https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019

Natural Hazards and Earth System Sciences
Discussions

Non-stationary extreme value analysis applied to seismic fragility assessment for nuclear safety analysis

Jeremy Rohmer[1], Pierre Gehl[1], Marine Marcilhac-Fradin[2], Yves Guigueno[2], Nadia Rahni[2], Julien Clément[2]

[1]BRGM, 3 av. C. Guillemin, 45060 Orléans Cedex 2, France
[2]Institute for Radiological Protection and Nuclear Safety, Fontenay-Aux-Roses, 92262, France

Correspondence to: Jeremy Rohmer (j.rohmer@brgm.fr)

Abstract. Fragility curves (FC) are key tools for seismic probabilistic safety assessments that are performed at the level of the nuclear power plant (NPP). These statistical methods relate the probabilistic seismic hazard loading at the given site and the required performance of the NPP safety functions. In the present study, we investigate how the tools of non-stationary extreme value analysis can be used to model in a flexible manner the tail behaviour of the engineering demand parameter as a function of the considered intensity measure. We focus the analysis on the dynamic response of an anchored steam line and of a supporting structure under seismic solicitations. The failure criterion is linked to the exceedance of the maximum equivalent stress at a given location of the steam line. A series of three-component ground-motion records (~300) were applied at the base of the model to perform non-linear time history analyses. The set of numerical results was then used to derive a FC, which relates the failure probability to the variation of peak ground acceleration (*PGA*). The probabilistic model of the FC is selected via information criteria completed by diagnostics on the residuals, which support the choice of the generalized extreme value GEV distribution (instead of the widely used log-normal model). The GEV distribution is here non-stationary and the relationships of the GEV parameters (location, scale and shape) are established with respect to *PGA* using smooth non-linear models. The procedure is data-driven, which avoids the introduction of any a priori assumption on the shape/form of these relationships. To account for the uncertainties in the mechanical and geometrical parameters of the structures (elastic stiffness, damping, pipeline thicknesses, etc.), the FC is further constructed by integrating these uncertain parameters. A penalisation procedure is proposed to set to zero the variables of little influence in the smooth non-linear models. This enables us to outline which of these parametric uncertainties have negligible influence on the failure probability as well as the nature of the influence (linear, non-linear, decreasing, increasing, etc.) with respect to each of the GEV parameters.

1 Introduction

A crucial step of any seismic Probability Risk Assessment (PRA) is the vulnerability analysis of structures, systems and components (SSC) with respect to the external loading induced by earthquakes. To this end, fragility curves (FC), which relate

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019


Natural Hazards and Earth System Sciences Discussions

the probability of an SSC to exceed a predefined damage state as a function of an intensity measure (*IM*) representing the hazard loading, are common tools. Formally, FC expresses the conditional probability $P_f(im) = P(EDP \geq th|IM = im)$ with respect to the *IM* value *im* and to the *EDP* engineering demand parameter obtained from structural analysis (e.g. force, displacement, drift ratio); *th* being an acceptable demand threshold.

FCs are applied on a large variety of different structures like residential buildings (e.g. Gehl et al., 2013), nuclear power plant (Zentner et al., 2017), wind turbines (Quilligan et al., 2012), underground structures (Argyroudis and Pitilakis, 2012), etc. Their probabilistic nature make them well suited for PRA applications, at the interface between probabilistic hazard assessments and event tree analyses, in order to estimate the occurrence rate of undesirable top events.

Different procedures exist to derive FCs (see e.g. an overview by Zentner et al., 2017). In the present study, we focus on the analytical approach, which aims at deriving a parametric cumulative distribution function (CDF) from data collected from numerical structural analyses. A common assumption in the literature is that the logarithm of *im* is normally distributed (e.g., Ellingwood, 2001) as follows:

$$P_f(im) = \Phi\left(\frac{\log(im) - \log(\alpha)}{\beta}\right), \quad (1)$$

where Φ is the standard normal cumulative distribution function, α is the median and β is lognormal standard deviation. The parameters of the normal distribution are commonly estimated either by maximum likelihood estimation (see e.g., Shinozuka et al., 2000) or by fitting a linear probabilistic seismic demand model in the log-scale (e.g., Banerjee and Shinozuka, 2008).

This procedure faces, however, limits in practice:

- *Limit (1)*: the assumption of normality may not always be valid in all situations as discussed by Mai et al. (2017) and Zentner et al. (2017). This widely-used assumption is especially difficult to justify when the considered EDP corresponds to the maximum value of the variable of interest (for instance maximum transient stress value), i.e. when the FC serves to model extreme values;
- *Limit (2)*: a second commonly-used assumption is the homoscedasticity of the underlying probabilistic model, i.e. the variance term β is generally assumed to be constant over the domain of the *IM*;
- *Limit (3)*: the assumption of linearity regarding the relation between the median and *IM* may not always hold valid as shown for instance by Wang et al. (2018) using artificial neural networks;
- *Limit (4)*: a large number of factors may affect the estimate of P_f in addition to *IM*; for instance epistemic uncertainties due to the identification/characterization of some mechanical (e.g., elastic stiffness, damping ratio, etc.) and geometrical parameters of the considered structure. This means that the failure probability is conditional on the vector of uncertain mechanical and geometrical factors $\boldsymbol{U}$ (in addition to *IM*), namely:

$$P_f(im, \boldsymbol{u}) = P(EDP \geq th|IM = im, \boldsymbol{U} = \boldsymbol{u}), \quad (2)$$

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019


Natural Hazards and Earth System Sciences Discussions

The current study aims at going a step forward in the development of seismic FCs by improving the procedure regarding the afore-mentioned limits. To deal with limit (1), we propose to rely on the tools of extreme value statistics (Coles, 2001) and more specifically on the Generalised Extreme Value (GEV) distribution, which can model different extremes' behaviour. Note that the focus is on the extremes related to EDP, not on the forcing, i.e. the analysis does not model the extremes of *IM* as it is done for current practices of probabilistic seismic hazard analysis (see e.g., Dutfoy, 2019). This means that no preliminary screening is applied, which implies that the FC derivation is conducted by considering both large and intermediate earthquakes, i.e. small-to-moderate to large *IM* values.

The use of GEV is examined using criteria for model selection like Akaike or Bayesian Information Criteria (Akaike, 1998; Schwarz, 1978). Limits (2) and (3) are addressed using tools for distributional regression (e.g., Koenker et al., 2013) within the general framework of Generalized Additive Model for Location, Scale and Shape parameter (GAMLSS; e.g., Rigby and Stasinopoulos, 2005). GAMLSS is very flexible in the sense that the mathematical relation of the median and variance in Eq. 1 can be learnt from the data via nonlinear smooth functions. GAMLSS can be applied to any parametric probabilistic model, and here to the GEV model as a particular case. This enables us to fit a non-stationary GEV model, i.e. a GEV model for which the parameters vary as a function of some covariates (here corresponding to *IM* and $\boldsymbol{U}$). The use of data-driven nonlinear smooth functions avoids introducing a priori model like linear or polynomial as many authors do (see an example by for sea level extremes by Wong (2018), and for temperature by Cheng et al., 2014). Finally, accounting for the epistemic uncertainties in the fitting process (limit (4)) raises the question of integrating a potentially large number of variables, which might hamper the stability and quality of the procedure for FC construction. This is handled with a penalisation procedure (Marra and Wood, 2011), which enables the analysist to screen the parametric uncertainties of negligible influence.

The paper is organised as follows. Section 2 describes the statistical methods to derive non-stationary GEV-based seismic fragility curves. Then, in Section 3, we describe a test-case related to the seismic fragility assessment for a steam line of a nuclear power plant. For this case, the derivation of FC is performed by considering the widely-used *IM* in the domain of seismic engineering, namely Peak Ground Acceleration (*PGA*). Finally, Section 4 applies the proposed procedure and discusses the results.

2 Statistical methods

In this section, we first describe the main steps of the proposed procedure for deriving the FC (Sect. 2.1). The subsequent sections provide technical details on the GEV probability model (Sect. 2.2), its non-stationary formulation and implementation (Sect. 2.3) within the GAMLSS framework and its combination with variable selection (Sect. 2.4).

