16th World Conference on Earthquake, 16WCEE 2017
Santiago Chile, January 9th to 13th 2017
Paper N° 1794
Registration Code: S- L1464704285

PROPOSED METHODOLOGY FOR DEFINING OPTIMAL INTENSITY MEASURES FOR EMPIRICAL TSUNAMI FRAGILITY FUNCTIONS

J. MACABUAG[1], T. ROSSETTO[2], I. IOANNOU[3],
A. SUPPASRI[4], D. SUGAWARA[5], B. ADRIANO[6], F. IMAMURA[7], I. EAMES[2], S. KOSHIMURA[7]

[1] *EngD Candidate, EPICentre, University College London, UK, macabuag@gmail.com*
[2] *Professor, EPICentre, University College London*
[3] *Post-doctoral Researcher, EPICentre, University College London*
[4] *Associate Professor, IRIDeS, Tohoku University, Sendai, Japan*
[5] *Museum of Natural and Environmental History, Shizuoka, Japan*
[6] *PhD Candidate, IRIDeS, Tohoku University*
[7] *Professor, IRIDeS, Tohoku University*

Abstract

Tsunami fragility functions for buildings provide a probabilistic link between tsunami intensity and building damage. They are a component of tsunami risk models, and so are vital for land-use and emergency planning, performance-based engineering, as well as human and financial loss estimation. When selecting or developing risk models, a Tsunami Intensity Measure (TIM) (e.g. inundation depth, velocity, force estimates etc) must be selected which provides the best possible representation of the damage potential of the tsunami inundation. However, there is no consensus as to which flow parameter is the most appropriate TIM to estimate fragility.

This paper presents a rigorous methodology using advanced statistical methods for the selection of the optimal TIM for fragility function derivation for any given dataset. This methodology is demonstrated using a unique, detailed, disaggregated damage dataset from the 2011 Great East Japan Earthquake and Tsunami, identifying the optimum TIM for describing observed damage for the case-study locations. Several advanced statistical methods are introduced which are novel in the fields of fragility analysis: multiple imputation for treatment of incomplete data-entries, semi-parametric Generalised Additive Models, and k-fold cross-validation for model comparison and optimization.

The methodology presented in this paper has application for researchers and risk modellers in the engineering, DRR and insurance industries; and the statistical methods presented have implications for fragility function derivation and selection for multiple hazards.

Keywords: Tsunami damage; Empirical fragility curves; Cross-validation; Multiple imputation; Intensity measures.

Introduction

For the purpose of making building damage predictions for future tsunamis it is generally unrealistic to define deterministic relationships (i.e. predicting with absolute certainty the exact damage caused by a tsunami to each of a population of buildings), and so statistical approaches are more appropriate. Tsunami fragility functions are statistical models which give the probability of damage exceedance (i.e. the probability that the damage state experienced by a building, ds, will be greater than or equal to a defined damage state, DS) as a function of the Tsunami Intensity Measure (TIM, a parameter used to define the flow conditions at each building location) (equation (1)). They are often presented as cumulative distribution functions derived by applying statistical model fitting techniques to building damage data.

$$P(ds \geq DS|TIM) = f(TIM) \tag{1}$$

The TIM should provide the best possible representation of the damage potential of the tsunami inundation. Tsunami-induced building damage can arise due to hydrostatic forces (including buoyancy), hydrodynamic effects (drag and bore impact) and debris (impact and damming). The severity of these effects are determined by a number of flow parameters, yet the majority of existing tsunami fragility curves adopt only the local maximum inundation depth as the TIM, often because it can be estimated from post-tsunami reconnaissance of buildings. Other parameters of the flow can be derived from inundation modelling, and existing studies use velocity and hydrodynamic force [1] [2], momentum flux (an indicator of drag force), moment of momentum flux (the product of momentum flux and inundation depth) [3] and energy head according to the Bernoulli Equation [4]. All force estimations in previous studies have been based on the standard form drag equation, and so do not account for alternative estimations such as equivalent hydrostatic methods [5], bore impact [6] or changes in flow regime [7]. Overall these studies do not show a consensus as to which flow parameter is the most appropriate TIM to estimate fragility.

Some existing studies attempt to define the optimal TIM based on qualitative visual assessments [8] or using small datasets [4]. A number of seismic studies have compared seismic Intensity Measures (IMs) using the criteria of “efficiency” [9], “sufficiency” [10] and “computability” [11], though these have often been based directly on continuous structural response (in terms of Engineering Demand Parameters (EDP)) derived from numerical analyses, and so the same methods are not appropriate for fragility functions formed on the discrete damage states of observational damage data. To date no existing study has compared efficiency of multiple TIMs based on empirical fragility curves fit to observed damage data.

This paper presents a proposed rigorous methodology using advanced statistical methods for the selection of the optimal TIM for fragility function derivation for any given dataset. This methodology is demonstrated using a unique, detailed, disaggregated damage dataset from the 2011 Great East Japan Earthquake and Tsunami, identifying the optimum TIM for describing observed damage for the case-study locations. The proposed methodology consists of the three steps shown in Figure 1 [12].

