Gaussian Kernel Methods for Seismic Fragility and Risk Assessment of Mid-Rise Buildings

Som LakshmiNarasimha Dhulipala

March 2021

Gaussian Kernel Methods for Seismic Fragility and Risk Assessment of Mid-Rise Buildings

Som LakshmiNarasimha Dhulipala

March 2021

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Article

Gaussian Kernel Methods for Seismic Fragility and Risk Assessment of Mid-Rise Buildings

Somayajulu L. N. Dhulipala

Facility Risk Group, Idaho National Laboratory, Idaho Falls, ID 83402, USA; lakshd5@vt.edu

Abstract: Seismic fragility functions can be evaluated using the cloud analysis method with linear regression which makes three fundamental assumptions about the relation between structural response and seismic intensity: log-linear median relationship, constant standard deviation, and Gaussian distributed errors. While cloud analysis with linear regression is a popular method, the degree to which these individual and compounded assumptions affect the fragility and the risk of mid-rise buildings needs to be systematically studied. This paper conducts such a study considering three building archetypes that make up a bulk of the building stock: RC moment frame, steel moment frame, and wood shear wall. Gaussian kernel methods are employed to capture the data-driven variations in the median structural response and standard deviation and the distributions of residuals with the intensity level. With reference to the Gaussian kernels approach, it is found that while the linear regression assumptions may not affect the fragility functions of lower damage states, this conclusion does not hold for the higher damage states (such as the Complete state). In addition, the effects of linear regression assumptions on the seismic risk are evaluated. For predicting the demand hazard, it is found that the linear regression assumptions can impact the computed risk for larger structural response values. However, for predicting the loss hazard with downtime as the decision variable, linear regression can be considered adequate for all practical purposes.

Keywords: seismic fragility; seismic risk; kernel methods; performance-based earthquake engineering; demand hazard

1. Introduction

In the recent past, both the vulnerability and the recovery of buildings under earthquakes have become important aspects to researchers and practitioners alike. Both these aspects are encompassed by the term resilience, the optimization of which is expected to restore the functionality of buildings as quickly as possible after an earthquake event [1]. Resilience is typically assessed by forecasting the functionality recovery path of buildings as a function of time after an earthquake event [2]. Central to an accurate assessment of infrastructure resilience are the fragility functions [3,4]. These functions drive the uncertainties in the times taken for the repair activities, which further influence the inferred resilience metrics [3,4]. Fragility functions derived using the probabilistic demand analysis procedure are also useful for the rapid seismic vulnerability assessment of structures [5,6].

Several methods are available to derive the fragility functions from seismic response analyses. The Incremental Dynamic Analysis (IDA) method scales a suite of ground motions until they each cause structural collapse and the IDA curves are inferred [7]. The Multiple Stripe Analysis (MSA) method is similar to the IDA method, execpt that ground motions are scaled to a common Intensity Measure (IM) level and the analyst has the option of choosing different ground motions for different IM levels [8]. Typically, the data obtained from IDA or MSA is analyzed using techniques such as Maximum Likelihood to derive the fragility functions [9]. The cloud analysis method establishes a statistical relationship between building response, typically the Peak Interstory Drift (PID),

and IM [10]. It is quite ubiquitous to use linear regression to establish a relation between PID and IM in a log-log space [10]. Fragility function is then inferred using:

$$P(PID > y \mid IM) = 1 - \Phi\left(\frac{\ln y - \ln \overline{PID}}{\sigma}\right) \tag{1}$$

where, y is the PID level of interest, Φ is a Gaussian cumulative distribution function, $\ln \overline{PID}$ is the predicted log PID through a linear relation with $\ln IM$, and σ is the standard deviation of this linear relation. In addition to a linear relation between $\ln PID$ and $\ln IM$, Equation (1) assumes that the regression residuals are Gaussian distributed with a constant standard deviation across the IM levels. Cloud analysis is a widely adopted method for inferring fragility functions [11]. Particularly, due to its simplicity and the option to use unscaled ground motions, unlike the other methods. Ground motion scaling has been debated in the Performance-Based Earthquake Engineering community, with some studies claiming that its practice may lead to biased response estimates [12,13].

While cloud analysis provides a convenient way to infer fragility functions, the assumptions underlying this method have been sometimes questioned. Ref. [14] found that a linear relation between log structural response and $\ln IM$ is inadequate at large IM levels, and consequently, they adopted a bilinear regression model. Ref. [15], while studying the seismic response of concrete frames, found that the standard deviation (σ) can subtly change with the IM level, oscillating around the constant standard deviation of linear regression. Ref. [16] found that the Gaussian distribution of regression residuals assumption can be inadequate for certain seismic applications, especially at large IM levels. Yet, the impacts systematically alleviating the linear regression assumptions have on fragility functions needs more carefully exploration. The degree to which the individual/compounded assumptions of linear regression affect the different damage states for a variety of mid-rise buildings (which make up a bulk of the building stock) has not been explored. Moreover, the impacts these assumptions have on the seismic risk of buildings assessed through the Performance-Based Earthquake Engineering framework (PBEE) [17] have not been investigated. There is a need to provide general guidance to researchers and practitioners on the uses and drawbacks of linear regression towards inferring the fragility functions and the subsequent risk.

This paper employs Gaussian kernel techniques to systematically alleviate the linear regression assumptions and explore the impacts on fragility and risk. Gaussian kernel techniques provide a form-free and data-driven means to capture the median variation of structural response, prediction standard deviation, and distribution of residuals as a function of the IM. Local regression with a Gaussian kernel is used to capture both the median variation of response and standard deviation given the IM level. Gaussian kernel density coupled with the K Nearest Neighbor algorithm is used to infer the variation in the distribution of residuals given the IM level. Fragility functions of three mid-rise buildings obtained using linear regression and incrementally alleviating the linear regression assumptions through Gaussian kernel techniques are compared. The three archetype buildings are Reinforced Concrete (RC) moment frame, steel moment frame, and wood shear wall. The impacts of systematically alleviating the linear regression assumptions on the seismic risk are also discussed. Seismic risk is characterized as demand hazard and loss hazard with downtime as the decision variable evaluated through the PBEE framework.

This paper is organized as follows. Section 2 discusses the archetype buildings and the ground motion records employed in this study. Section 3 discusses Gaussian kernel techniques for predicting the median seismic response, variation of the prediction standard deviation, and variation of the residuals distribution with the IM level. Sections 4 and 5 discuss the impacts a completely data-driven approach have on fragility functions, demand hazard, and loss hazard.

2. Case Study Description

The archetype buildings considered, the ground motion records used, and the subsequent seismic response analyses results are discussed in this section.

2.1. Structural Models

Three archetypes of mid-rise buildings are considered for this study. The first archetype is the seven story Van Nuys reinforced concrete building with minor structural repairs post the 1971 San Fernando earthquake [18]. Seismic response of this building has been extensively studied during the developmental phases of the PEER framework for PBEE [19]. Being a pre-code building, its ductility may be inferior to RC buildings designed per contemporary codes, and hence its consideration would be interesting for this study. Using OpenSees [20], Kalkan [21] modeled the 2-D moment frame in the East–West direction which consists of eight bays. Material nonlinearity was accounted through the Steel01 and the Concrete01 materials in OpenSees. Distributed plasticity was considered using fiber sections; this also makes the model computationally more expensive than using concentrated plasticity modeling through plastic hinges. The fundamental oscillator period of the model is about 1.5 s. Paspuleti [22] further verified the time history response of this model by comparing it with the recorded time history of the instrumented Van Nuys building during the 1994 Northridge earthquake.

The second archetype is a four-story steel building designed by Lignos [23] for high-code in Los Angeles, California. Using OpenSees, Eads [24] modeled the 2D moment frame in the East–West direction which consists of two bays. Material nonlinearity is considered through a concentrated plasticity model with the plastic hinges located at the reduced sections of the beams. These plastic hinges are modeled as rotational springs using the modified Ibarra–Medina–Krawinkler (IMK) bilinear model [25]. Parameters for implementing the IMK model are inferred from experiments conducted on steel specimens by Lignos [23]. A leaning column approach is used to account for geometric nonlinearity. The fundamental oscillator period of this model is about 1.33 s.