2.1 Overall procedure

To derive the seismic FC, the following overall procedure is proposed:

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019


Natural Hazards and Earth System Sciences Discussions

- Step 1 consists in analysing the validity of using the GEV distribution with respect to alternative probabilistic models (and the normal distribution of Eq. 1 in particular). This can be done based on the criteria for model selection like Akaike or Bayesian Information Criteria (Akaike, 1998; Schwarz, 1978), respectively denoted AIC and BIC. Since the constructed models use penalisation for the smoothness, we use the formulation provided by Wood et al. (2016: Sect. 5) to account for the smoothing parameter uncertainty;
- Depending on the results of step 1, step 2 aims at fitting the non-stationary GEV model using the double penalisation formulation described in Sect. 2.2 and 2.3;
- Step 3 aims at producing some diagnostic information about the fitting procedure and results. The first diagnostic test uses the QQ plot of the model deviance residuals (conditional on the fitted model coefficients and scale parameter) formulated by Augustin et al. (2012). If the model distributional assumptions are met then the QQ plot should be close to a straight line. The second diagnostic test relies on a transformation of the data to a Gumbel distributed random variable (e.g. Beirlant et al., 2004) and on an analysis of the corresponding Gumbel QQ plot;
- Step 4 aims at analysing the partial effect of each input variable (i.e. the smooth non-linear term, see Eq. 4 in Sect. 2.3) to assess the influence of the different GEV parameters;
- Step 5 aims at deriving the seismic FC by evaluating the failure probability $P_f(im, \boldsymbol{u}) = P(EDP \geq th | IM = im, \boldsymbol{U} = \boldsymbol{u})$.

For step 5, the following procedure is conducted to account for the mechanical and geometrical uncertainties:

- Step 5.1: the considered *IM* is fixed at a given value;
- Step 5.2: for the considered *IM* value, a large number (here chosen at n=1000) of mechanical and geometrical parameters are randomly and uniformly sampled within their respective bounds of variation;
- Step 5.3: for each of the randomly selected parameters and for the considered *IM* value, the failure probability is estimated;
- Return to step 5.1.

The procedure of Step 5 provides a set of n FCs from which we can derive the median FC as well as the uncertainty bands based on the pointwise confidence intervals at different levels.

2.2 GEV distribution

The cumulative distribution function (CDF) of the Generalized Extreme Value (GEV) probability model holds as follows:

$$\mathrm{P}(\mathrm{EDP} \leq \mathrm{edp}) = \exp\left(-\left(1 + \xi\left(\frac{\mathrm{edp}-\mu}{\sigma}\right)\right)^{-1/\xi}\right), \quad (3)$$

where *edp* is the variable of interest; μ, σ and ξ are the GEV location, scale, and shape parameters, respectively. Depending on the value of the shape parameter, the GEV distribution presents an asymptotic horizontal behaviour for ξ<0 (i.e. the

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019

Natural Hazards and Earth System Sciences
Discussions

asymptotically–bounded distribution, which corresponds to the Weibull distribution); unbounded when ξ>0 (i.e. high probability of occurrence of great values can be reached, which corresponds to the Fréchet distribution); and intermediate in the case of ξ=0 (Gumbel distribution).

Fig. 1a illustrates the behaviour of the GEV density distribution for μ=12.5, σ=0.25 and different ξ values: the higher ξ, the heavier the tail. Fig. 1b,c further illustrates how changes in the other parameters (respectively the location and the scale) affect the density distribution. The location primarily translates the whole density distribution, while the scale affects the tail and to a lesser extent (for the considered case) the mode.

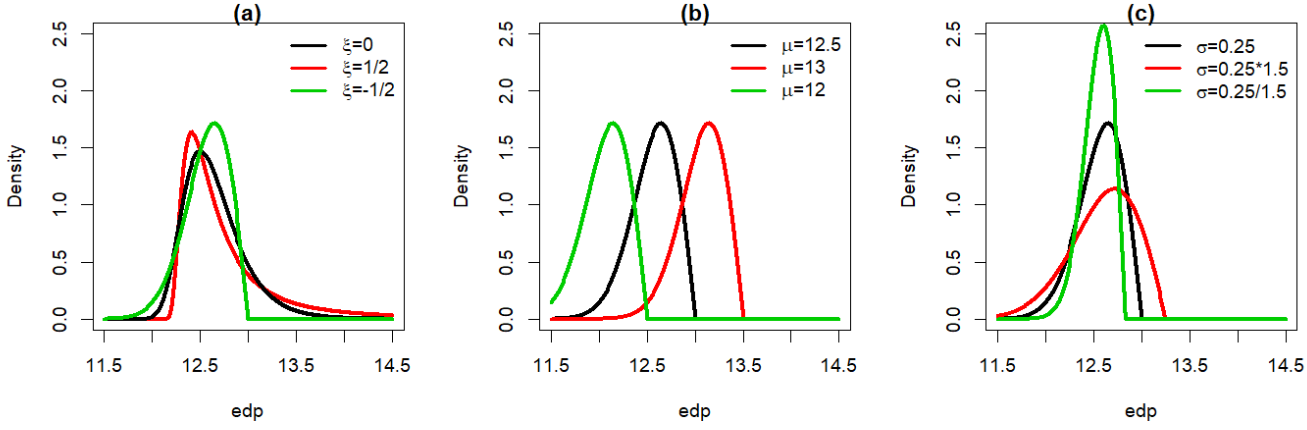


Figure 1: Behaviour of the GEV density distributions depending on the changes in the parameter value: (a) ξ (with μ fixed at 12.5, and σ fixed at 0.25); (b) μ (with ξ fixed at 0.5, and σ fixed at 0.25); (c) σ (with μ fixed at 12.5, and ξ fixed at 0.5).

2.3 Nonstationary GEV

The GEV distribution is assumed to be nonstationary in the sense that the GEV parameters $\boldsymbol{\theta}$=(μ,σ,ξ) vary as a function of **x** the vector of input variables, which include *IM* and the uncertain input variables ***U*** (as described in the introduction). The fitting is performed within the general framework of Generalized Additive Model for Location, Scale and Shape parameter (GAMLSS; e.g., Rigby and Stasinopoulos, 2005). Since the scale parameter satisfies σ>0, we preferably work with its log-transformation, which is denoted lσ. In the following, we assume that $\boldsymbol{\theta}$ follows a semi-parametric additive formulation as follows:

$$\eta_\theta(\boldsymbol{x}) = \sum_{j=1}^{J} f_j(x_j), \quad (4)$$

where J is the number of functional terms that is generally inferior to the number of input variables (see Sect. 2.3), f_j(.) corresponds to a univariate smooth non-linear model described as follows:

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019


Natural Hazards and Earth System Sciences
Discussions

$$f_j(x) = \sum_b \beta_{jb} b_b(x), \quad (5)$$

with $b_b(.)$ the thin plate spline basis function (Wood, 2003) and $\boldsymbol{\beta}_j$ the regression coefficients for the considered smooth function.

These functional terms (termed as partial effect) hold the information of each parameter's individual effect on the considered GEV parameter. The interest is to model the relationship between each GEV parameter and the input variables flexibly. Alternatives approach would assume a priori functional relationships (like linear or of polynomial form), which may not be valid.

The model estimation consists in evaluating the regression coefficients $\boldsymbol{\beta}$ (associated to the GEV parameters $\boldsymbol{\theta}$) by maximizing the log-likelihood $l(.)$ of the GEV distribution. To avoid overfitting, the estimation is based on the penalized version of $l(.)$ to control the roughness of the smooth functional terms (hence their complexity) as follows:

$$\underset{\beta}{\operatorname{argmax}} \left(l(\boldsymbol{\beta}) - \frac{1}{2} \sum_j \lambda_j \boldsymbol{\beta}^T \boldsymbol{S}^j \boldsymbol{\beta} \right), \quad (6)$$

where λ_j controls the extent of the penalisation (i.e. the trade-off between goodness-of-fit and smoothness), and $\boldsymbol{S}^j$ is a matrix of known coefficients (such that the terms in the summation measure the roughness of the smooth functions). Computational methods and implementation details are detailed in (Wood et al., 2016 and references therein).