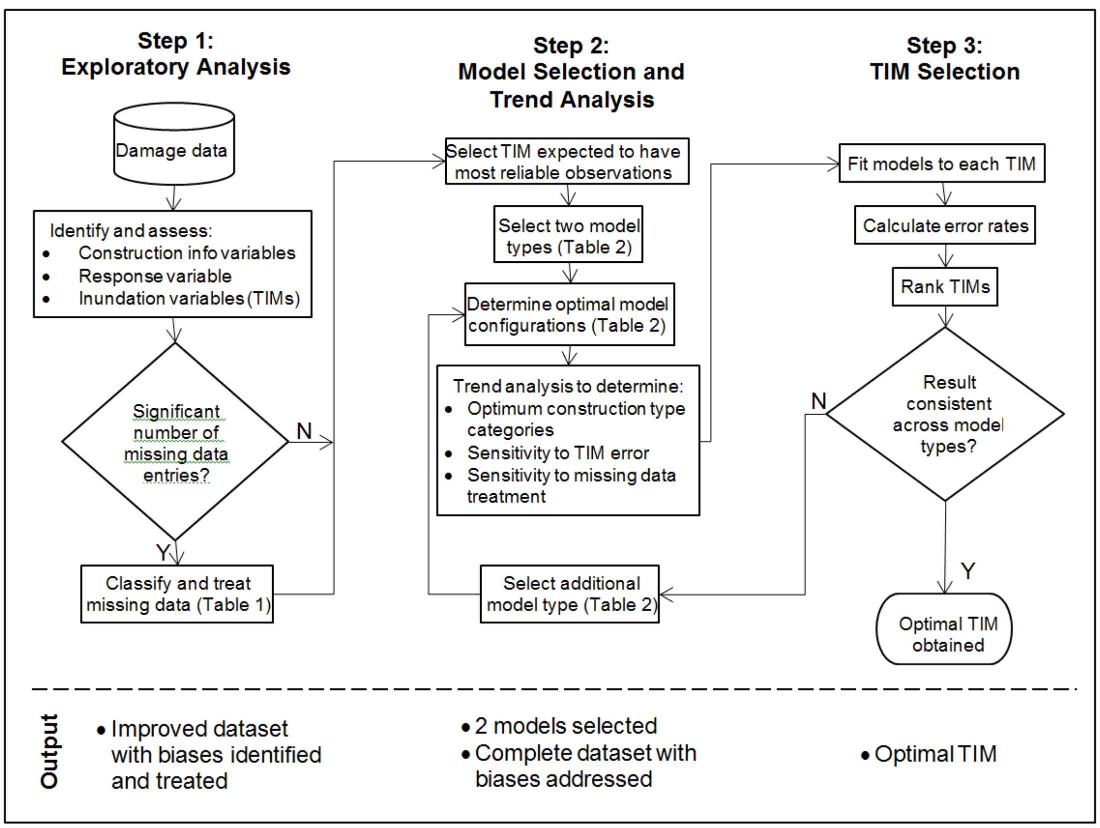


Figure 1: Methodology flow chart with expected outputs at each Step [12].

Step 1. Exploratory Analysis of Data Quality

The aim of Step 1 is to identify the response and explanatory variables from the information available in the database, and to identify and treat any underlying bias. This is achieved in two stages:

1. Identify and assess the following categories of variables:
   a. Building construction variables.
   b. The response variable: Damage state definitions and distributions.
   c. Inundation variables: Identify and validate TIMs.
2. Classify and treat incomplete data-entries.

Items 1 above is discussed in [12], [13] and are demonstrated for the case study database below.

Regarding treatment of missing data, previous studies generally conduct complete-case analysis [14], i.e. they remove any partial data, such as buildings of unknown material, from their fragility analysis. However, not dealing with missing data leads to a loss of statistical power and bias if the missing data is informative. According to the guidelines set out by [15], the approach to be used for dealing with missing data depends on whether the data is Missing Completely At Random (MCAR), Missing At Random (MAR), or Missing Not At Random (MNAR) (Table 1). MCAR refers to the case where the data is missing purely by chance, in which case complete-case analysis may be conducted without introducing bias in the results. MNAR refers to the case where the missing information is related to the reason that the information is missing (e.g. if wooden buildings had been removed from the dataset because they were wooden), in which case complete case analysis would introduce bias and missing data cannot be estimated, and so the dataset must be supplemented with additional information to address this before fragility analysis can be conducted. MAR refers to the case where the information is not missing completely at random but can be accounted for by using other attributes, in which case the missing data may be estimated by Multiple Imputation (MI) techniques. MI involves replacing missing observed data with substituted values estimated multiple times via stochastic regression models built on the other attributes (used as explanatory variables), with all of the imputations being combined in order to derive the final estimate.

Classification	Method of Identification	Recommended Action
Missing Completely At Random (MCAR)	Test whether the missing data distribution is the same as for the complete dataset (Kolmogorov-Smirnoff test for disaggregated data, or χ^2-test for aggregated data).	Conduct Complete-Case Analysis (i.e. remove datapoints with missing information and perform regression analysis on the remaining dataset), or estimate missing data using Multiple Imputations techniques.
Missing Not at Random (MNAR)	Is the missing information related to the reason that the information is missing?	Fragility analysis cannot be conducted without introducing bias. Revisit data-collection process to complete missing data.
Missing at Random (MAR)	Not MCAR or MNAR.	Estimate missing data using Multiple Imputations techniques.

Table 1: Classification and treatment of missing data.

Case Study: Building Damage Dataset

The building damage data used in this paper is taken from the Great East Japan Earthquake (2011) building damage database compiled by Japan's Ministry of Land Infrastructure Tourism and Transport (MLIT). The database is comprised of relevant information (including the number of floors, construction material, and building usage) for each individual building (circa 250,000) located within the inundation area of the GEJE, though information is generally not included for every field for each building. All buildings are allocated a damage state from DS0 to DS6 based on the damage scale presented in Table 2, and assigned an observed inundation depth. As discussed by [16] DS5 and DS6 do not represent progressively worse damage states, and so for the fragility function derivation in this study, damage states 5 and 6 are combined and collectively termed as DS5*.

DS1: Minor Damage Inundation below ground floor. Possible to use immediately after minor floor and wall cleanup.		**DS2: Moderate Damage** The building is inundated less than 1m above the floor. Possible to use after moderate repairs.	
DS3: Major Damage The building is inundated more than 1m above the floor (below the ceiling). Possible to use after major repairs.		**DS4: Complete Damage** The building is inundated above the ground floor level. Major work is required for re-use of the building.	
DS5*: Irreparable			
DS5: Collapsed The key structure is damaged. Not repairable.		DS6: Washed Away The building is completely washed away except for the foundation. Not repairable.	

Table 2 - Damage state definitions used by the Japanese Ministry of Land Infrastructure Tourism and Transport following the 2011 Great East Japan Earthquake and Tsunami. Descriptions from Japan Cabinet Office (2013), usage descriptions are after Suppasri et al. (2014).

In the present study three case-study locations are considered: Ishinokami, Onagawa, and Kesennuma (shown in Figure 2), representing 80%, 15%, and 5%, respectively, of the combined dataset (67,125 buildings). A closer look at the data shows that the distributions of buildings with different construction materials is similar for the three towns and that together they provide a better coverage of a range of inundation depths, and hence it is reasonable to combine the data from the three towns in order to provide a larger dataset, so enabling greater confidence in the derived fragility curves.