The third archetype is a four-story wooden building designed for high-code by Jayamon [26]. We modeled the lateral system, which is a shear wall, using OpenSees. Material nonlinearity of the shear wall is considered through the SAWS model [27]. The SAWS model requires a set of ten parameters to be input. These parameters were inferred by Jayamon [26] using the program CASHEW [27]. A leaning column approach is employed to consider geometric nonlinearity. The model has a fundamental oscillator period of about 0.53 s.

2.2. Ground Motion Records

This study employs a generic set of 380 unscaled ground motions obtained from the PEER NGA-West2 database [28]. Figure 1a presents the magnitude-distance scatter plot of the records. The mean magnitude, distance, and PGA values are 6.67, 15.82 Km, and 0.29 g, respectively. Figure 1b presents the response spectra and the geometric mean of the response spectra of the records. It is noted that the record set used for analysis does not make any distinction between near- and far-field recordings or the two horizontal components of a ground motion. Some studies have investigated the influence of these characteristics of ground motions on the seismic response of structures [29,30]. However, the motive of this study is to investigate a completely data-driven approach to compute fragility and risk. Since the same general set of ground motions is used across all the statistical analyses that are subsequently performed, a common basis for comparisons between the results is established.
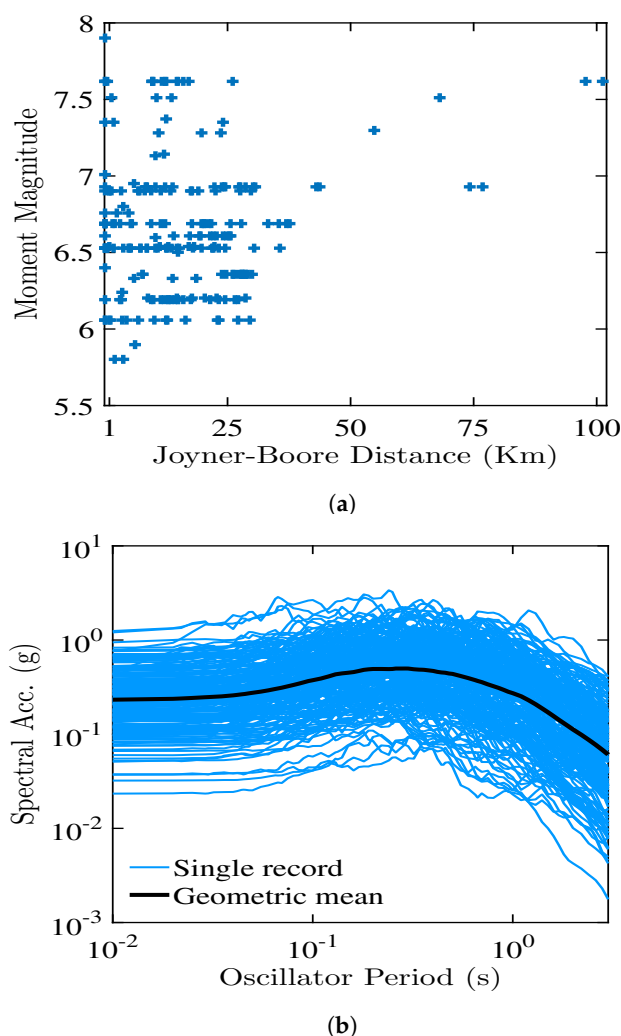
(**a**)

(**b**)

Figure 1. (**a**) Magnitude-distance values of the ground motions considered in this study. (**b**) Response spectra of the ground motions considered in this study.

2.3. Seismic Response Analyses

Seismic intensity is quantified through the spectral acceleration at the fundamental oscillator period of the archetype since this IM is found to be both efficient and sufficient for predicting the global responses [31]. The global response is quantified using Peak Inter-story Drift (PID). Those response simulations which led to numerical instabilities (i.e., non-convergence of the solution algorithms) or excessive PIDs were excluded from further statistical analysis. Figure 2 presents the seismic response results of the three archetype buildings used for this study. The RC moment frame, being designed to pre-code standards, exhibits larger seismic PIDs and earlier onset of nonlinearities than the other two archetypes. The wood shear wall exhibits the lowest seismic drifts and later onset of nonlinearities. A high-code design and lower dead loads may contribute to this behavior of wood shear wall. Behavior of the steel moment frame falls in between the other two archetypes.
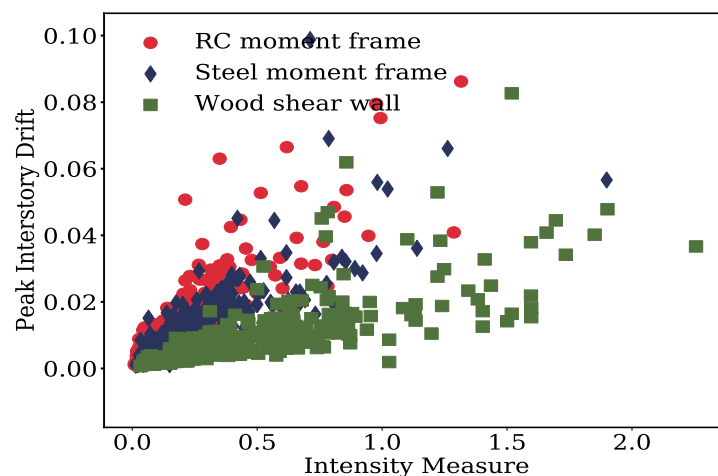
Figure 2. Peak Interstory Drift versus Intensity Measure plots of the three archetype buildings considered in this study. The IM considered for the RC frame, steel frame, and wood shear wall are *Sa* (1.5 s), *Sa* (1.33 s), and *Sa* (0.53 s), respectively.

3. Data-Driven Seismic Response Modeling Using Gaussian Kernels

In this section, kernel regression is applied to predict the median responses and the standard deviation around the predictions of the archetype buildings as a function of the IM. Any predictive model for response whose standard deviation depends on the IM is termed a heteroscedastic model [32]. In addition, kernel techniques are applied to characterize the distribution of residuals around the median predictions. The median response prediction, the standard deviation around the predictions, and the distribution of residuals are all critical for the development of fragility functions.

3.1. Overview of the Kernel Function in Kernel Regression

A non-parametric regression technique termed kernel regression is adopted. Kernel regression predicts the median response (i.e., PID) as a function of the IM by locally averaging the responses near the input IM level. A kernel function is used for weighting the responses towards predicting the local median response. Any function $K(u)$ that integrates to one and which is symmetric for all values of *u* is a kernel function. While the first property of a kernel ensures that it is a probability density function, the second property sees that the average of this function is equal to that of the response sample used [33]. Many choices for the kernel function exist and a Gaussian kernel is quite popular. It is mathematically expressed as:

$$\mathcal{K}(u) = \frac{1}{h\sqrt{2\pi}}\, e^{-u^2/(2\,h^2)} \tag{2}$$

where *h* is the bandwidth. The bandwidth is an important parameter which controls the smoothness of the kernel regression, with a low value resulting in very local (i.e., spiky or jagged) predictions of the median response. Ref. [34] discuss the influence of the bandwidth on damage predictions of buildings after earthquakes. A Gaussian kernel is a popular choice for the kernel function and it is adopted in this paper. Furthermore, for a uni-dimensional predictor (i.e., the *IM*), the choice of the kernel by itself does not significantly affect the performance of kernel regression in relation to other attributes such as the bandwidth.

3.2. Predictive Models for Median Structural Response

Given seismic response analysis results represented by the PID and the IM of interest, predictive model using kernel regression is expressed as:

$$\ln(\overline{PID}) = m(\ln IM) + \varepsilon \tag{3}$$

where i is the index of the PID-IM pair, $m(.)$ is a smooth, but unknown, function of $\ln IM$ that needs to be determined, and ε is the residual of the prediction. The function $m(.)$ in Equation (3) is equivalent to the expectation of $\ln PID$ conditional on $\ln IM$ [i.e., $E(\ln PID|\ln IM)$]. Logarithms of PID and IM are considered in Equation (3), in line with most previous studies on developing predictive models for seismic fragility evaluation. The goal in kernel regression is to infer the function $m(.)$ locally at various IM levels with aim of capturing a completely data-driven variation of PID with IM. Unlike linear regression that fits a constant trend across all $\ln IM$ levels, kernel regression infers a trend which is specific to the neighborhood of the required IM level.