2.4 Variable selection

The introduction of the penalisation coefficients in Eq. 6 has two effects: they can penalize how "wiggly" a given term is (i.e. it has a smoothing effect) and they can penalize the absolute size of the function (i.e. it has a shrinkage effect). The second effect is of high interest to screen out input variables of negligible influence. However, the penalty can only affect the components that have derivatives, i.e. the set of smooth non-linear functions termed as the "range space". Completely smooth functions (including constant or linear functions), which belong to the "null space" are however not influenced by Eq. 6. For instance, for one-dimensional thin plate regression splines, a linear term might be left in the model, even when the penalty value is very large (as $\lambda \rightarrow \infty$); this means that the afore-described procedure does not ensure that an input variable of negligible influence will completely be filtered out of the analysis (with corresponding regression coefficient shrunk to zero). The consequence is that Eq. 6 does not usually remove a smooth term from the model altogether (Marra and Wood, 2011). To overcome this problem, a double-penalty procedure was proposed by Marra and Wood (2011) based on the idea that the space of a spline basis can be decomposed in the sum of two components, one associated with the functions in the penalty null space and the other with the penalty range space. This double-penalty procedure is adopted in the following.

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019

Natural Hazards and Earth System Sciences
Discussions

3 Application case

This section provides details on the test-case on which the proposed statistical methods (Sect. 2) for the derivation of FCs are demonstrated. The numerical model of the main steam line of a nuclear reactor is described in Sect. 3.1. A set of ground-motion records (Sect. 3.2) is applied to assess the seismic fragility of this essential component of a nuclear power plant.

3.1 Structural model

The 3-D model of a steam line and its supporting structure (i.e., the containment building, see schematic overview in Fig. 2a), previously assembled by Rahni et al. (2017) in the CAST3M finite-element software (Combescure et al., 1982), is introduced here as an application of the seismic fragility analysis of a complex engineered object. The containment building consists of a double-wall structure: the inner wall (reinforced pre-stressed concrete) and the outer wall (reinforced concrete) are modelled with multi-degree-of-freedom stick elements (see Fig. 2b). The steel steam line is modelled by means of beam elements, representing pipe segments and elbows, as well as several valves, supporting devices and stops at different elevations of the supporting structure.

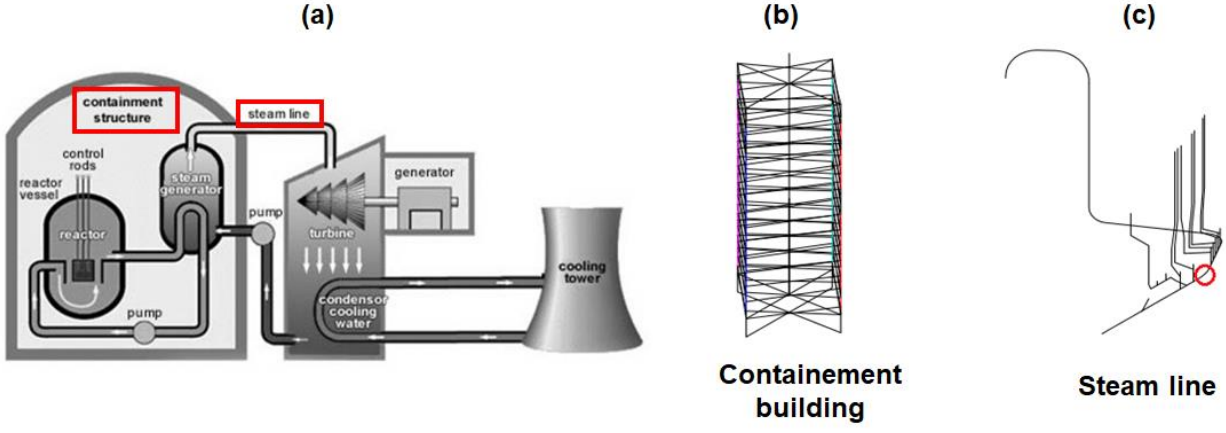


Figure 2: (a) Schematic overview of a nuclear power plant (adapted from nucleus.iaea.org); The red rectangles indicate the main components represented in the structural model. (b) Stick model of the containment building; (c) Steam line beam model, originally built by Rahni et al. (2017). The red circle indicates the location of the vertical stop.

The objective of the fragility analysis is to check the integrity of the steam line: one of the failure criteria identified by Rahni et al. (2017) is the effort calculated at the location corresponding to a vertical stop along the steam line (Fig. 2c). Failure is assumed when the maximum transient effort exceeds the stop's design effort, i.e. $EDP \geq 775$ kN (i.e. 13.56 in log-scale). The model also accounts for epistemic uncertainties due to the identification of some mechanical and geometrical parameters; namely the Young's modulus of the inner containment, the damping ratio of the structural walls and of the steam line, and the thickness of the steam line along various segments of the assembly. The variation range of the ten selected parameters,

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019


Natural Hazards and Earth System Sciences Discussions

constituting the vector $\boldsymbol{U}$ of uncertain factors (see Eq. 2), is detailed in Table 1. A uniform distribution is assumed for these parameters following the values provided by Rahni et al. (2017).

Table 1. Input parameters of the numerical model, according to Rahni et al. (2017).

Variable	Description	Uniform distribution interval
E_{IC}	Young's Modulus – Inner containment	[27700 – 45556] MPa
ξ_{RPC}	Damping ratio – reinforced pre-stressed concrete	[4 – 6] %
ξ_{RC}	Damping ratio – reinforced concrete	[6 – 8] %
e_1	Pipe thickness – Segment #1	[29.8 – 38.3] mm
e_2	Pipe thickness – Segment #2	[33.3 – 42.8] mm
e_3	Pipe thickness – Segment #3	[34.1 – 43.9] mm
e_4	Pipe thickness – Segment #4	[33.3 – 42.8] mm
e_5	Pipe thickness – Segment #5	[53.4 – 68.6] mm
e_6	Pipe thickness – Segment #6	[34.1 – 43.9] mm
ξ_{SL}	Damping ratio – steam line	[1 – 4] %

3.2 Dynamic structural analyses

A series of non-linear time-history analyses are performed on the 3-D model by applying ground-motion records (i.e., acceleration time-histories) at the base of the containment building in the form of a 3-component loading. In the CAST3M software, the response of the building is first computed, and the resulting displacement time-history along the structure is then applied to the steam line model, in order to record the effort demands during the seismic loading.

To this end, natural ground-motion records are selected and scaled using the conditional spectrum method described by Lin et al. (2013). In this manner, the scaling of a set of natural records is carried out to some extent, while preserving the consistency of the associated response spectra. The steps of this procedure hold as follows:

- *Selection of scaling levels*: the acceleration response spectrum is chosen to be conditioned at the period T_1=0.38s, which corresponds to the fundamental mode of the structural model. Six scaling levels are selected and correspond to return periods from 20 years to 20,000 years.
- *Identification of reference earthquakes*: a reference earthquake scenario is identified for each scaling level by means of a probabilistic seismic hazard disaggregation at the studied site. The corresponding target response spectrum is then generated.

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019

Natural Hazards and Earth System Sciences
Discussions

- *Selection of ground-motion records that are spectrum compatible* using the procedure by Jayaram et al. (2011): records are selected from the PEER database[1] (Ancheta et al., 2014), which consist of 30 records for each of the 6 scaling levels; i.e., in total, 180 ground-motion records.

Two distinct cases are considered for the derivation of FCs, depending on whether parametric uncertainties are included in the statistical model or not:

- *Case #1* (*without parametric uncertainties*): A first series of numerical simulations are performed by keeping the mechanical and geometrical parameters fixed at their best estimate values, i.e. the mid-point of the distribution intervals detailed in Table 1. The 180 ground-motion records are applied to the deterministic structural model, resulting in a database of 180 *IM-EDP* points, with *PGA* chosen as the *IM*.
- *Case #2* (*with parametric uncertainties*): A second series of numerical simulations are performed by accounting for parametric uncertainties. This is achieved by randomly varying the values of the mechanical and geometrical input parameters of the numerical model (Table 1) using the Latin Hypercube Sampling technique (Mc Kay et al., 1979). A total number of 360 numerical simulations are performed (using 180 ground-motion records).