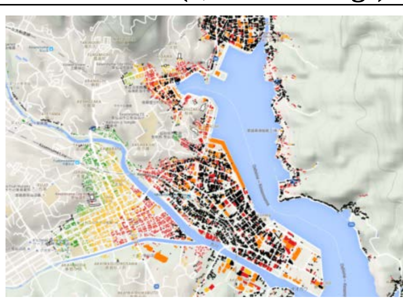
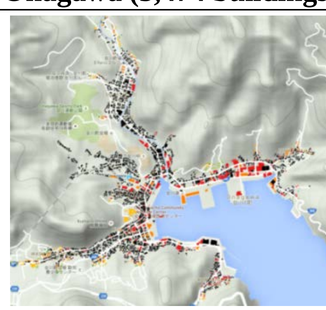
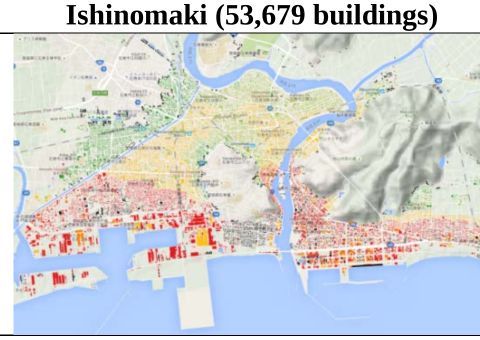
Kesennuma (9,979 buildings)	Onagawa (3,474 buildings)	Ishinomaki (53,679 buildings)
		

Figure 2: Case-study locations. GIS images have buildings coloured according to their observed damage state (right), where: white buildings indicate no damage (DS0), black indicates that buildings have been washed away (DS6) and all other damage states are coloured based on a scale from green (DS1) to red (DS5).

Producing fragility curves for each construction material requires splitting the data into small datasets for some materials (e.g. reinforced concrete, RC, buildings represent only 1.8% of the data, spread over the 5 damage states), which can result in larger uncertainty associated with the model parameter estimates. Inspection of the data shows that damage state distributions for wood and masonry (typically associated with non-engineered constructions) are very similar to each other. The same can be observed of the damage distributions for (RC) and steel (engineered) buildings. Comparison between the damage state distributions of engineered and non-engineered buildings instead shows significant differences. Hence, in this study fragility curves are developed for engineered and non-engineered structures, in order to account for the significant differences in the fragilities of such buildings, whilst maintaining large enough datasets to avoid greatly increasing uncertainty in the model parameter estimates.

Buildings of unknown construction material make up 18.1% of the total dataset within the inundated area, representing a significant proportion of the data and so it is necessary to analyse this missing data further so as to avoid the introduction of bias. If the missing data were MCAR (see Table 1) then there should be no relationship between the buildings that have missing material data and other attributes such as the building height, size and use. However, analysis of building footprint sizes (Figure 3) suggests that engineered buildings (RC and steel) are generally larger than non-engineered buildings (wood and masonry), with buildings of unknown material representing the smallest footprints. This suggests that many buildings of unknown material may represent non-engineered buildings. A Kolmogorov-Smirnoff test is conducted and confirms that footprint areas for the buildings of unknown material are not of the same distribution as for the total dataset (i.e. they have different probability density functions). Therefore, the missing building material data is not MCAR. MNAR would refer to, for example, if wooden buildings are more likely to have missing material data because they are wooden. However, there is no reason to believe that all the missing material data can be associated with either the engineered or non-engineered construction types. Hence, the missing building material data is not considered MNAR. MAR would be the case where, for example, small buildings are more likely to have missing material data, but this has nothing to do with material after accounting for size. This is more likely to be the case here, and hence we adopt a Multiple Imputation (MI) approach to assign building data for which construction material information is missing to either the engineered or non-engineered building categories.

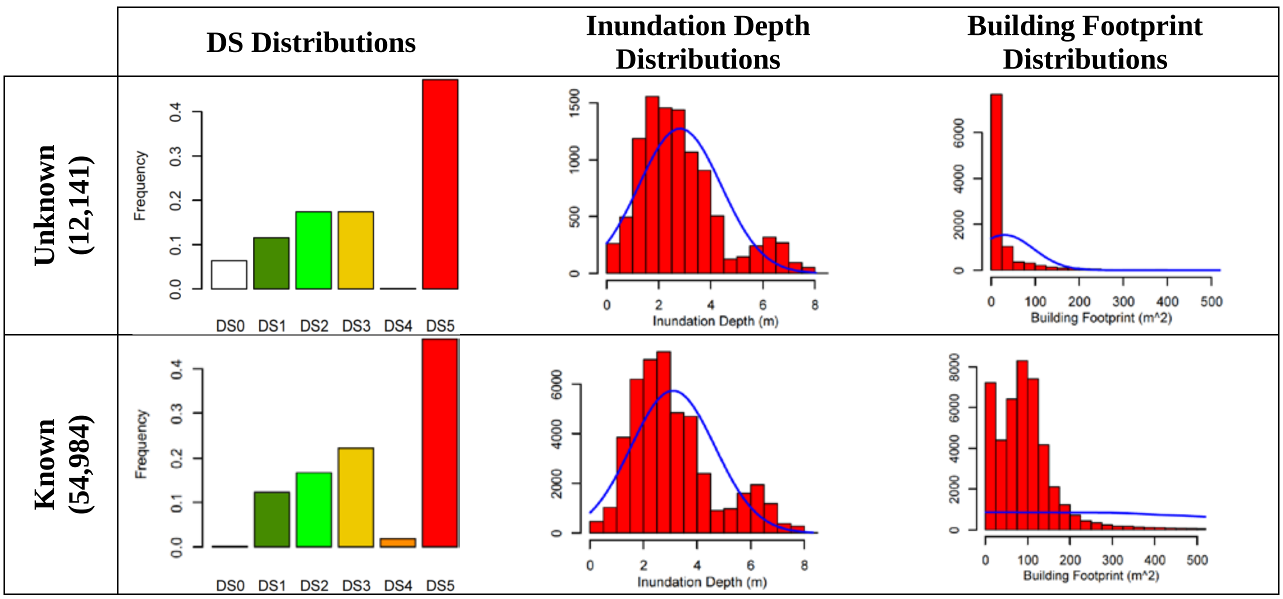


Figure 3: Damage State distributions, showing that buildings of unknown material type have a greater proportion of undamaged (DS0) buildings than buildings of known material type. Histograms and normal curves for building inundation depths and footprint areas for buildings of unknown (top) and known material (right).

In order to conduct MI, which attributes should be used for the imputation? It has already been shown that building footprint is an indicator of construction material. Figure 3 also shows that buildings of unknown material show a large proportion of undamaged (DS0) buildings. Visualization of building location by construction material shows no obvious spatial correlation of the unknown buildings. However, a Kolmogorov-Smirnov test performed on the observed inundation depths for unknown and known materials indicates that there is a very low probability (<5%) that the two datasets are drawn from the same underlying distribution (indeed Figure 3 shows that the distributions of inundation depths for buildings of unknown material do have a slight increase in the number of buildings at low simulated inundation depths). In addition, building usage information shows some correlation with construction material.

Therefore, Multiple Imputation analysis, with 4 imputations, is conducted in order to estimate building material based on footprint area, damage state, building use, and observed inundation depth. The effect of imputation on results is investigated further in Step 2.