The mechanism with which kernel regression captures the local PID variation as a function of the IM is explained briefly, but interested readers are referred to Racine [33] for an exhaustive explanation. Kernel regression employs kernel functions to estimate the average PID value by using the PID-IM pairs only in the neighborhood of the input IM level. A popular variant of kernel regression is the local constant regression which simply predicts a local average value of PID as a smooth function of $\ln IM$. This smoothness is controlled by the bandwidth parameter h in the Gaussian kernel function (Equation (2)). The definition for $m(.)$ is quite intuitive since it is equivalent to a statistical expectation [33]:

$$m(\ln IM) = \frac{\sum_{i=1}^{N} \mathcal{K}(\ln IM_i - \ln IM) * \ln PID_i}{\sum_{i=1}^{N} \mathcal{K}(\ln IM_i - \ln IM)} \tag{4}$$

where $\mathcal{K}(.)$ is the Gaussian kernel function defined in Equation (2). It is noted that the numerator in Equation (4) is a weighted average with more weight assigned to those PID_i values whose corresponding IM_i values are closer to the input IM. The denominator is a normalizing constant. Equation (4) is termed a local constant regression because it predicts an almost constant value of PID at the boundaries of the PID-IM data set. Local linear regression is next discussed to alleviate this limitation.

The basic idea in a local linear regression is to fit a model of the form [33]:

$$\ln(PID_i) = \gamma_1 + \gamma_2 (\ln IM_i - \ln IM) + \varepsilon_i \tag{5}$$

where γ_1 and γ_2 are the coefficients . The regressor [i.e., $(\ln IM_i - \ln IM)$] has this form so that the intercept γ_1 is equal to the required function $m(\ln IM)$. Minimizing the following expression gives an estimate of γ_1 [or $m(\ln IM)$] [33]:

$$\min_{\gamma_1, \gamma_2} \left[\left(\ln(PID_i) - \gamma_1 - \gamma_2 (\ln IM_i - \ln IM) \right)^2 \mathcal{K}(\ln IM_i - \ln IM) \right] \tag{6}$$

Tere is a closed form estimate of γ_1 [or $m(\ln IM)$] and interested readers are referred to Racine [33] for more information. Whereas local constant regression extrapolates a constant value of PID beyond the PID-IM dataset, a local linear regression extrapolates a log-linear function of the IM.

As mentioned previously, the bandwidth parameter h in the kernel function ($\mathcal{K}(.)$ in Equations (4) and (6)) influences the smoothness of the median response prediction. With such an important role played by this parameter, many rules for optimal bandwidth selection have been proposed. In general, the different optimal bandwidth rules should result in somewhat similar estimates of the median response, if not identical. Two popular rules for bandwidth selection are employed in this paper: least squares cross validation

(LS-CV) and Akaike Information Criterion (AIC). Least squares cross validation employs a leave-one-out cross validation error to evaluate the alternative bandwidth values [35]:

$$CV(h) = \frac{1}{N} \sum_{i=1}^{N} \left(Y_i - m_{-i}(\ln IM; h) \right) \tag{7}$$

where $m_{-i}(\ln IM; h)$ is the kernel regression estimate by leaving out the ith datum (i.e., $\ln PID_i, \ln IM_i$) for a certain value of the bandwidth h. The $CV(h)$ is computed for several values of h and that value of the bandwidth is adopted which results in least the CV. Akaike Information Criterion selects the optimal bandwidth based on information theoretic principles in relation to kernel regression, as discussed by Loader [35].

Figure 3 presents the predictive modeling results of standard linear, local linear, and local constant regressions. For the latter two regression types, results concerning the two optimal bandwidth rules are also presented. First, let us consider the different regression types under LS-CV bandwidth criterion and discuss the AIC bandwidth later. At the edges of the dataset, local constant regression is predicting somewhat constant values of the PID. This is seen at larger IM values of the RC and the steel moment frames (Figure 3a,b, respectively) and at smaller IM values of the wood shear wall (Figure 3c). This behavior is undesirable for fragility function development since it is well known that PID generally increases with the IM when considering non-collapse cases only. Local linear regression predicts a median PID that is linearly increasing with the IM (in the log-log space), albeit without necessarily the same slope at different IM levels. For the wood shear wall (Figure 3c), local linear predicts median PIDs that are quite similar to linear regression. While this behavior is mostly true for the RC and steel moment frames too (Figure 3a,b), at large IM levels (>0.6 g and >0.45 g, respectively), local linear predicts PIDs with a different slope than linear regression. As presented more clearly in Figure 4, this change in slope is supported by another linear regression performed on data with IM greater than 0.6 g and 0.45 g for the RC and steel moment frames, respectively. For the non-ductile RC moment frame (Figure 4a), this increase in slope can be attributed to early on set of softening effects prior to collapse, leading to increased median PIDs [36]. For the ductile steel moment frame (Figure 4b), this decrease in slope can be attributed to slight period elongation and yielding, leading to slightly reduced median PIDs [14].

The two optimal bandwidth estimation methods are seen to result in similar estimates of the median PID for local linear, in general, in Figure 3. However, the differences at low and high IM levels under the RC moment frame makes us prefer one method more over the other. At these IM levels, the AIC method is resulting in abrupt changes in the estimated PID values. The LS-CV method is resulting in smoother median estimates.