Fig. 3 provides the evolution of l*EDP* (log-transformed of *EDP*) versus l*PGA* (log-transformed *PGA*) for both cases. We can note that only a few simulation runs (5 for Case #1 and 8 for Case #2) lead to the exceedance of the acceptable demand threshold.

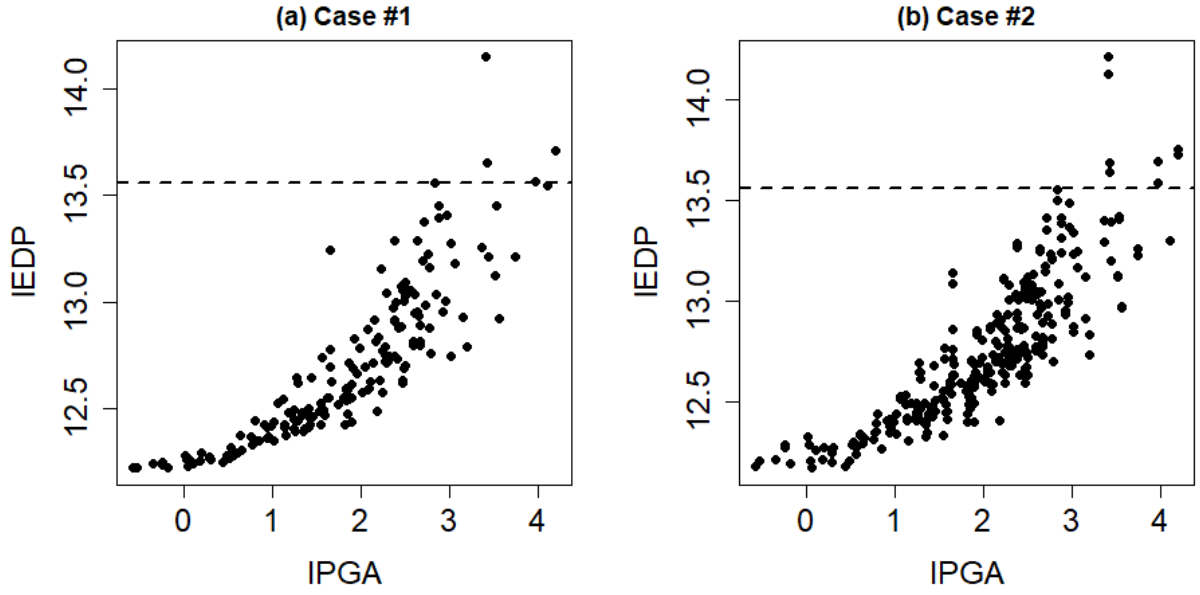


Figure 3: Evolution of l*EDP* (log-transformed *EDP*) as a function of l*PGA* (log-transformed *PGA*) for Case #1 (a) without parametric uncertainty, and for Case #2 (b) with parametric uncertainty. The horizontal dashed line indicates the acceptable demand threshold.

[1] https://ngawest2.berkeley.edu/

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019


Natural Hazards and Earth System Sciences Discussions

4 Applications

In this section, we apply the proposed procedure to both cases described in Sect. 3.2. Sect. 4.1 and 4.2 respectively describes the application for deriving the FCs without and with parametric uncertainty. The analysis is here focused on the log-transformed *PGA* (denoted l*PGA*) to derive the FC.

4.1 Case #1 Derivation of seismic FC without parametric uncertainties

A series of different probabilistic models (Table 2) were fitted to the database of *IM-EDP* points described in Sect. 3.2 (Fig. 3a). Three different probabilistic models (Normal, Tweedie, GEV) and two types of effects on the probabilistic model's parameters were tested (linear or non-linear). Note that the Tweedie distribution corresponds to a family of exponential distributions which takes as special cases the Poisson distribution and the Gamma distribution (Tweedie, 1984).

Table 2. Description of the probabilistic model used to derive the FCs

Model name	Probability model	Type of relationship
NOlin1	Normal	Linear effect on the mean
NOlin2	Normal	Linear effect on the mean and variance
NOsmo1	Normal	Non-Linear smooth effect on the mean
NOsmo2	Normal	Non-Linear smooth effect on the mean and the variance
GEVlin1	GEV	Linear effect on the location
GEVlin2	GEV	Linear effect on the location and scale
GEVlin3	GEV	Linear effect on the location, scale and shape
GEVsmo1	GEV	Non-Linear smooth effect on the location
GEVsmo2	GEV	Non-Linear smooth effect on the location and scale
GEVsmo3	GEV	Non-Linear smooth effect on the location, scale and shape
TWElin1	Tweedie	Linear effect on the location
TWEsmo1	Tweedie	Non-Linear smooth effect on the location

The corresponding AIC and BIC were evaluated (Fig. 4), which shows that GEVsmo2 model should preferably be selected. These criteria suggest that both models, GEVsmo3 and GEVsmo2 are valid (as indicated by the low values for AIC and BIC). The analysis of the regression coefficients of GEVsmo3 shows that the penalisation procedure imposes all coefficient of the shape parameters to be zero, which indicates that l*PGA* only acts on the location and scale parameters. This provides further

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019

Natural Hazards and Earth System Sciences
Discussions

support in favour of GEVsmo2, i.e. a GEV distribution with non-linear smooth term for the location and scale parameters only. The estimated shape parameter reaches here a constant value of 0.07 (+/-0.05), hence indicating a behaviour close to the Gumbel domain. We also note that the analysis of the AIC and BIC values would have favoured the selection of NOsmo2 if the GEV model had not been taken into account, i.e. a heteroscedastic log-normal FC.

260

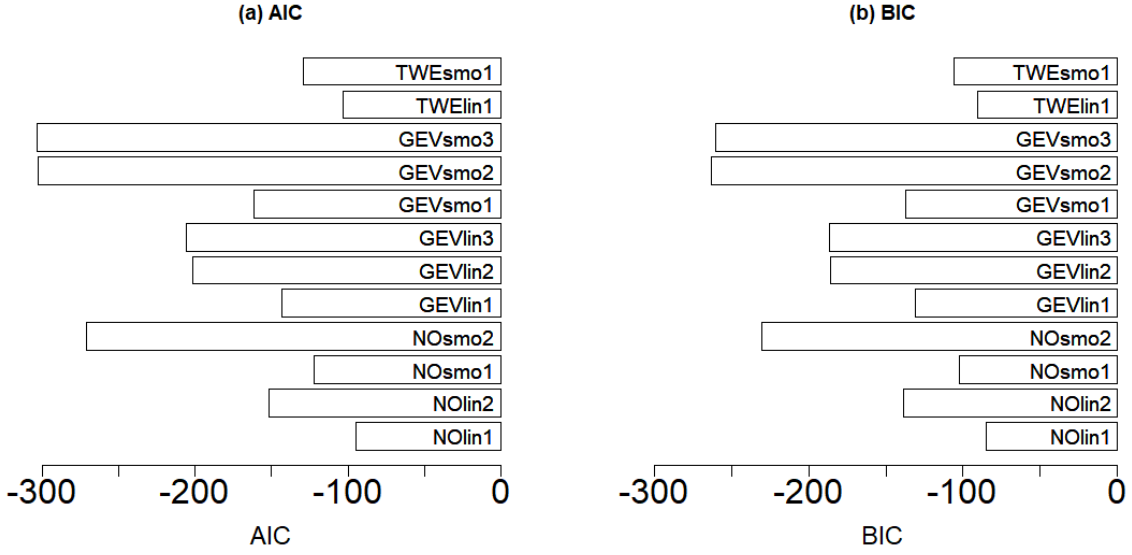


Figure 4: Model selection criteria (AIC (a) and BIC (b)) for the different models described in Table 2 considering the derivation of a FC without parametric uncertainty.

265 The examination of the QQ plots (Fig. 5a) of the model deviance residuals (conditional on the fitted model coefficients and scale parameter) shows a clear improvement of the fitting; in particular for large theoretical quantiles above 1.5 (the dots better aligned along the first bisector in Fig. 5b). The Gumbel QQ plot (Fig. 5c) also indicates a satisfactory fitting of the GEV model.