Case Study: Tsunami Inundation Simulation Data

To supplement the observed inundation depth data, a numerical inundation simulation is conducted for the case-study locations to derive additional TIMs. The optimum TIM for this dataset will be defined in Step 3 from the TIMs shown in Table 3.

TIM1-TIM6 have already been discussed in the context of existing studies. The drag force is proportional to the local momentum flux and so is proportional to TIM4. TIM7 is an equivalent quasi-steady force proposed by [7] which is evaluated via two different flow regimes (sub-critical and choked flow) determined by Froude Number (readers are referred to [7] for the calculation procedure). All of the simulated TIM values are calculated at the geometrical centres of each building footprint for each time-step of the simulation, and the peak values extracted, with the exception of the equivalent peak momentum flux (MF_{equiv}, TIM5) and quasi-steady force estimation (F_{QS}, TIM7) both of which are calculated using the separate peak depth and peak velocity values (which do not occur at the same time).

Tsunami Intensity Measure		Symbol	Description
TIM1	Observed inundation depth	h_{obs}	Peak *observed* inundation depth
TIM2	Simulated inundation depth	h_{sim}	Peak *simulated* inundation depth
TIM3	Flow speed	v	Peak simulated velocity (v_{peak})
TIM4	Momentum flux	MF	$(hv^2)_{peak}$
TIM5	Equivalent peak momentum flux	MF_{equiv}	$(h)_{peak}.(v^2)_{peak}$
TIM6	Froude number	Fr	$\left({}^{v}/_{\sqrt{gh}}\right)_{peak}$
TIM7	Equivalent quasi-steady force	F_{QS}	Alternative steady-state force estimation considering choked and sub-critical flow [7]

Table 3: Alternative TIMs considered in this investigation.

The numerical tsunami inundation model is presented in detail and validated by [19], [20], and the tsunami source model is presented in [21]. The wave propagation and inundation calculation solves discretized non-linear shallow-water equations [22], [23] over six computational domains in a nested grid system (Figure 4), accounting for flow resistance via the Manning's roughness coefficient (n, taken as 0.025 to represent resistance from obstacles in the urban case study areas).

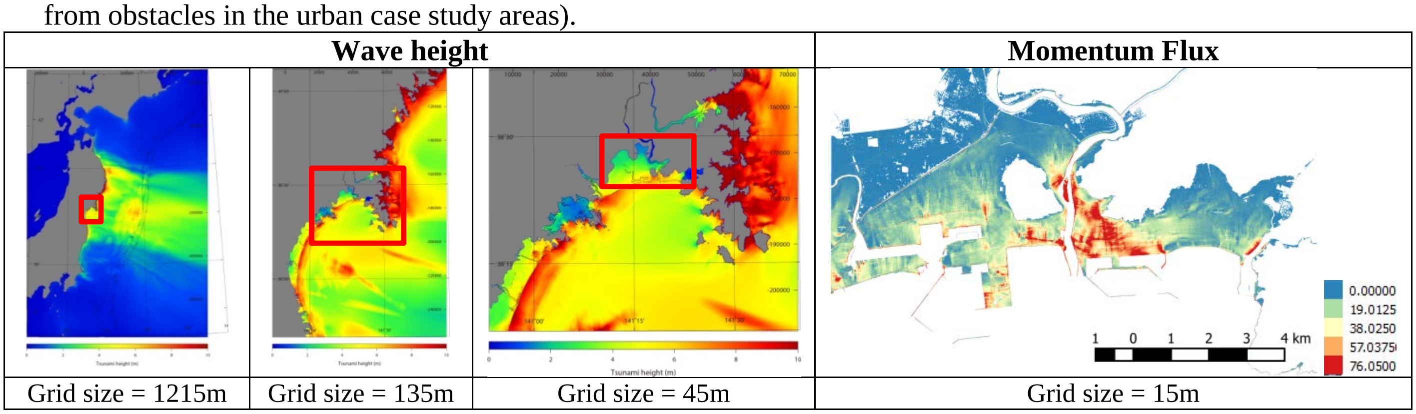


Figure 4: Some of the computational domains for the nested grid wave propagation and inundation model used for Ishinomaki (dx indicates the grid size). Example results for grid size = 15m inundation simulation are shown for momentum flux (TIM4). The results shown are the peak values for each grid square over the simulation period.

Step 2. Statistical Model Selection and Trend Analysis

The aim of Step 2 is to select appropriate models with which to conduct the TIM comparison of Step 3, and to use these models to supplement the exploratory analysis of Step 1. The outcome of Step 2 is to have selected at least two models and their optimum configurations, and to have a complete dataset with biases addressed. The statistical models can be parametric, semi-parametric, or non-parametric (Table 4). Step 2 consists of the stages shown in Figure 1.

Class	Model	Configuration Options	Selection Method	Reference
Parametric	~~Ordinary Least Squares (OLS)~~	(OLS not suitable for fragility function derivation)		Suppasri et al. (2012a); Suppasri et al. (2009); Tanaka & Kondo (2015)
	Generalised Linear Model (GLM) Or Cumulative Link Model (CLM)	Transformation of explanatory variables	AIC*	Charvet et al. (2015); Leelawat (2014); Muhari (2015); Reese (2011)
		Link function	AIC	
		Ordered or partially ordered models	LRT	
Semi-parametric	Generalised Additive Model (GAM)	Transformation of explanatory variables	AIC*	Wood (2006)
		Link function	AIC	
		Number of knots	KFCV**	
Non-parametric	Kernal Smoother	(See reference for information on fitting these models)		Noh et al. (2014)

*It is noted that fragility functions are generally fit to the natural logarithm of the explanatory variable .
**If conducting trend analysis using GAMs it is recommended to simply select a preliminary number of knots (e.g. 4 knots).

Table 4: Statistical models considered for TIM comparison. AIC = Akaike Information Criteria (Rossetto et al., 2014), LRT = Likelihood Ratio Test (Rossetto et al., 2014), KFCV = K-Fold Cross-Validation (introduced below).

The vast majority of existing fragility curves use Ordinary Least Squares (OLS) parameter estimation to fit Normal or Lognormal Cumulative Distribution Functions (CDFs) to aggregated model data [1]. However, [20] quantitatively shows that OLS regression is inappropriate for fragility function derivation as several of the linear model assumptions are violated by the data and the aggregation of data results in information being lost, leading to reduced predictive accuracy and increased uncertainty by an amount which is dependent on the aggregation approach. For these reasons OLS models may not be used in the methodology proposed in this paper.