Per the previous discussion, local constant regression predicts a constant median PID beyond the PID-IM dataset and local linear regression with AIC bandwidth may lead to abrupt changes in the median PID estimates at low and high IM levels. Local linear regression with LS-CV bandwidth seems to perform more satisfactorily and will be considered for further evaluation of the fragility functions and the demand and loss hazards.
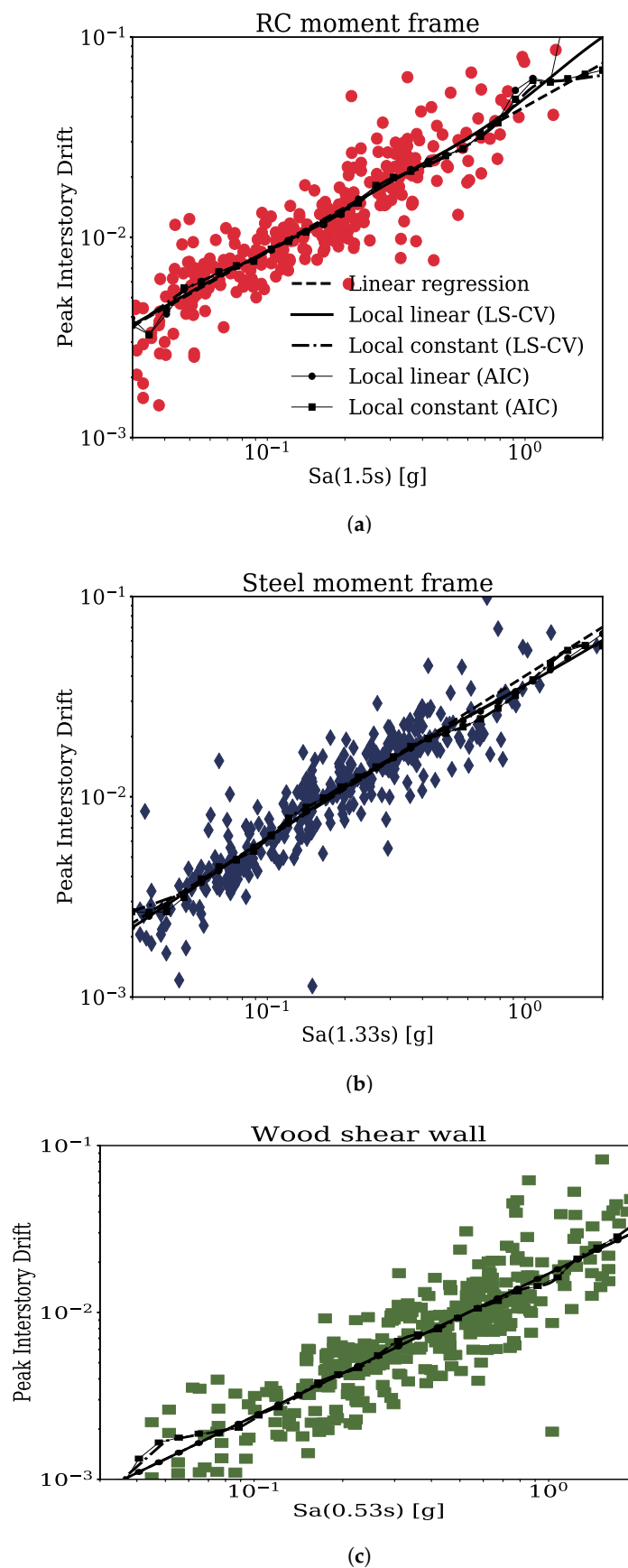
Figure 3. Performance of linear regression, local linear regression, and local constant regression for predicting Peak Interstory Drift considering: (**a**) RC moment frame; (**b**) Steel moment frame; and (**c**) Wood shear wall. For the latter two regression types, Least Squares Cross Validation and Akaike Information Criterion methods are considered for determining the optimal bandwidth.
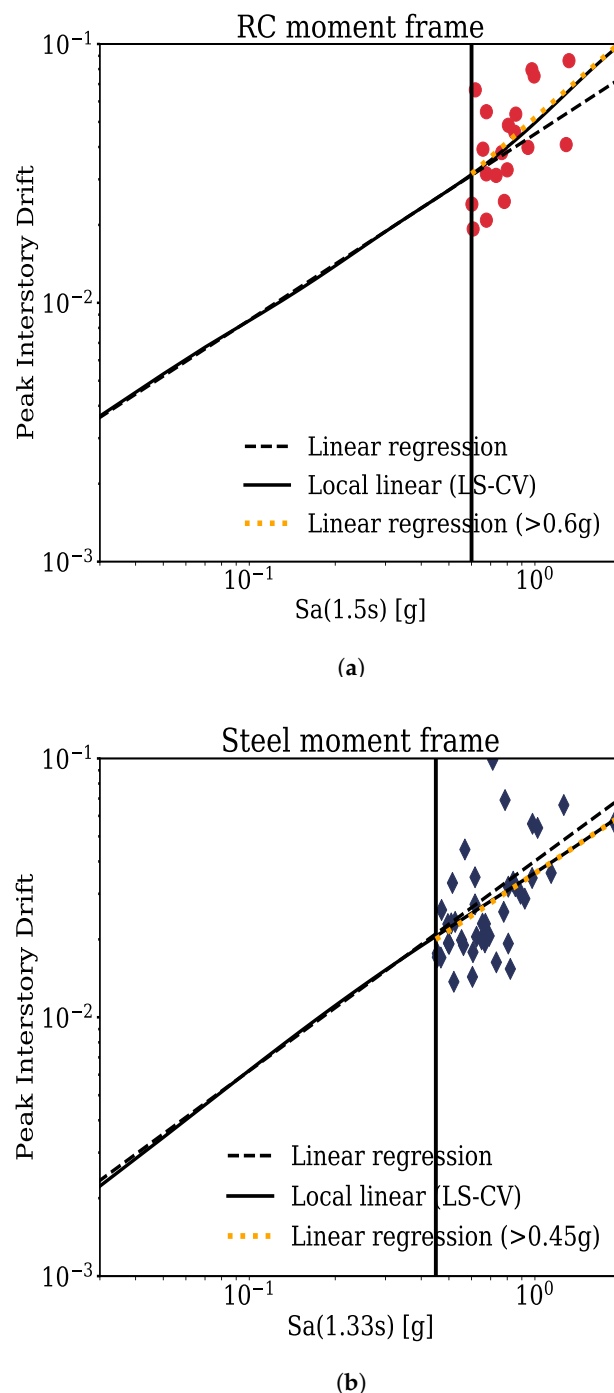
(**a**)

(**b**)

Figure 4. Comparison of linear regression and local linear regression (LS-CV bandwidth) beyond the IM levels 0.6 g and 0.45 g for the (**a**) RC and (**b**) steel moment frames, respectively. Separate linear regressions performed for those data points beyond these IM levels are also presented. While for the RC frame, there are 30 data points beyond 0.6 g, the steel frame had 43 data points beyond 0.45 g.

3.3. Predictive Models for Standard Deviation around the Median Response

In linear regression for cloud analysis, it is widely assumed that the standard deviation around the median is constant with the IM level. Or in other words, the average squared error between the recorded and the predicted ln PID (i.e., $\bar{\varepsilon}^2$) does not change with the IM value. This assumption need not hold true in all cases and to overcome this, the squared

errors obtained from the median PID prediction can be modeled as function of the IM using kernel regression:

$$\bar{\varepsilon}^2 = s(\ln IM_i) + v_i \tag{8}$$

where $s(.)$ is again an unknown smooth function that needs to be estimated and v is the error. Notice that square root of the median error squared is the standard deviation, by definition. It is desirable to model the squared error resulting from the median PID predictive model instead of the error itself to ensure positivity of the standard deviation. Alternative choices exist for inferring the function $s(.)$. We consider local linear and local constant regressions with $LS - CV$ and AIC optimal bandwidths.

Figure 5 presents the results for the three archetype buildings considering the four alternative modeling choices of the squared error. While the four modeling choices lead to somewhat similar median squared error predictions, there are some differences that warrant discussion. Local linear is resulting in a linear variation of the squared error near the edges of the dataset. This is clearly evident at large IMs for the steel frame (Figure 5b) and at small and large IMs for the wood shear wall (Figure 5c). This behavior is again due to local linear extrapolating a linear trend beyond the dataset. However, it is conservative to extrapolate a constant value of the median squared error as there is exists no justification for any other variation. From this respect, local constant regression is more desirable. Under local constant regression, the two alternative bandwidth methods are leading to very similar median errors. For evaluating the fragility functions and the downtime hazard subsequently, the AIC bandwidth will be considered since the median estimates are flatter near the edges of the dataset than LS-CV (especially for the RC frame; Figure 5a).

3.4. Characterizing the Distribution of Peak Interstory Drift Prediction Residuals

Another assumption made by linear regression is, the PID prediction residuals (ε_i) are Gaussian distributed. Some studies have noted that this assumption need not be valid [16]. The Gaussian kernel density enables a data driven characterization of the distribution of these residuals without making any assumptions about the functional form of the distribution (An important distinction should be noted here. A Gaussian or a Normal distribution is parametric defined by the median and the standard deviation. A Gaussian kernel density, however, is non-parametric and is designed to take the form that best describes the data without a fixed functional form). A Gaussian kernel density is defined by:

$$f(\varepsilon) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{K}(\varepsilon - \varepsilon_i) \tag{9}$$