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019

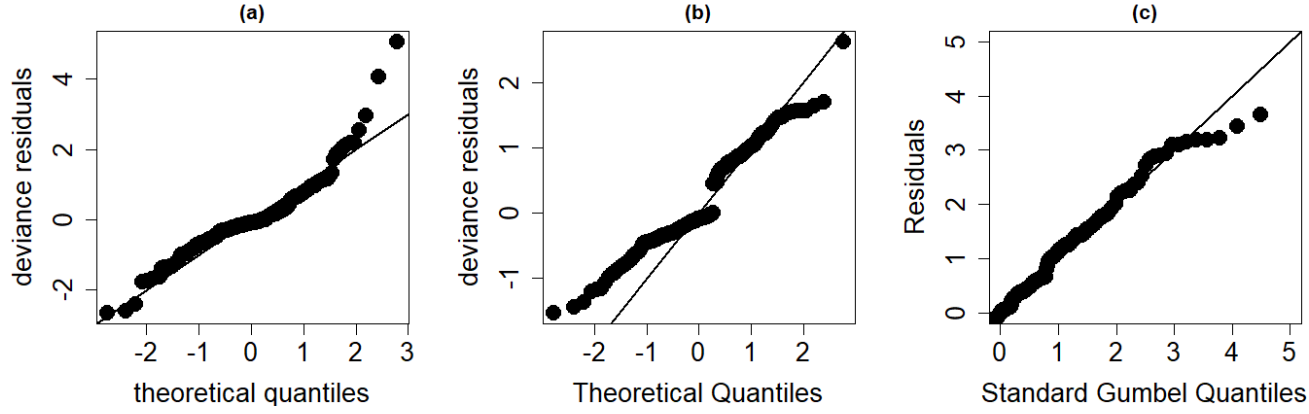


Figure 5: QQ plots to check the validity of the considered model with respect to: (a) the deviance residuals for the NOsmo2 model (b) the deviance residuals for the GEVsmo2 model without parametric uncertainty; (c) the Gumbel quantiles.

Fig. 6a,b respectively provides the evolution of the partial effect with respect to the location and to the scale parameter. We note that the assumption of the relationship between *EDP* and l*PGA* is non-linear (contrary to the widely-used assumption). An increase in l*PGA* both induces an increase of μ and of lσ, hence a shift of the density (as illustrated in Fig. 1b) and an impact on the tail (as illustrated in Fig. 1c).

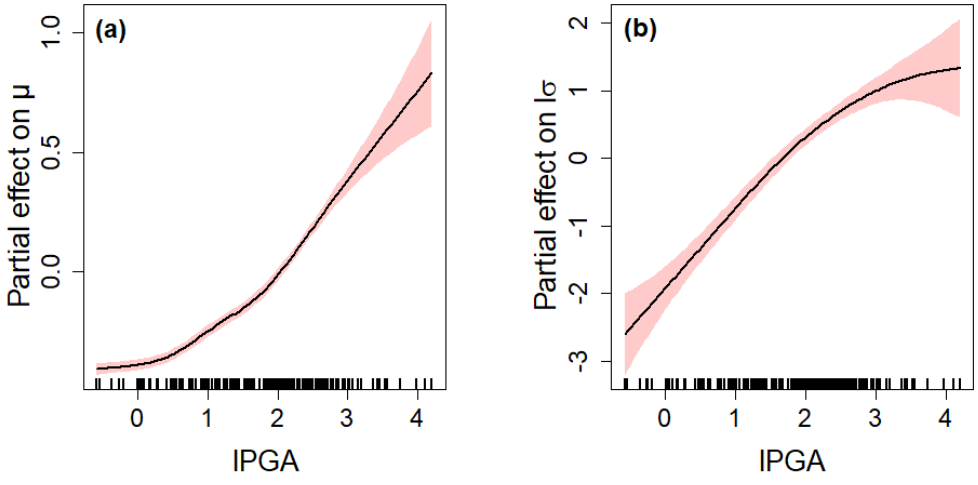
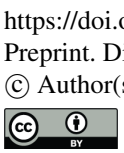


Figure 6: Partial effect of (a) PGA on the GEV location parameter; (b) PGA on the log-transformed GEV scale parameter. The red-coloured bands are defined by 2 standard errors above and below the estimate.

Based on Fig. 6, we evaluate the failure probability (Eq. 2) to derive the corresponding GEV-based FC, which is compared to the one based on the normal assumption (Fig. 7). This shows that the failure probability would have been under-estimated by

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019



Natural Hazards and Earth System Sciences
Discussions

a factor of ~4 if the selection of the probability model had not been applied (i.e. if the widespread assumption of normality had been used).

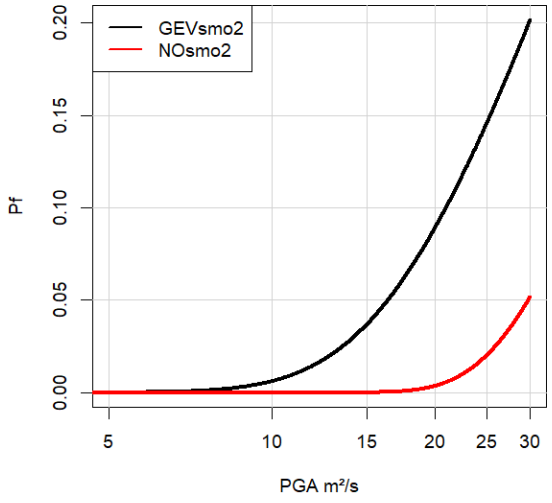


Figure 7: Fragility Curve (that relates the failure probability P_f to PGA) considering the GEV-based approach with non-linear smooth terms of the location and of the scale parameter (termed as GEVsmo2, black line) and considering the approach based on the normal assumption (termed as NOsmo2, red line).

4.2 Case #2 Derivation of seismic FC with parametric uncertainties

In this case, the FCs were derived by accounting not only for l*PGA* but also for 10 additional uncertain parameters (Table 1).

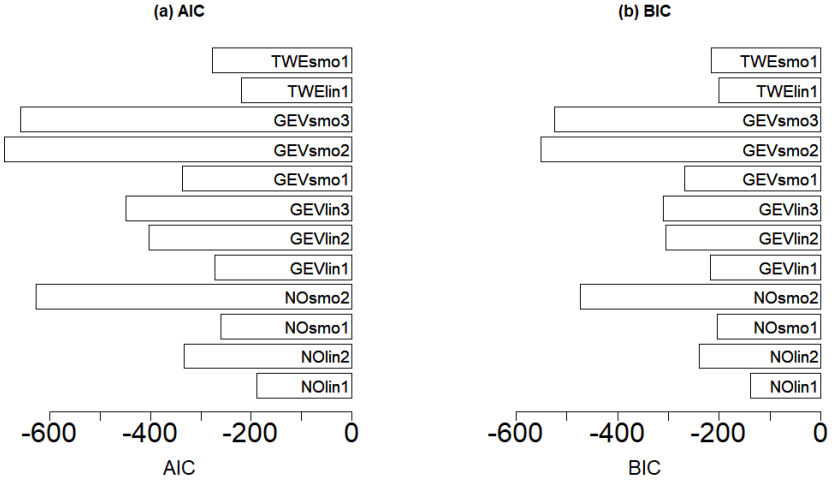


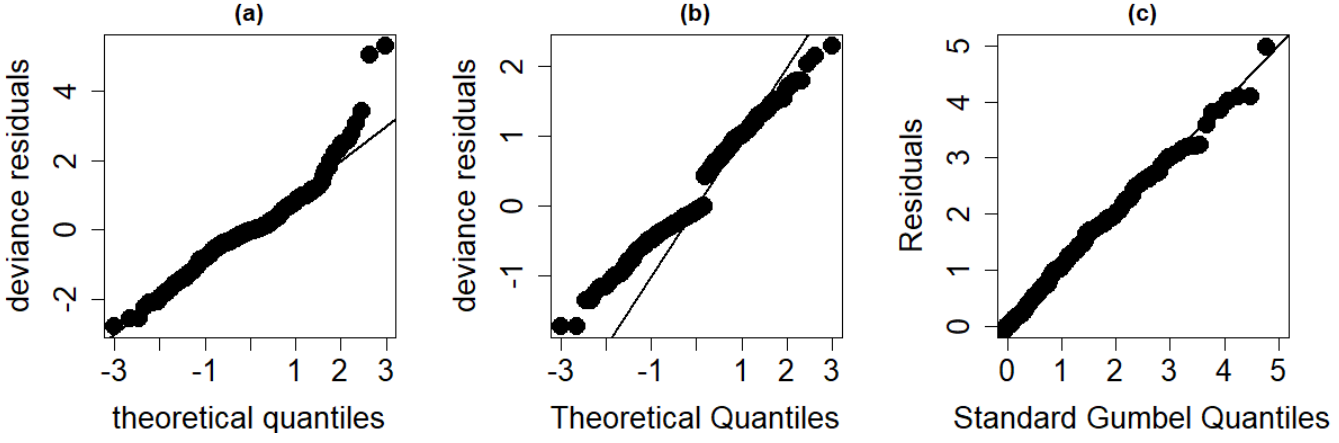
Figure 8: Model selection criteria (AIC (a) and BIC (b)) for the different models described in Table 2 considering the derivation of a vector-based FC with parametric uncertainty.