Generalised Linear Models (GLMs) allow for a relaxation of some of the linear model assumptions by relating the mean of a response variable ($E(y_i)=\mu_i$) to the explanatory variables (x_i) via an arbitrary link function (g) (where the subscript i refers to the i^{th} building) [16]. The link function is selected dependent on the distribution of the response variable, typically transforming the response such that g(μ) is a continuous variable bounded by $[-\infty,+\infty]$ (generally probit, logit or cloglog link functions are used for fragility functions). As such, GLMs can be used for variables with distributions other than the Gaussian distribution assumed in OLS linear regression models, and so where damage data is either binary (damaged/not damage) or ordinal (falling into one of several discrete damage states) GLMs are recommended as an improvement over OLS for deriving fragility curves [16], [31].

In Step 2 further exploratory analysis of the available database can be conducted by fitting ordinal cumulative link models to the data. Fragility curves corresponding to each damage state (DS1-DS5*) are determined by assigning a damage response indicator, *ds*, to each building, which is considered to follow a multinomial distribution. Each building is also assigned a TIM value, x_j. The main advantage of this model over separate GLMs fitted to binary data, is its ability to use all available information regarding the data in the database, it recognises that the damage is an ordinal categorical variable and accounts for the main conclusions of the exploratory analysis [32]. The optimal link function can be selected by comparing AIC statistics, as in [16]. The model equations are given in Table 5 for a probit link function (the inverse standard cumulative normal distribution), where β_0 and β_1 are the unknown regression parameters (the intercept and slope, respectively) estimated by a maximum likelihood optimisation algorithm. Uncertainty may be quantified using bootstrap methods [16]. Multinomial data can be assessed using either partially-ordered or ordered models. For ordered models the slope parameters (β_1 in Table 5) are assumed to be equal for all damage states so as to avoid undesirable effects such as the crossing of curves. Partially-ordered models relax this assumption. The decision of whether to use ordered or partially-ordered models can me made via the Likelihood Ratio Test [31].

Random Component	$ds = \{0,1,2,3,4,5^{*}\}, \quad ds \mid x_j \sim Multinomial\left(P\left(ds = DS_i \mid TIM = x_j\right)\right)$ Where, $P\left(ds = DS_i \mid TIM = x_j\right) = \begin{cases} 1 - P\left(ds \geq DS_i \mid x_j\right) & i = 0 \\ P\left(ds \geq DS_i \mid x_j\right) - P\left(ds \geq DS_{i+1} \mid x_j\right) & 0 < i < N_{DS} \\ P\left(ds \geq DS_i \mid x_j\right) & i = N_{DS} \end{cases}$
Systematic Component	$probit\left(P\left(ds \geq DS_i \mid TIM = x_j\right)\right) = \beta_{0,i} + \beta_{1,i} x_j$
Parameter Estimation	where $\beta_{0,i}$, $\beta_{1,i}$ are estimated via Maximum Likelihood

Table 5: Components of partially-ordered Cumulative Link Models (CLM) with probit link function.

Generalised Additive Models (GAMs) [29] are semi-parametric models that fit GLMs in a piecewise regression system with a number of separation points (or knots). Whilst there are dangers in using non-parametric and semi-parametric methods for prediction purposes, they are suitable for comparing the influence of different explanatory variables (TIMs) to describe response variable observations. However, an issue with non- and semi-parametric models is that they are susceptible to over-fitting, and their appropriateness in the context of fragility analysis has not yet been demonstrated. There are methods for overcoming overfitting (demonstrated in Step 3), and so GAMs are recommended (alongside CLMs) to conduct the TIM comparison of Step 3.

Case Study: Further Exploratory Analysis and Model Selection

In this stage several statistical model types and model configurations are investigated and the models used for the TIM comparison of Step 3 are selected. It is noted that the TIM used in this section is the observed inundation depth reported in the MLIT database, and therefore this investigation is independent of the inundation simulation.

Curves are constructed for engineered and non-engineered building categories and the influence of these construction material groups is examined by fitting the cumulative link model shown in Table 5 to the data corresponding to the two material groups. The confidence intervals are quantified using bootstrap methods employed by [16] based on 1,000 iterations. Figure 5 shows that fragility curves for engineered and non-engineered buildings differ in both slopes and intercepts, and so it is appropriate to consider these material groups separately. Consequently, the TIM comparison of Step 3 is conducted for each material group separately, and results are compared.

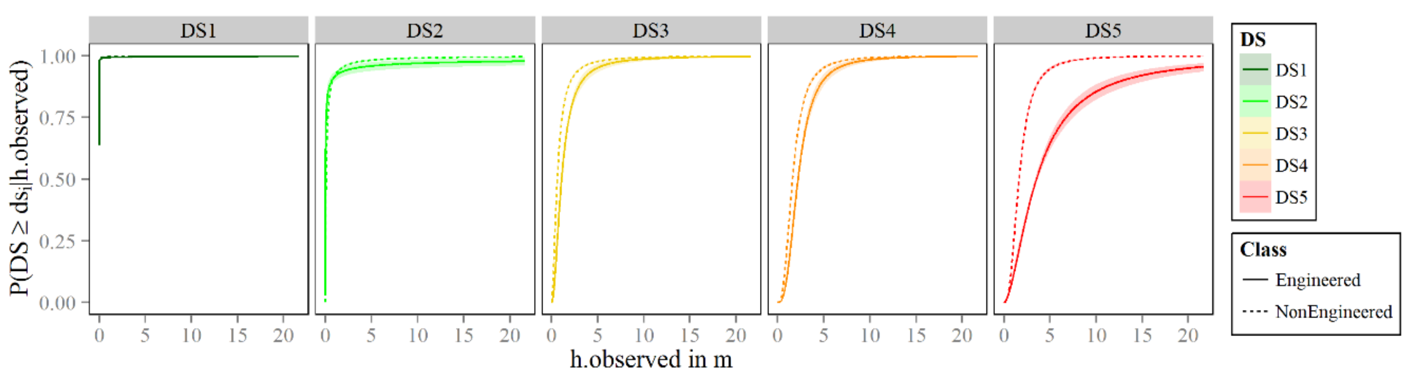


Figure 5: Comparison of fragility curves for engineered and non-engineered material groups, for each damage state, formed on disaggregated data.

The more complex model (partially-ordered model, M1.1) will always fit the data as well as or better than the simpler model (M1.2). The LRT results given in Table 6 confirm that there is less than a 1% chance that the improvement of fit for the more complex model could be observed by random chance, and therefore the partially ordered model is to be used for the TIM comparison in Step 3.

Model	logLikelihood	Likelihood Ratio Statistic	P (χ^2)
Partially Ordered Model	-4964.4	1734	<2.2e-16 ***
Ordered Model	-5831.4		

Table 6: Likelihood Ratio Test results for ordered and partially ordered models, showing that the partially ordered model provides a significantly better fit than the ordered.