where, $\mathcal{K}(.)$ is the Gaussian kernel function defined in Equation (2). As with kernel regression, the choice of the bandwidth parameter h plays a crucial role in controlling the smoothness of the Gaussian kernel density. Smaller bandwidths lead to a large variance defined by highly localized predictions of the density [overfitting]. Larger bandwidth leads to a large bias defined by highly generalized predictions of the density [underfitting]. Plug-in methods or cross validation methods are commonly employed to select the optimal bandwidth [35]. Plug-in methods such as the Silverman's optimal bandwidth are based on thumb rules and can lead to inaccuracies in bandwidth estimation. We therefore employ a cross validation based bandwidth selection by minimizing the Mean Integrated Squared Error (MISE). Minimizing the MISE achieves a balance between overfitting and underfitting.
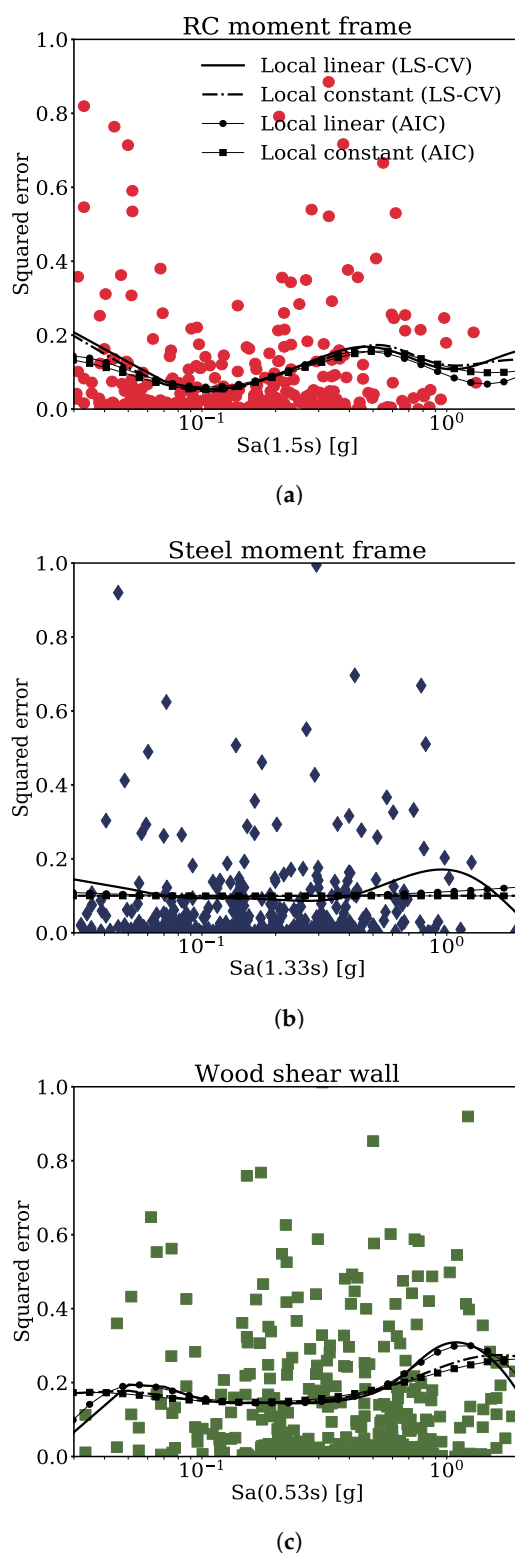
Figure 5. Performance of local linear regression and local constant regression for predicting the standard deviation variation with the IM level considering: (**a**) RC moment frame; (**b**) Steel moment frame; and (**c**) Wood shear wall. Least Squares Cross Validation and Akaike Information Criterion methods are considered for determining the optimal bandwidth. It is noted that square root of the median squared error is the standard deviation.

The distribution of residuals from applying linear or local linear regressions to predict the median PID can be characterized as a Gaussian kernel density. However, this characterization still assumes that the residuals distribution, despite being non-parametric, does

not change with the input IM. In other words, the variation in standard deviation of the residuals with the input IM level cannot be captured. To overcome this limitation, there are two options: use a kernel density estimation with variable bandwidth as a function of the input IM [37] or use a K Nearest Neighbor (KNN) approach to select K residuals nearest to an input IM level and apply kernel density estimation on these nearest residuals. The latter approach is applied in this paper as it allows us to explicitly see the variation of the standard deviation with the IM, in addition to the variation of the distribution of the residuals.

The KNN kernel density approach first selects K neighboring residuals nearest to the input IM level. This parameter K is similar in function to the bandwidth in kernel regression in that it controls the locality with which the variation in distribution of residuals is captured. In practice, the parameter K is mostly selected based on the thumb rule $K = 2\sqrt{N}$ rounded to the nearest integer [38]. While this rule of thumb is mostly used in practice, we wanted to also evaluate alternative values of K using a quantitative approach to avoid over-fitting. However, quantitative approaches for selecting K in the KNN method for density estimation are scarce, if not non-existent. We hence designed a cross validation scheme to check the predictive capability of the standard deviation variation given a K value. Standard deviation of the K values near the IM level governs the bandwidth which further influences the quality of the kernel density estimate [33]. So, if the standard deviation can be predicted well without over fitting, it is highly likely that the same holds when characterizing the distribution of residuals and its variation using kernel density estimation.

The cross validation scheme, for a given K value, proceeds as follows:

- At each IM level, K residuals nearest to the IM level are selected using a nearest neighbor algorithm with uniform weights.
- The K samples are randomly divided into training and test subsets in 77–33 proportion thirty times.
- The standard deviation variation across the different IM levels is independently computed for the training and the test subsets across all the thirty partitions.
- Then, the sum of squared differences (SSD) between the standard deviation from training and test subsets is averaged across the thirty partitions:

$$SSD = \frac{1}{30}\sum_{i=1}^{M}(\sigma_i^{train} - \sigma_i^{test})^2 \tag{10}$$

where, M is the number of IM levels and σ is the standard deviation.
- For a low K, the SSD value would be high due to overfitting. As the K increases, the SSD starts decreasing as enough samples are available in the training and test subsets to predict similar standard deviation variations across the IM levels. The K at which the reduction in SSD starts being small is the one required for subsequent analysis.

Figure 6 presents the SSD metric as a function of the number of neighbors for the three archetypes. Across all the archetypes, the SSD metric is becoming approximately a constant for higher values of K. The vertical line in Figure 6 represents the number of neighbors recommended by the rule of thumb, which is 38. It is seen that at $K = 38$, the average SSD is a constant for the RC frame, and almost a constant for the other two archetypes. So, the rule of thumb maybe used for selecting K across all the archetypes. However, to be on the conservative side, 38, 45, and 50 are the K values selected for the RC frame, steel frame, and the wood shear wall, respectively. These selected values are represented as scattered points in Figure 6.
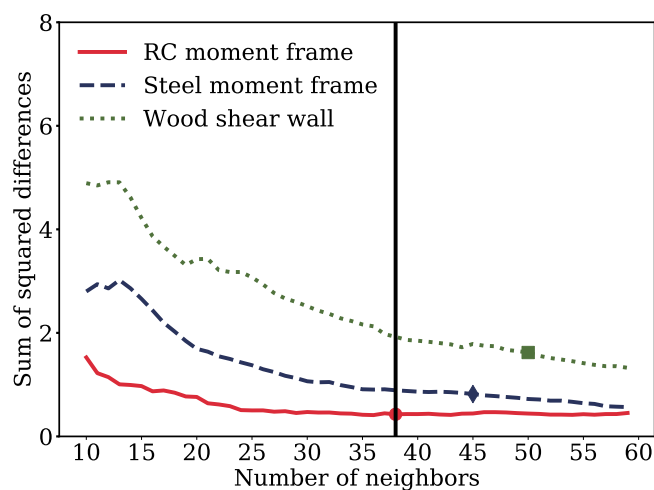
Figure 6. Variation of the average SSD metric with the number of neighbors. The scattered points represent the number of neighbors used for analysis. The vertical line represents the optimal number of neighbors suggested by a rule of thumb.

Figure 7 presents the distributions of PID residuals computed using three cases: (1) Gaussian distribution under linear regression; (2) kernel density under local linear; and (3) kernel density with KNN under local linear. When using a kernel density with KNN, the residuals distribution changes with the input IM level. Some observations can be made from Figure 7. First, cases 1 and 2 are resulting in similar distributions of the residuals which indicates that residuals distribution of local linear regression is approximately a Gaussian distribution, on an average. Second, case 3 is resulting in residuals that are again approximately Gaussian distributed at most IM levels for the RC frame and at some IM levels for the steel frame (Figure 7a,b, respectively), although with a changing standard deviation across the IM levels. Furthermore, the residuals distributions at different IM levels are seen to fluctuate around the case 2 distribution. Third, under case 3, the residuals deviate from a Gaussian distribution at large IM levels for the steel moment frame and at some intermediate IM levels for the wood shear wall. The impacts these observations have on the fragility functions and the downtime hazard will be explored subsequently.