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019

Natural Hazards and Earth System Sciences
Discussions

295 The AIC and BIC values for the different probabilistic models (described in Table 2) were evaluated (Fig. 8), which shows that GEVsmo2 model should preferably be selected. This indicates that the location and scale parameters are non-linear smooth functions of *IM* and of the uncertain parameters. The estimated shape parameter reaches here a constant value of -0.24 (+/- 0.06), hence indicating a Weibull tail behaviour. Similarly as for the analysis without parametric uncertainties (Sect. 4.1), we note that the AIC and BIC values would have favoured the selection of NOsmo2 if the GEV model had not been taken into
300 account.

The examination of the QQ plots (Fig. 9) of the model deviance residuals (conditional on the fitted model coefficients and scale parameter) shows an improvement of the fitting; in particular for large theoretical quantiles above 1.0 (the dots better aligned along the first bisector in Fig. 9b). The Gumbel QQ plot (Fig. 9c) also indicates a very satisfactory fitting of the GEV model.



305

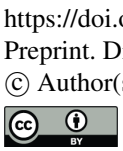
Figure 9: QQ plots to check the validity of the considered model with respect to: (a) the deviance residuals for the NOsmo2 model (b) the deviance residuals for the GEVsmo2 model with parametric uncertainty; (c) the Gumbel quantiles.

Fig. 10 provides the evolution of the partial effects with respect to the location parameter. Several observations can be made:

310
- Fig. 10a show quasi-similar partial effect for l*PGA* (Fig. 6a);
- Three among the ten uncertain parameters were filtered out by the procedure of Sect. 2.4, namely two mechanical parameters (the damping ratio of reinforced pre-stressed concrete ξ_{RPC}, and the damping ratio of the steam line ξ_{SL}) and one geometrical parameter (the pipe thickness of segment #2). As an illustration, Fig. 10e depicts the partial effect of a parameter, which was identified as of negligible influence: here, the partial effect of e_2 is shrunk to zero;

315
- Three thickness parameters (e_1, e_4, e_5) present an increasing linear effect on μ (Fig. 10d,g,h);
- Two parameters (the Young's Modulus of the inner containment E_{IC} and the thickness e_3) present a decreasing linear effect on μ (Fig. 10b,f);
- The damping ratio of the reinforced concrete ξ_{RC} presents a non-linear effect with a minimum value at around 0.0725 (Fig. 10c);

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019



Natural Hazards and Earth System Sciences
Discussions

- The thickness e_6 presents a non-linear effect with a maximum value at around 0.04 (Fig. 10i).

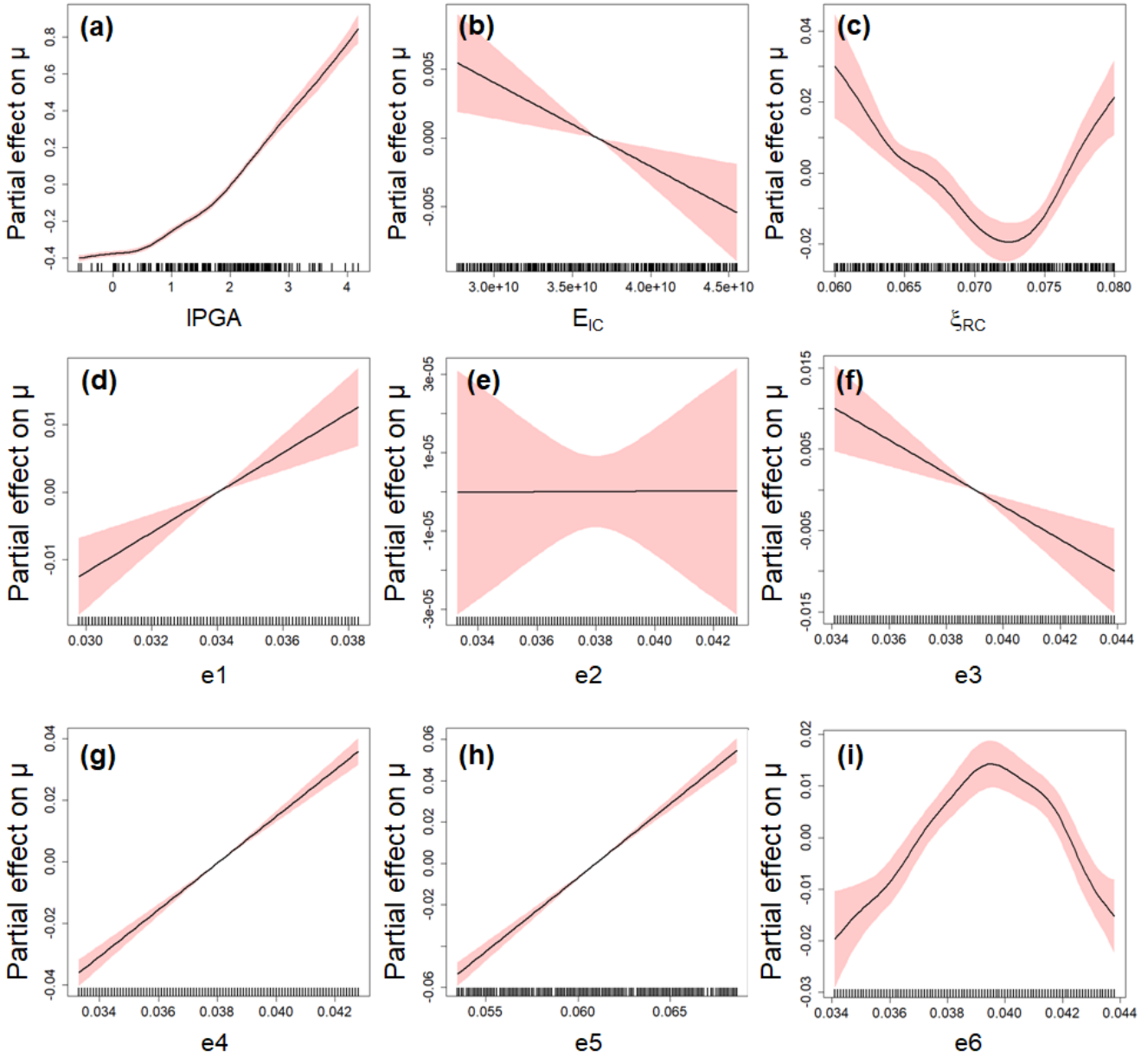
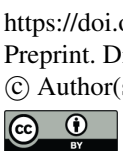


Figure 10: Partial effect on the GEV location parameter. The red-coloured bands are defined by 2 standard errors above and below the estimate.

Fig. 11 provides the evolution of the partial effects with respect to the (log transformed) scale parameter. We show here that a larger number of input parameters were filtered out by the selection procedure i.e. only the thickness e_5 is selected as well as

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019



Natural Hazards and Earth System Sciences
Discussions

the damping ratios of the concrete structures. The partial effects are all non-linear, but with larger uncertainty than for the location parameter (compare the widths of the red-coloured uncertain bands in Fig. 10 and 11). We also note that the partial effect for l*PGA* is quasi-similar to Fig. 6b.

