Forbes C, Evans M, Hastings N, Peacock B (2011) Statistical distributions. Wiley series in probability and statistics, 4th edn. Wiley, New York
- Fritz HM, Phillips DA, Okayasu A, Shimozono T, Liu H, Mohammed F, Skanavis V, Synolakis CE, Takahashi T (2012) The 2011 Japan tsunami current velocity measurements from survivor videos at Kesennuma Bay using LIDAR. *Geophys Res Lett* 39(7):L00G23
- Green SB (1991) How many subjects does it take to do a regression analysis? *Multivar Behav Res* 26(3):499–510
- 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
- Hu B, Palta M, Shao J (2006) Properties of R^2 statistics for logistic regression. *Stat Med* 25:1383–1395
- Hughes SA (1993) Physical models and laboratory techniques in coastal engineering. World Scientific, London
- Imamura F (1996) Review of tsunami simulation with a finite difference method, long-wave runup models. In: Yeh H, Liu P, Synolakis C (eds) World Scientific, pp 25–42 (ISBN981-02-2909-7)
- Kim J-H (2009) Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal* 53:3735–3745
- Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the international joint conference on artificial intelligence (IJCAI 1995) Montreal, Canada
- Koshimura S, Oie T, Yanagisawa H, Imamura F (2009) Developing fragility function for tsunami damage estimation using numerical model and post-tsunami data from Banda Aceh, Indonesia. *Coast Eng J* 51:243–273
- Leelawat N, Suppasri A, Charvet I, Imamura F (2014) Building damage from the 2011 Great East Japan Tsunami: quantitative assessment of influential factors. *Natural Hazards*, published online, A new perspective on building damage analysis. doi:10.1007/s11069-014-1081-z
- Leone F, Lavigne F, Paris R, Denain JC, Vinet F (2011) A spatial analysis of the December 26th, 2004 tsunami-induced damages: lessons learned for a better risk assessment integrating buildings vulnerability. *Appl Geogr* 31:363–375
- Lloyd T (2014) An experimental investigation of tsunami forces on coastal structures. PhD thesis, University College London
- Mader CL (2004) Numerical modelling of water waves, 2nd edn. CRC Press, London
- Mas E, Koshimura S, Suppasri A, Matsuoka M, Matsuyama M, Yoshii T, Jimenez C, Yamazaki F, Imamura F (2012) Developing Tsunami fragility curves using remote sensing and survey data of the 2010 Chilean Tsunami in Dichato. *Nat Hazards Earth Syst Sci* 12:2689–2697
- Mc Cullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman & Hall/CRC, Boca Raton
- Myung J (2003) Tutorial on maximum likelihood estimation. *J Math Psychol* 47(1):90–100
- Pudasaini SP (2012) A general two-phase debris flow model. *J Geophys Res* 117:1–28
- Pudasaini SP (2014) Dynamics of submarine debris flow and tsunami. *Acta Mech* 225:2423–2434
- Pudasaini SP, Dominik B (2009) Energy considerations in accelerating rapid shear granular flows. *Nonlinear Process Geophys* 16:399–407
- Reese S, Cousins WJ, Power WL, Palmer NG, Tejakusuma IG, Nugrahadi S (2007) Tsunami Vulnerability of buildings and people in South Java—field observations after the July 2006 Java tsunami. *Nat Hazards Earth Syst Sci* 7:573–589
- Reese S, Bradley BA, 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(2011):156–173
- Rennie JDM, Srebro N (2005) Loss functions for preference levels: regression with discrete ordered labels. In: Proceedings of the international joint conference on artificial intelligence (IJCAI 2005) Edinburgh, UK
- Rossetto T, Peiris N, Pomonis A, Wilkinson SM, Del Re D, Koo R, Gallocher S (2007) The Indian Ocean Tsunami of December 26, 2004: observations in Sri Lanka and Thailand. *Nat Hazards* 42(1):105–124
- Rossetto T, Ioannou I, Grant DN (2014) Guidelines for empirical vulnerability assessment. GEM technical report. GEM Foundation, Pavia

- Ruangrassamee A, Yanagisawa H, Foytong P, Lukkunaprasit P, Koshimura S, Imamura F (2006) Investigation of Tsunami-induced damage and fragility of buildings in Thailand after the December 2004 Indian Ocean Tsunami. *Earthq Spectra* 22:377–401
- Suppasri A, Koshimura S, Imamura F (2011) Developing tsunami fragility curves based on the satellite remote sensing and the numerical modeling of the 2004 Indian Ocean tsunami in Thailand. *Nat Hazards Earth Syst Sci* 11(1):173–189
- Suppasri A, Mas E, Koshimura S, Imai K, Harada K, Imamura F (2012a) Developing tsunami fragility curves from the surveyed data of the 2011 Great East Japan tsunami in Sendai and Ishinomaki Plains. *Coast Eng J* 54(1):1250008 (Special Anniversary Issue on the 2011 Tohoku Earthquake Tsunami)
- Suppasri A, Koshimura S, Matsuoka M, Gokon H, Kamthonkiat D (2012b) Application of remote sensing for tsunami disaster. In: Chemin Y (eds) *Remote sensing of planet earth*. InTech, ISBN 978-953-307-919-6
- Suppasri A, Charvet I, Imai K, Imamura F (2013a) Fragility curves based on data from the 2011 Great East Japan tsunami in Ishinomaki city with discussion of parameters influencing building damage. *Earthq Spectra*. doi:[10.1193/053013EQS138](https://doi.org/10.1193/053013EQS138)
- Suppasri A, Mas E, Charvet I, Gunasekera R, Imai K, Fukutani Y, Abe Y, Imamura F (2013b) Building damage characteristics based on surveyed data and fragility curves of the 2011 Great East Japan tsunami. *Nat Hazards* 66(2):319–341
- Suppasri A, Muhari A, Futami T, Imamura F, Shuto N (2014) Loss functions for small marine vessels based on survey data and numerical simulations of the 2011 Great east Japan Tsunami. *J Waterway Port Coast Ocean Eng* 140:5
- Zheng B, Agresti A (2000) Summarizing the predictive power of a generalized linear model. *Stat Med* 19:1771–1781