Figure 8 presents the standard deviation variation with the IM level. This variation is computed using two methods: (1) local constant regression applied to predict the mean squared error, as discussed in Section 3.3; and (2) the KNN technique discussed previously. Both these methods use the local linear regression to predict ln $\overline{PID}$. Furthermore, presented in Figure 8 is the constant standard deviation predicted by linear regression. The following observations are made. Methods 1 and 2 both predict variations that oscillate around the constant standard deviation predicted by linear regression. For method 2, this was evident in Figure 7, where the residuals densities at different IM levels characterized by KNN kernel density oscillate around the Gaussian distribution considering all the samples. In Figure 8, for the RC moment frame and the wood shear wall, both methods result in a similar variation of standard deviation. For the steel moment frame, while method 1 does not result in any variation, method 2 predicts a subtle variation around the constant standard deviation that lacks a specific trend. Method 1 tries to fit a smooth trend to the standard deviation variation. In contrast, method 2 infers the standard deviation variation using the neighborhood of data points close to the input IM. This can make method 2 sensitive to the standard deviation variation more so than method 1. Nevertheless, whether or not these subtle differences in the standard deviation variations will have an impact on the fragility functions is explored subsequently.
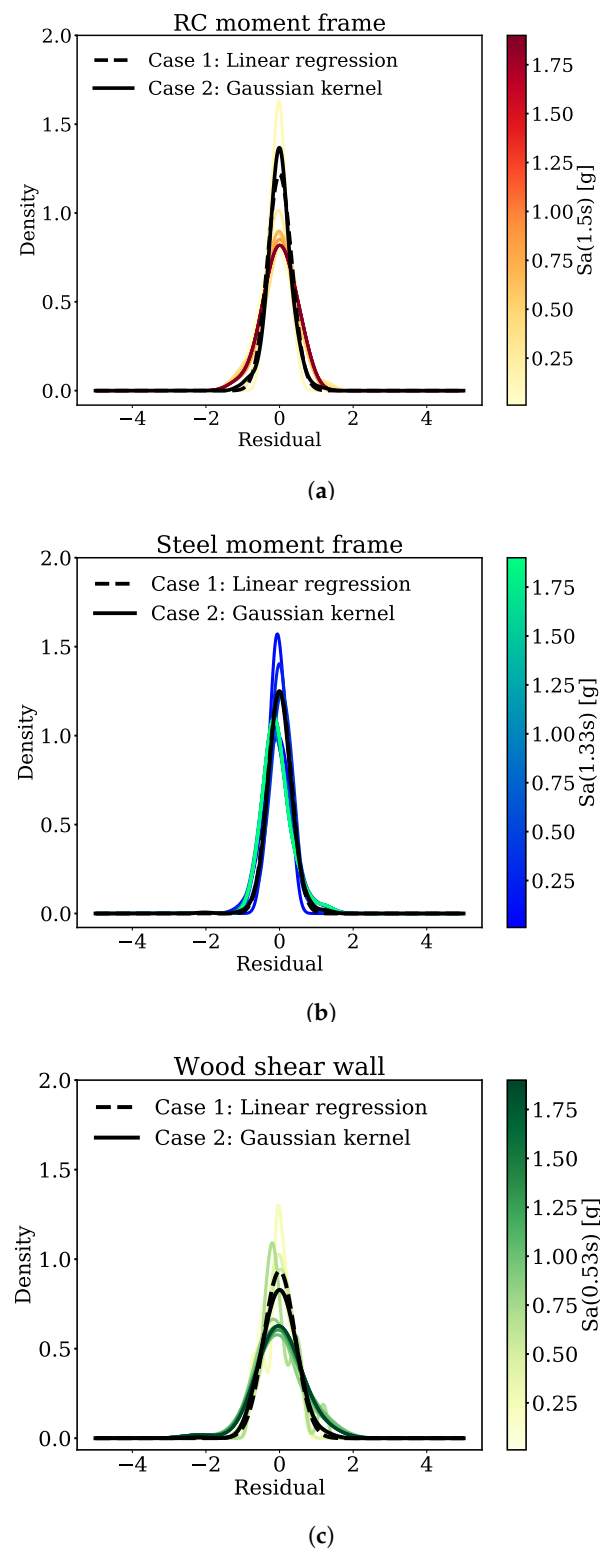
(**a**)

(**b**)

(**c**)

Figure 7. Variation of the distribution of the PID residuals as a function of the IM level characterized using KNN kernel density (the colored plots). The black dashed and solid plots, respectively, represent constant residuals distribution characterized by a Gaussian distribution and a Gaussian kernel density.
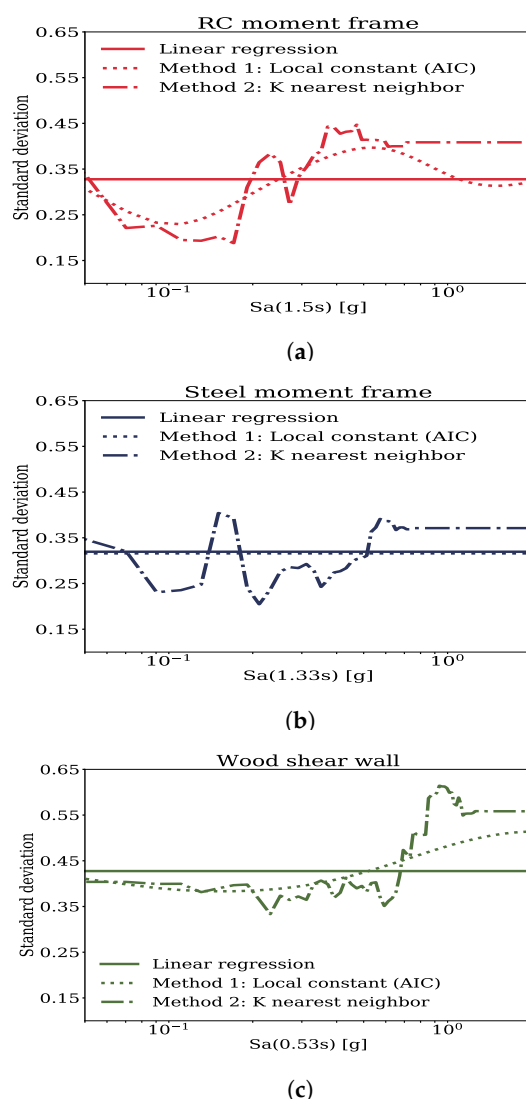
(a)

(b)

(c)

Figure 8. Variation of standard deviation with the IM level predicted by local constant regression (AIC optimal bandwidth) and KNN algorithm. The constant standard deviation predicted by linear regression is also presented.

4. Impacts on Fragility Functions

In this section, the impacts of a data-driven approach to predict median PID, standard deviation, and the distribution of residuals on the fragility functions are explored.

4.1. Cases and Drift Limits for Fragility Evaluation

Five cases for developing the fragility functions, as presented in Table 1, are considered. The consideration of these fives cases allows us to systematically explore the influence of relieving the assumptions made by linear regression on the fragility functions. Case 1 is the standard linear regression analysis employed by most studies. Case 2 alleviates the assumptions that the PID-IM relationship is linear in a log-log space and that the residuals are Gaussian distributed; however, the standard deviation is still assumed to be constant across the IM levels. From the discussion corresponding to Figure 7, characterizing the distribution of all the local linear residuals either as a Gaussian or a kernel density is expected to have the same influence on fragilities. Case 3 alleviates the constant standard deviation assumption, but it assumes that residuals follow a Gaussian distribution at all IM levels. Case 4 is similar to Case 3, except that the KNN method is used to compute the standard deviation variation. Case 5 alleviates all the assumptions made by linear regression: log-linear relationship, constant standard deviation, and Gaussian distributed residuals.

Table 1. Cases for fragility and downtime hazard evaluation.

	Predictive Model for Mean Response	Predictive Model for Standard Deviation	Distribution of Residuals	Plot Type
Case 1	Log-linear	Constant	Gaussian distribution	Circles
Case 2	Kernel regression: local linear (LS-CV)	Constant	Gaussian kernel density	Dashdot
Case 3	Kernel regression: local linear (LS-CV)	Kernel regression: local constant (AIC)	Gaussian distribution	Dotted
Case 4	Kernel regression: local linear (LS-CV)	K nearest neighbor	Gaussian distribution	Dashed
Case 5	Kernel regression: local linear (LS-CV)	K nearest neighbor	Gaussian kernel density	Solid

Notes. LS-CV: Least Squares Cross Validation bandwidth; AIC: Akaike Information Criterion bandwidth.