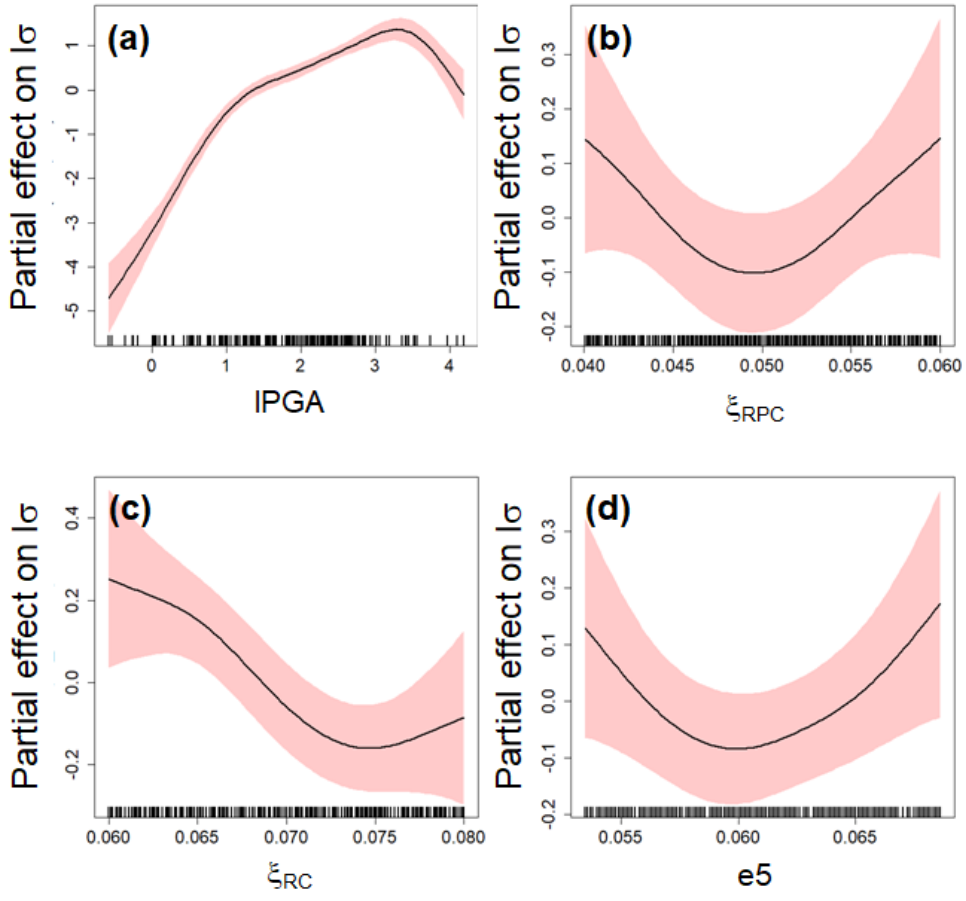


Figure 11: Partial effect on the log-transformed GEV scale parameter. The red-coloured bands are defined by 2 standard errors above and below the estimate.

Based on the results of Fig. 10 and 11, the FC is derived by accounting for the mechanical and geometrical uncertainties by following the procedure in Sect. 2.1 (step 5). Fig. 12a presents the median FC together with the uncertainty induced by the mechanical and geometrical parameters. We show that the GEV-based FC is less steep than for the one without uncertainty (Fig. 6): this is mainly related to the value of the shape parameter (close to Gumbel for case #2 without uncertainty and of Weibull behaviour for case # 1 with uncertainty). Fig. 12a also outlines that the uncertainty on the mechanical and geometrical parameters have a non-negligible influence as shown by the width of the uncertainty bands. Compared to the widely-used

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019

Natural Hazards and Earth System Sciences
Discussions

assumption of normality (Fig. 12b), we reach the same conclusions as for Sect. 4.1, namely an under-estimation of the failure probability when using this assumption, here by a factor of ~3 (by comparing the median FCs).

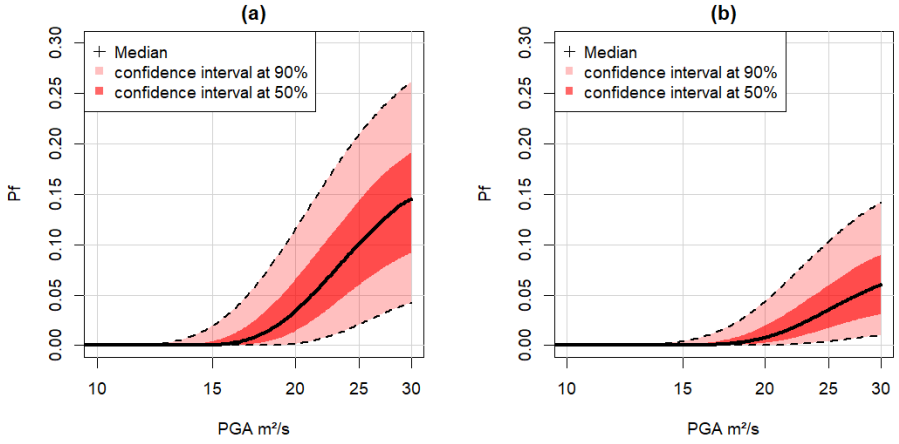


Figure 12: Fragility curve (that relates the failure probability Pf to PGA) considering the uncertain mechanical and geometrical parameters. (a) GEV-based FC; (b) FC based on the normal assumption. The red-coloured band are defined based on the pointwise confidence intervals derived from the set of FCs (see text for details).

The interest of incorporating the mechanical parameters directly in the equation of the FC is the ability to study how the FC in Fig. 12 evolves as a function of the parametric uncertainties, hence to identify regions of the parameters' values leading to large failure probability. This is illustrated in Fig. 13, where the FC is modified depending on the value of the thickness e_4. Here larger e_4 induces a steeper FC. This appears to be in agreement with the increasing effect of e_4 as shown in Fig. 10g.

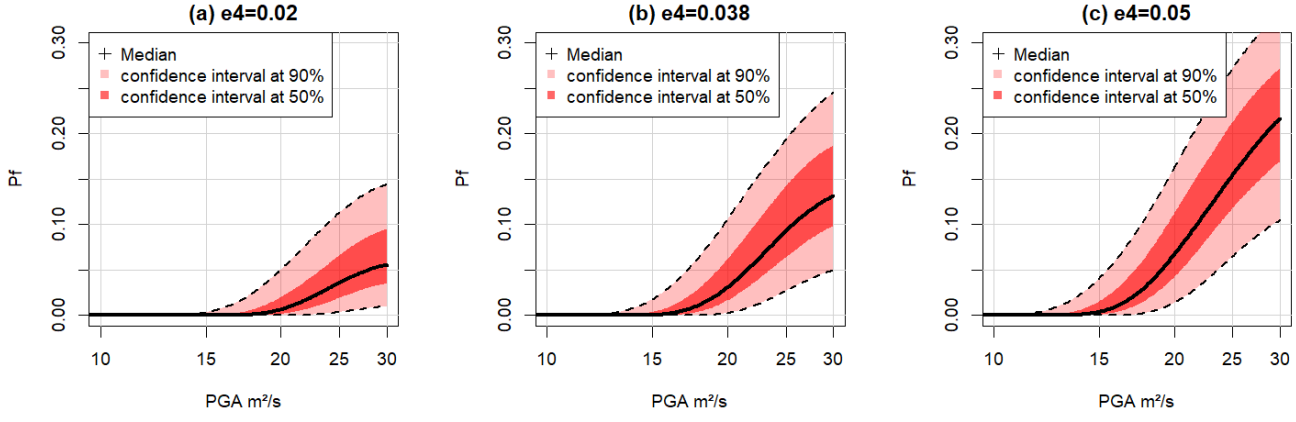


Figure 13: Fragility curve (that relates the failure probability P_f to PGA) by considering the thickness e_4 fixed at a constant value of: (a) 0.02m; (b) 0.038m; (c) 0.05m, while accounting for the variability of the other uncertainties.