As GAMs are a piecewise system of GLMs, and as overfitting can be avoided using cross-validation sub-sensitivity analysis (demonstrated in Section **Error! Reference source not found.**) GAMs are also selected (alongside CLM model M1.1) to conduct the TIM comparison of Step 3.

Step 3. Comparison of Tsunami Intensity Measures

The optimum TIM is defined by fitting the models selected in Step 2 to the completed dataset and assessing the goodness of fit for each TIM. Goodness of fit measures based on examination of residuals (e.g. R^2, LRT and AIC) are biased by overfitting, indicating a better fit to the underlying population than is really the case, and cannot be used to directly compare cumulative models with separate models, nor to compare models formed on aggregated and disaggregated data. Cross-validation techniques overcome these issues and so are recommended here.

Cross-validation is an improvement over simply plotting the residuals, as it attempts to indicate the prediction error (i.e. the proportion of incorrectly classified outcomes) that would be experienced on data that has not been used to form the statistical model. *K*-fold cross-validation creates *K*-fold partitions in the total dataset, and for each of *K* validation experiments uses one fold as the testing set (a different one each time), and the remaining data as the training set. The average of the error rates for all iterations gives an estimate of the true prediction error rate. Cross validation has been used to estimate tsunami fragility curve prediction error rates by [26] and [2], who also propose a penalized error estimation method for multinomial models that is used in this study (shown in (2)). In (2), N_{DS} refers to the number of damage states (6 in this case, including DS0), and the predicted damage state ($ds_{j,predicted}$) for the j^{th} observation is taken as the damage state that has the greatest probability of occurrence. The optimum TIM for a given data set is that for which the fit models provide the lowest error rate.

$$Error\ rate(multinomial) = \frac{1}{K}\sum_{k=1}^{K}\left[\sum_{j}^{N_{test\ set}} \frac{\left|ds_{j,predicted} - ds_{j,observed}\right|}{N_{DS}-1}\right] \quad (2)$$

where $ds_{j,predicted} = \mathrm{argmax}_{DS_i \in \{DS_0 : DS_6\}} P\left(ds = DS_i \middle| TIM = x_j\right)$

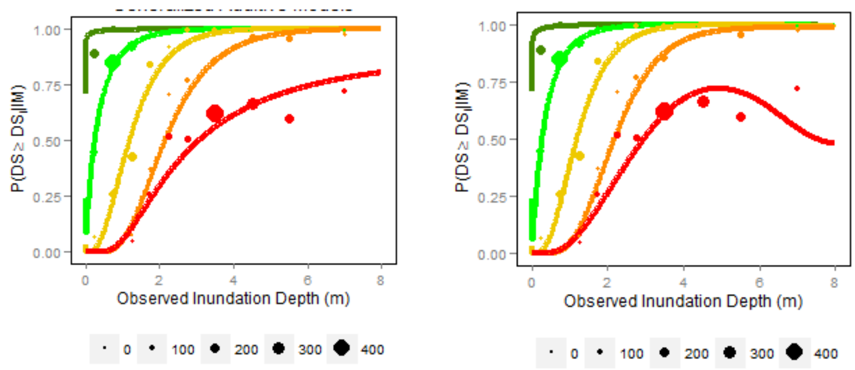
Case Study: Semi-Parametric Model Optimization and Intensity Measure Comparison

This section compares several Tsunami Intensity Measures (TIMs) in their ability to describe the observed damage data. First it is demonstrated how cross-validation can be used to define the number of knots to be used for GAMs fit to the data each TIM. Partially-ordered probit models (model M1.1) are fit to the disaggregated data of the MLIT building damage database for each of the TIMs identified in Table 3, and their relative fits are compared using prediction error rates estimated via 10-fold cross-validation. The same procedure is then conducted using Generalised Additive Models (GAMs), with the number of knots for each model selected using the sub-sensitivity analysis shown below. Finally, the TIMs are ranked by their predictive error rates for both the CLM and GAM model groups. Note that for all models, the penalized error rate is repeatedly estimated until the difference between the running average and that of the previous iteration reduces to below 10^{-5}.

Cross-validation techniques are less biased by overfitting than techniques that simply consider residuals, and so comparison of the cross-validation error rates can be used to select the optimal non-parametric or semi-parametric model (e.g. to select the number of knots when using GAMs) for each TIM. For this dataset, Figure 6 shows that for a series of GAMs fit to observed inundation depth, the model using 4 knots provides the lowest error rate and so provides the optimal fit over GAMs with more knots, which exhibit signs of overfitting. Note that this sub-sensitivity is repeated so as to identify the optimal GAM model for each TIM in turn.

Knots	1	2	3	4	5	6	7	8
Error Rate	10.48%	10.47%	10.48%	10.35%	10.39%	10.40%	10.62%	10.52%

Table 7: 10-fold cross-validation error rates for GAMs fit to observed inundation depth (TIM1) over a range of knots. The best and worst models are shown in Figure 6.



(a) **(b)**

Figure 6: Comparison of fragility curves for Generalised Additive Models (probit link function) fit to observed inundation depth (TIM1) with 4 and 7 knots, showing optimal and over-fit curves respectively. Note that aggregated datapoints are shown for graphical reference, but have not been directly used in the regression analysis, which has been conducted on the imputed disaggregated dataset.

Table 8 compares the prediction error rates for CLMs and GAMs fit to each additional TIM for engineered buildings. For engineered buildings the quasi-steady force estimation (F_{QS}) and simulated inundation depth (h_{sim}) appear to give the best fit. The fact that the results for CLMs and GAMs are similar suggest that the results are not model-specific.

	Alternative Intensity Measures						Optimal IM	
	TIM2	TIM3	TIM4	TIM5	TIM6	TIM7	1st	2nd
	h_{sim}	v	MF	MF_{equiv}	Fr	F_{QS}		
CLMs	16.0%	22.9%	17.3%	16.2%	27.5%	15.3%	F_{QS}	h_{sim}
GAMs	13.4%	19.9%	16.6%	15.7%	24.3%	14.1%	h_{sim}	F_{QS}

Table 8: Engineered Buildings: Comparison of prediction error rates for partially ordered cumulative link models. The colour scale indicates the goodness of fit, with the lowest error rates (indicating the best fit) shown in green.

Velocity and Froude number alone are consistently the worst TIMs. However, F_{QS} (a function of h, v and Fr) generally performs better than the traditional force measure of momentum flux (a function of h and v only). This implies that the construction of future empirical and analytical fragility functions based on the GEJE dataset are that force should be used as a TIM, where either force accounts for the flow regime (for 2D curves) or an indicator of the flow regime (e.g. Froude Number) should be investigated as an additional TIM (for fragility surfaces).