For computing the fragility functions at different damage states, structural drift limits are necessary. These drift limits, presented in Table 2, were inferred from the HAZUS technical manual [39]. In line with the discussion in Section 2.1, the high-code drift limits are used for the steel moment frame and the wooden shear wall. For the RC moment frame, the pre-code limits are considered. In addition, only for the Complete damage states, the high-code limit is considered to explore the differences between the five cases (Table 1) at larger PID levels.

Table 2. Damage states and drift limits for fragility evaluation.

Structure Type	DS 1: Slight	DS 2: Moderate	DS 3: Extensive	DS 4: Complete
RC moment frame	0.0027	0.0043	0.0107	0.0267 0.0533 (High)
Steel moment frame	0.004	0.008	0.02	0.0533
Wood shear wall	0.004	0.012	0.04	0.1

Notes: Damage State is denoted by DS ● These drift limits were inferred from FEMA [39] ● RC moment frame: Pre-code limits; Steel moment frame and wood shear wall: High code limits ● For the RC moment frame, under the Complete damage state, the high-code limit is also considered.

4.2. Computing the Fragility Functions Using the Different Cases

Case 1 in Table 2 is the standard linear regression. Hence, Equation (1) can be directly used to compute the fragilities. Even for Cases 3 and 4, Equation (1) can be still used with a couple of modifications: (1) the median PID as a function of the IM ($\ln \overline{PID}$) is predicted using local linear regression (Equation (3)) instead of the standard linear regression; (2) the standard deviation (σ) is also a function of the IM instead of being a constant. For Cases 2 and 5, since the residuals distribution is characterized as a Gaussian kernel density and not as a Gaussian distribution, applying Equation (1) is infeasible. Alternatively, the following equation can be used to evaluate the fragilities for these two cases:

$$P(PID > y \mid IM) = 1 - \int_{-\infty}^{\varepsilon} f(\varepsilon) \, d\varepsilon \tag{11}$$

where ε is prediction residual (i.e., $\ln y - \ln \overline{PID}$) and $f(.)$ is kernel density estimate obtained from Equation (9). While for Case 2, all the prediction residuals are used in the above equation, for Case 5, only those residuals that are close to the input IM level, as inferred by the KNN algorithm, are used.

4.3. Results

Figure 9 presents the fragility functions for the five cases discussed earlier considering the moderate, severe, and complete damage states. For the RC moment frame (Figure 9a–c), considering the pre-code drift limits, all five cases are resulting in similar fragility estimates. Case 5 is resulting in a non-smooth estimate of the fragility due to a data-driven characterization of the median PID, standard deviation, and distribution of residuals. If analysts prefer a smooth fragility, they may consider the use of smoothing functions on the fragility resulting from Case 5. The pre-code drift limits are small such that they do not reveal convincingly the impacts of alleviating the assumptions made by linear regression. This is where the high-code drift limit applied to the complete damage state (Figure 9c) may help. Fragilities from Cases 2–5 are encompassing Case 1 fragility. This is due to local linear regression predicting lager median PIDs at high IM levels than Case 1 (also refer Figure 4a). While Cases 2–5, which are only different in terms how they characterize the PID probability density, are resulting in somewhat similar estimates of the fragility, Case 5 appears to be a bit more different than the others. This implies that the compounded effects of alleviating all the assumptions of linear regression can have more impact on the fragility than alleviating any of those individually.
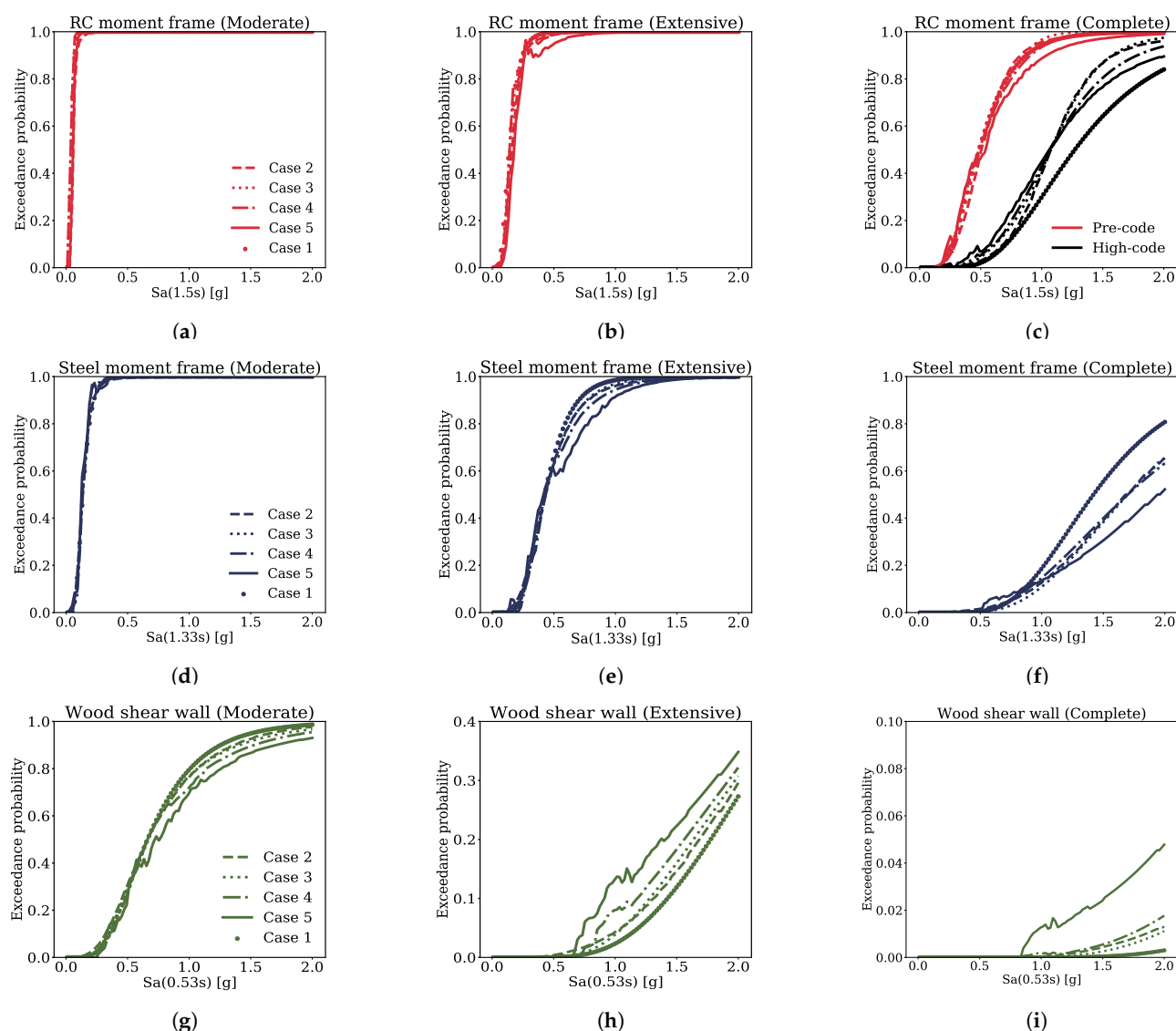

Figure 9. Fragility functions for the different structure types and damage states. For a description of the cases for fragility evaluation, refer to Table 1.

For the steel moment frame (Figure 9d–h), all five cases are resulting in similar fragilities under the moderate and extensive damage states. It is for the complete damage state that we observe some differences (Figure 9f). Case 1 is encompassing Cases 2–5 due to local linear regression predicting lesser median PIDs at large IM levels (refer Figure 4b). Case 5 is resulting in slightly different estimate of fragility than Cases 2–3. This is again due to the compounded effects of non-constant standard deviation along with non-Gaussian distribution of PID residuals having more influence on the fragility than any of those have individually. Even for the wood shear wall (Figure 9g–i), a similar conclusion d holds. However, unlike the previous two cases, the differences in fragilities are not significantly influenced by the median PID predictions since it was noted in Figure 3c that both linear and local linear regressions are very similar. These differences in fragilities are influenced by the non-constant standard deviation and the non-Gaussian distribution of residuals. Furthermore, it is only for the wooden shear wall that we observe non-negligible differences between the fragilities under the extensive damage state. The higher drift limits given a damage state than the other two archetypes (Table 2) contribute to these differences.