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019


Natural Hazards and Earth System Sciences Discussions

5 Discussion and conclusion

The current study has focused on the problem of seismic fragility curve derivation for nuclear power plant safety analysis. We propose a procedure based on the non-stationary GEV distribution to model, in a flexible manner, the tail behaviour of the EDP as a function of the considered *IM*. The key ingredient is the use of non-linear smooth functional *EDP-IM* relationships that are learnt from the data (to overcome limits 2 and 3 as highlighted in the introduction), which avoids the introduction of any a priori assumption on the shape/form of these relationships. The application to a nuclear power plant's steam line and its supporting structure highlights how the application of the (log-)normal fragility curve (to overcome limit 1) would under-estimate the failure probability compared to the GEV-based one. This result brings an additional element against the uncritical use of the (log-)normal fragility curve (see discussions by Karamlou and Bocchini, 2015; Mai et al., 2017; Zentner et al., 2017, among others). We further proposed to incorporate the mechanical and geometrical parameters in the derivation of the fragility curve (using advanced penalisation procedures). This enables us to outline which of these parametric uncertainties have negligible influence on the failure probability as well as the nature of the influence (linear, non-linear, decreasing, increasing, etc.) with respect to each of the GEV parameters (to overcome limit (4)).

The current study should be considered a feasibility assessment of non-stationary GEV tools for FC derivations. Several lines of future research are however identified. From an earthquake engineering viewpoint, the proposed procedure has focused on a single *IM* (here *PGA*), but any other *IM*s could easily be incorporated similarly as for the mechanical and geometrical parameters to derive vector-based fragility curves as done by Gehl et al. (2019) using the same structure. The proposed penalisation approach can be seen as a valuable option to solve a recurrent problem in this domain, namely the identification of most important *IM*s (see discussion by Gehl et al., 2013 and references therein).

From a methodological viewpoint, the derivation of the fragility curves was done using the best estimate of the partial effects. Due to the limited number of observations, these functional terms are associated to uncertainties, which should be integrated in the analysis. Bayesian techniques within framework of GAMLSS (Umlauf et al., 2018) are promising options to overcome this problem. Finally, combining the proposed penalisation with tools of global sensitivity analysis (like Borgonovo et al., 2013) is worth testing in future work to improve the interpretability of the procedure.

Author contributions

JR designed the concept with input from PG. M M-F, YG, NR, JC designed the structural model and provided to PG for adaptation, and implementation on the described case. PG performed the dynamical analyses. JR undertook the statistical analyses and wrote the paper, with input from PG.

Competing interests

The authors declare that they have no conflict of interest.

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019

Code/Data availability

Code are available upon request to the first author.

Acknowledgements

This study has been carried out within the NARSIS project, which has received funding from the European Union's H2020-
390 Euratom Programme under grant agreement N° 755439.

References

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L.: NGA-West2 database, Earthquake Spectra 30, 989–1005, 2014.

Argyroudis, S.A., and Pitilakis, K.D.: Seismic fragility curves of shallow tunnels in alluvial deposits, Soil Dyn. Earthq. Eng.,
395 35, 1–12, 2012.

Akaike, H.: Information theory and an extension of the maximum likelihood principle, in: Selected papers of hirotugu akaike (pp. 199-213), Springer, New York, NY, 1998.

Banerjee, S., and Shinozuka, M.: Mechanistic quantification of RC bridge damage states under earthquake through fragility analysis, Prob. Eng. Mech., 23(1), 12–22, 2008.

400 Borgonovo, E., Zentner, I., Pellegri, A., Tarantola, S., & de Rocquigny, E.: On the importance of uncertain factors in seismic fragility assessment, Reliability Engineering & System Safety, 109, 66-76, 2013.

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer, London, UK, 2001.

Cheng, L., AghaKouchak, A., Gilleland, E., and Katz, R. W. Non-stationary extreme value analysis in a changing climate, Climatic change, 127(2), 353-369, 2014.

405 Combescure, A., Hoffmann, A., and Pasquet, P.: The CASTEM finite element system, in: Finite Element Systems, 115-125, Springer, Berlin, Heidelberg, 1982.

Dutfoy, A.: Estimation of Tail Distribution of the Annual Maximum Earthquake Magnitude Using Extreme Value Theory, Pure and Applied Geophysics, 176(2), 527-540, 2019.

Ellingwood, B.R.: Earthquake risk assessment of building structures, Reliab. Eng. Sys. Safety, 74(3), 251–262, 2001.

410 Gehl, P., Seyedi, D. M., and Douglas, J.: Vector-valued fragility functions for seismic risk evaluation, Bulletin of Earthquake Engineering, 11(2), 365-384, 2013.

Jayaram, N., Lin, T., Baker, J.W.: A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance, Earthquake Spectra, 27(3), 797-815, 2011.

Karamlou, A., Bocchini, P.: Computation of bridge seismic fragility by large-scale simulation for probabilistic resilience
415 analysis, Earthquake Eng. Struct. Dyn., 44(12), 1959–1978, 2015.

https://doi.org/10.5194/nhess-2019-400
Preprint. Discussion started: 13 December 2019

Koenker, R., Leorato, S., Peracchi, F.: Distributional vs. Quantile Regression. Technical Report 300, Centre for Economic and International Studies, University of Rome Tor Vergata, Rome, Italy, 2013.

Lin, T., Haselton, C.B., Baker, J.W.: Conditional spectrum-based ground motion selec-tion. Part I: hazard consistency for risk-based assessments, Earthquake Engineering & Structural Dynamics, 42(12), 1847-1865, 2013.

Marra, G. and Wood, S.N.: Practical variable selection for generalized additive models, Computational Statistics and Data Analysis, 55, 2372-2387, 2011.

Mai, C., Konakli, K., and Sudret, B.: Seismic fragility curves for structures using non-parametric representations. Frontiers of Structural and Civil Engineering, 11(2), 169-186, 2017.

McKay, M., Beckman, R., and Conover, W.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21:239–245, 1979.

Quilligan, A., O Connor, A., and Pakrashi, V.: Fragility analysis of steel and concrete wind turbine towers, Engineering Structures, 36, 270–282, 2012.

Rahni, N., Lancieri, M., Clement, C., Nahas, G., Clement, J., Vivan, L., Guigueno, Y., Raimond, E.: An original approach to derived seismic fragility curves – Application to a PWR main steam line, in: Proceedings of the International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA2017), Pittsburgh, PA, 2017.

Rigby, R.A., Stasinopoulos, D.M.: Generalized additive models for location, scale and shape. J. of the Royal Statistical Society: Series C (Applied Statistics), 54(3), 507-554, 2005.

Schwarz, G.: Estimating the Dimension of a Model, Annals of Statistics 6, 461–464, 1978.

Shinozuka, M., Feng, M., Lee, J., and Naganuma, T.: Statistical analysis of fragility curves. J. Eng. Mech., 126(12), 1224–1231, 2000.

Tweedie, M. C. K.: An index which distinguishes between some important exponential families. Statistics: Applications and New Directions, in: Proceedings of the Indian Statistical Institute Golden Jubilee International Conference (J. K. Ghosh and J. Roy (Eds.)), 579-604. Calcutta: Indian Statistical Institute, 1984.

Umlauf, N., Klein, N., and Zeileis, A.: BAMLSS: bayesian additive models for location, scale, and shape (and beyond), Journal of Computational and Graphical Statistics, 27(3), 612-627, 2018.

Wang, Z., Pedroni, N., Zentner, I., and Zio, E.: Seismic fragility analysis with artificial neural networks: Application to nuclear power plant equipment, Engineering Structures, 162, 213-225, 2018.

Wong, T.E.: An Integration and Assessment of Nonstationary Storm Surge Statistical Behavior by Bayesian Model Averaging, Advances in Statistical Climatology, Meteorology and Oceanography, 4(1/2), 53-63, 2018.

Wood S. Thin-plate regression splines, J. R. Stat. Soc. (B), 2003, 65, 95–114, 2003.

Wood, S. N., Pya, N., and Säfken, B.: Smoothing parameter and model selection for general smooth models, Journal of the American Statistical Association, 111(516), 1548-1563, 2016.

Zentner, I., Gündel, M., and Bonfils, N.: Fragility analysis methods: Review of existing approaches and application, Nuclear Engineering and Design, 323, 245-258, 2017.