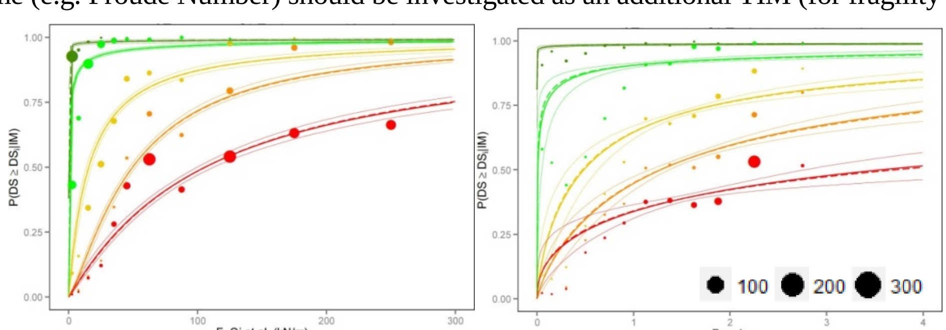


Figure 7: The derived fragility functions (partially-ordered CLMs with probit link functions fitted to ln|TIM|) for engineered buildings for the best (left) and worst (right) performing TIMs (F_{QS} and Fr, respectively), showing narrower confidence intervals for the better performing TIM (F_{QS}).

Summary and Conclusion

This paper has collated, compared and expanded on the current state-of-the-art methodologies for tsunami fragility assessment, in order to present a rigorous methodology for the selection of the optimal TIM for empirical fragility function derivation for any given dataset. This methodology is demonstrated using a unique, detailed, disaggregated damage dataset from the 2011 Great East Japan Earthquake and Tsunami. Exploratory analysis is conducted on a detailed, disaggregated building damage dataset, unique in the fields of both tsunami and seismic fragility assessment.

Buildings of unknown construction material present a significant proportion of the case study dataset (18.2%) and so in order to avoid the introduction of bias when producing fragility curves by material, missing material data is estimated using multiple imputation techniques. The first stage of fragility assessment consists of a sensitivity analysis of several statistical methods for fragility curve derivation, so as to select at least two statistical models with which to conduct the TIM comparison. General conclusions are drawn regarding the suitability of various models and the methods used to select between them, with Cumulative Link Models and Generalised Additive Models selected for the TIM comparison. Partially-ordered probit models are derived for several TIMs and their 10-fold cross-validation results are compared. The same procedure was repeated using GAMs to show that the results are not model-specific. It is shown that both inundation depth and a quasi-steady force estimation which differentiates between flow regimes (F_{QS}) consistently provide the best fit to the observed damage.

Several advanced statistical methods novel in the field of tsunami fragility assessment were introduced: missing data classification and Multiple Imputation techniques for missing data estimation, semi-parametric models (Generalised Additive Models), and K-fold cross-validation for model comparison. The main conclusions can be summarized as follows:

Exploratory Analysis

1. <u>Missing data</u> can only be removed if it can be shown to be Missing Completely At Random (shown not to be not the case for the 2011 MLIT Japan data).
2. <u>Multiple Imputation</u> (MI) has been shown to be an acceptable method for estimating missing data, and is recommended for use on future fragility studies where data cannot be shown to be Missing Completely At Random.

Statistical Modelling

3. <u>*K*-fold cross-validation</u> (KFCV) is shown to be a suitable method for comparing model fits for various model types, and the methodology for conducting this for multinomial models is demonstrated. It is recommended that KFCV be used for evaluation of model fits in future fragility studies.
4. <u>Semi-parametric methods</u> are seen to be suitable for comparative fragility assessments, and the issue of over-fitting can be avoided through the use of cross-validation techniques, as demonstrated.

The methodology presented in this paper has application for researchers and risk modellers in the engineering, DRR and insurance industries; and the statistical methods presented have implications for fragility function derivation and selection for multiple hazards.

Acknowledgements

The first author was funded by EPSRC Engineering Doctorate Programme and the Willis Research Network, and travel to conduct this collaborative work was kindly funded by the Sasakawa Foundation and the Earthquake Engineering Field Investigation Team travel grant. Co-author Ioanna Ioannou was funded by the EPSRC Challenging Risk grant. We would like to thank Dr Ioannis Kosmidis and Prof Richard Chandler of the Department of Statistical Science, UCL, for advising on the statistical analysis conducted in this study. We would also like to acknowledge the many years of successful collaboration between EPICentre, UCL (UK) and IRIDeS, Tohoku University (Japan) which has made this work possible.

References

[1] S. Koshimura, T. Oie, H. Yanagisawa, and F. Imamura, "Developing fragility functions for tsunami damage estimation using numerical model and post-tsunami data from banda aceh, indonesia," *Coast. Eng. Journal, , Japan Soc. Civ. Eng.*, vol. 51, no. 3, pp. 243–273, 2009.

[2] I. Charvet, A. Suppasri, H. Kimura, D. Sugawara, and F. Imamura, "Fragility estimations for Kesennuma City following the 2011 Great East Japan Tsunami based on maximum flow depths,velocities and debris impact, with evaluation of the ordinal model's predictive accuracy," *Nat. Hazards*, vol. 79, no. 3, pp. 2073–2099, 2015.

[3] N. Tanaka and K. Kondo, "Fragility curve of region of wooden building washout due to tsunami based on hydrodynamic characteristics of the Great East Japan Tsunami," *Ocean Eng.*, 2015.

[4] H. Kreibich, K. Piroth, I. Seifert, H. Maiwald, U. Kunert, J. Schwarz, B. Merz, and a. H. Thieken, "Is flow velocity a significant parameter in flood damage modelling?," *Nat. Hazards Earth Syst. Sci.*, vol. 9, no. 5, pp. 1679–1692, Oct. 2009.

[5] MLIT, "Concerning Setting the Safe Structure Method for Tsunamis which are Presumed when Tsunami Inundation Occurs - Note 1318," 2011.

[6] I. N. Robertson and H. R. Riggs, "OMAE2011-49487 Tsunami Bore Forces on Walls," in *Proceedings of the ASTM 2011 30th Internal Conference on Ocean, Offshore and Arctic Engineering*, 2011.

[7] Z. X. Qi, I. Eames, and E. R. Johnson, "Force acting on a square cylinder fixed in a free-surface channel flow," *J. Fluid Mech.*, vol. 756, pp. 716–727, Sep. 2014.

[8] H. Park, D. Wiebe, D. Cox, and K. Cox, "TSUNAMI INUNDATION MODELING: SENSITIVITY OF VELOCITY AND MOMENTUM FLUX TO BOTTOM FRICTION WITH APPLICATION TO BUILDING DAMAGE AT SEASIDE, OREGON," *Coast. Eng.*, pp. 1–12, 2014.

[9] N. Luco and C. A. Cornell, "Structure-Specific Scalar Intensity Measures for Near-Source and Ordinary Earthquake Ground Motions," *Earthq. Spectra*, vol. 23, no. 2, pp. 357–392, 2007.

[10] H. Ebrahimian, F. Jalayer, A. Lucchini, F. Mollaioli, and G. Manfredi, "Preliminary ranking of alternative scalar and vector intensity measures of ground shaking," *Bull. Earthq. Eng.*, Apr. 2015.

[11] P. Giovenale, C. A. Cornell, and L. Esteva, "Comparing the adequacy of alternative ground motion intensity measures for the estimation of structural responses," *Earthq. Eng. Struct. Dyn.*, vol. 33, no. 8, pp. 951–979, Jul. 2004.

[12] J. Macabuag, T. Rossetto, I. Ioannou, A. Suppasri, D. Sugawara, B. Adriano, F. Imamura, I. Eames, and S. Koshimura, "A Proposed Methodology for Deriving Tsunami Fragility Functions for Buildings using Optimum Intensity Measures," *Nat. Hazards*, pp. 1–25.

[13] A. Suppasri, I. Charvet, J. Macabuag, T. Rossetto, N. Leelawat, P. Latcharote, and F. Imamura, "Building Damage Assessment and Implications for Future Tsunami Fragility Estimations," in *Handbook of Coastal Disaster Mitigation for Engineers and Planners*, Elsevier, Ed. 2015, pp. 147–178.

[14] A. Suppasri, E. Mas, I. Charvet, R. Gunasekera, K. Imai, Y. Fukutani, Y. Abe, and F. Imamura, "Building damage characteristics based on surveyed data and fragility curves of the 2011 Great East Japan tsunami," *Nat. Hazards*, vol. 66, no. 2, pp. 319–341, Nov. 2013.

[15] J. H. Ware, D. Harrington, D. J. Hunter, and R. B. D'Agostino, "Missing Data," *N. Engl. J. Med.*, vol. 367, no. 14, pp. 1353–1354, Oct. 2012.

[16] I. Charvet, I. Ioannou, T. Rossetto, A. Suppasri, and F. Imamura, "Empirical fragility assessment of buildings affected by the 2011 Great East Japan tsunami using improved statistical models," *Nat. Hazards*, Mar. 2014.

[17] Japan Cabinet Office, "Residential Disaster Damage Accreditation Criteria Operational Guideline.pdf," 2013.

[18] A. Suppasri, I. Charvet, I. Kentaro, and F. Imamura, "Fragility curves based on data from the 2011 Great East Japan tsunami in Ishinomaki city with discussion of parameters influencing building damage," *Earthq. Spectra*, vol. 31, no. 2, pp. 841–868, 2014.

[19] B. Adriano and S. Koshimura, "Understanding the extreme tsunami inundation in onagawa town by the 2011 tohoku earthquake, its effects in urban structures and coastal facilities," *Coast. Eng. J.*, 2016.

[20] J. Macabuag, T. Rossetto, I. Ioannou, A. Suppasri, D. Sugawara, B. Adriano, F. Imamura, I. Eames, and S. Koshimura, "Investigation of Optimum Intensity Measures and Advanced Statistical Methods for Constructing Tsunami Fragility Curves for Buildings," *Nat. Hazards*, vol. (tbc), 2016.

[21] K. Satake, Y. Fujii, T. Harada, and Y. Namegaya, "Time and Space Distribution of Coseismic Slip of the 2011 Tohoku Earthquake as Inferred from Tsunami Waveform Data," *Bull. Seismol. Soc. Am.*, vol. 103, no. 2B, pp. 1473–1492, May 2013.

[22] F. Imamura, E. Gica, T. Takahashi, and N. Shuto, "Numerical Simulation of the 1992 Flores Tsunami: Interpretation of Tsunami Phenomena in Northeastern Flores Island and Damage at Babi Island," *Pure Appl. Geophys.*, vol. 144, pp. p555–568, 1995.

[23] A. Suppasri, E. Mas, S. Koshimura, K. Imai, K. Harada, and F. Imamura, "Developing Tsunami Fragility Curves From the Surveyed Data of the 2011 Great East Japan Tsunami in Sendai and Ishinomaki Plains," *Coast. Eng. J.*, vol. 54, no. 01, pp. 1250008–1, 2012.

[24] A. Suppasri, S. Koshimura, and F. Imamura, "Tsunami fragility curves and structural performance of building along the Thailand coast," no. 2008, pp. 3–8, 2009.

[25] A. Suppasri, E. Mas, S. Koshimura, K. Imai, K. Harada, and F. Imamura, "Developing Tsunami Fragility Curves From the Surveyed Data of the 2011 Great East Japan Tsunami in Sendai and Ishinomaki Plains," *Coast. Eng. J.*, vol. 54, no. 01, pp. 1250008–1, 2012.

[26] A. Muhari, I. Charvet, F. Tsuyoshi, A. Suppasri, and F. Imamura, "Assessment of tsunami hazards in ports and their impact on marine vessels derived from tsunami models and the observed damage data," *Nat. Hazards*, Apr. 2015.

[27] N. Leelawat, A. Suppasri, I. Charvet, and F. Imamura, "Building damage from the 2011 Great East Japan tsunami: quantitative assessment of influential factors," *Nat. Hazards*, vol. 73, no. 2, pp. 449–471, Feb. 2014.

[28] S. Reese, B. a. Bradley, J. Bind, G. Smart, W. Power, and J. Sturman, "Empirical building fragilities from observed damage in the 2009 South Pacific tsunami," *Earth-Science Rev.*, vol. 107, no. 1–2, pp. 156–173, Jul. 2011.

[29] S. Wood, *Generalized Additive Models: An Introduction with R*. CRC Press, London, 2006.

[30] H. Y. Noh, D. Lallemant, and A. S. Kiremidjian, "Development of empirical and analytical fragility functions using kernel smoothing methods," *Earthq. Eng. Struct. Dyn.*, 2014.

[31] T. Rossetto, I. Ioannou, D. N. Grant, and T. Maqsood, "Guidelines for Empirical Vulnerability Assessment Report produced in the context of the Vulnerability Global Component project," 2014.

[32] I. Charvet, I. Ioannou, T. Rossetto, a. Suppasri, and F. Imamura, "Empirical fragility assessment of buildings affected by the 2011 Great East Japan tsunami using improved statistical models," *Nat. Hazards*, vol. 73, no. 2, pp. 951–973, Mar. 2014.