The fragilities for Cases 3 and 4 are quite similar across the archetype-damage state combinations. This indicates that, if the all the prediction residuals are characterized as a Gaussian distribution, inferring the standard deviation variation using either local constant regression or KNN method results in negligible differences in the corresponding fragilities. This inference also means that, any differences between the fragilities of Case 3 (or 4) and Case 5 are due to the non-Gaussian distribution of residuals given an IM level. Compared to the RC moment frame, the steel moment frame and the wood shear wall exhibit more differences between Case 3 (or 4) and Case 5. This observation aligns with the discussion made corresponding to Figure 7 that the residuals distributions for the RC moment frame follow a Gaussian distribution more faithfully at different IM levels than the other two archetypes.

To summarize, two main conclusions can be drawn: (1) the compounded effect of alleviating the assumptions made by linear regression is greater than that of the individual effects; (2) the linear regression assumptions influence the extensive damage state to some extent and the complete damage state to a greater extent, but may not influence the other damage states, in general. The impacts of these conclusions on the demand and loss hazards are next investigated.

5. Impacts on Seismic Risk Quantified through the Demand and Loss Hazards

5.1. Results: Demand Hazard

Demand hazard given the fragility functions is computed using:

$$\lambda(PID > y) = \int P(PID > y | IM = im) \, |d\lambda(IM > im)| \tag{12}$$

where $\lambda(PID > y)$ and $\lambda(IM > im)$ are the demand and seismic hazards expressed as probability of exceedances in fifty years. Seismic hazard curves for spectral acceleration at the oscillator periods of the archetypes were inferred from the OpenSHA software [40] for a site in Los Angeles, CA (34.053° N; 118.243° W). Ground motion prediction equation proposed by Boore and Atkinson 2008 [41] was used assuming a shear wave velocity averaged over the top 30 m depth as 420 m/s. The USGS/CGS 1996 Adj. Cal. Earthquake Rupture Forecast was used. These hazard curves are presented in Figure 10.
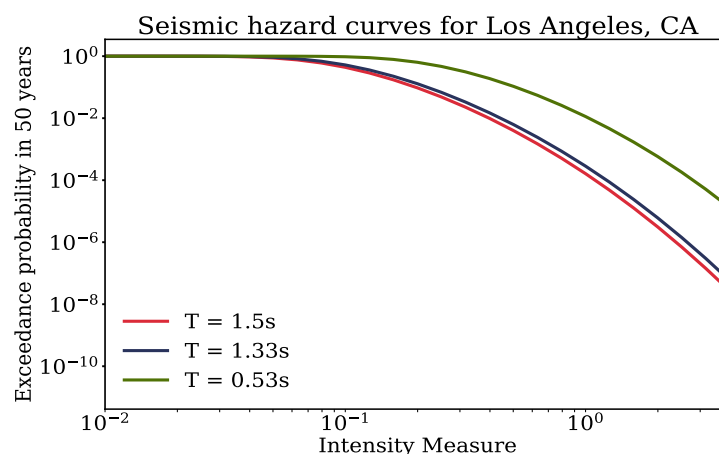
Figure 10. Seismic hazard curves for a site in Los Angeles, CA, for the oscillator periods 1.5 s, 1.33 s, and 0.53 s which are the fundamental periods of the RC frame, steel frame, and wood shear wall, respectively.

Figure 11 presents the demand hazards for the archetype buildings considering the five cases for fragility evaluation discussed in Table 1. It is noted that for PID values less than about 0.02, all the five cases are leading to quite similar demand hazards across the archetypes. For greater values of PID, demand hazards for the RC frame and the wood shear wall show some differences between the five cases. Results corresponding to case 5 for these two archetypes align with the conclusion drawn from Section 4.3 that the compounded effect of alleviating the linear regression assumptions is greater than the individual effects. Demand hazards for the steel moment frame, however, show lesser variation between the five cases. Differences in the demand hazards between the five cases depend on the IM levels at which differences in the fragility functions manifest. If such differences in the fragility functions appear at larger IM levels, they tend to be downplayed by the integration with the seismic hazard in Equation (12).

5.2. Results: Loss Hazard

Loss hazard with downtime as the decision variable is expressed as the probability of exceeding a downtime value in fifty years. The PEER equation for PBEE is used for computing the loss hazard [17]:

$$\lambda\left(T^* > t\right) = \int_{im} \sum_{i=1}^{N_{ds}} P(T^* > t | DS = ds_i) \, P(DS = ds_i | IM = im) \, |d\lambda(IM > im)| \quad (13)$$

where T^* is the downtime and t is the required downtime level, $P(T^* > t | DS = ds_i)$ is the conditional probability of exceeding a required downtime level, N_{ds} is the number of damage states, and $\lambda(IM > im)$ is the seismic hazard expressed as a probability of exceeding an IM level in fifty years. The quantity $P(T^* > t | DS = ds_i)$ is estimated by assuming that the downtime given a damage state is log-Normally distributed. The mean downtime values for the slight, moderate, severe, and complete damage states are taken as 20, 90, 360, and 480 days, respectively. These values were obtained from the HAZUS technical manual for commercial buildings [39]. In line with [3], a log standard deviation of 0.75 is assumed for all the damage states.
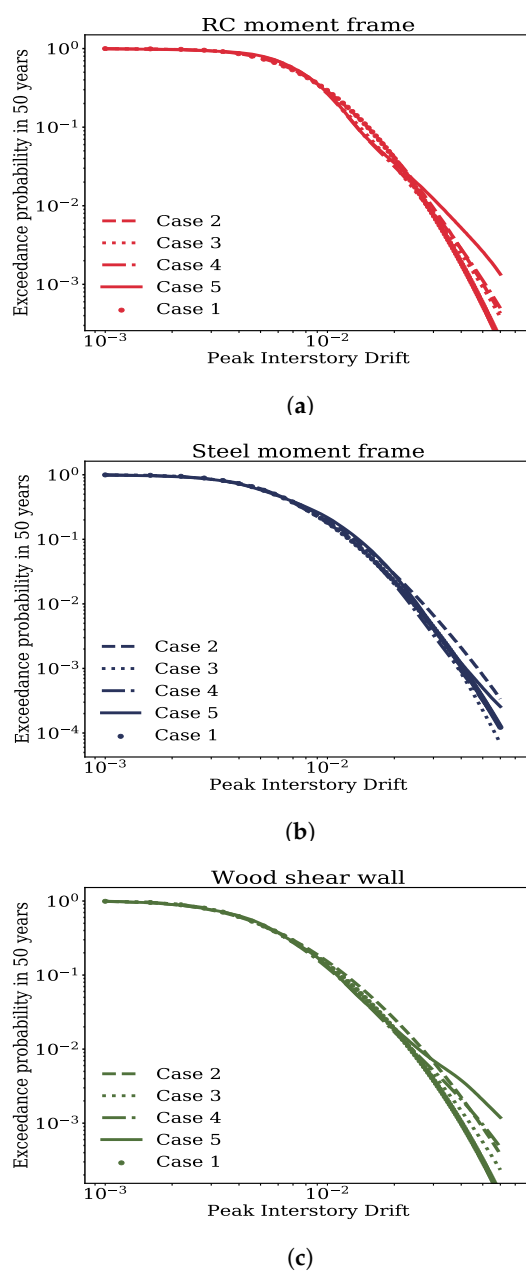
(a)

(b)

(c)

Figure 11. Demand hazard for the three considered archetypes. For a description of the cases for demand hazard evaluation, refer to Table 1.

Figure 12 presents the loss hazard in terms of the probability of exceeding a downtime level in 50 years. Five cases for fragility evaluation discussed in Table 1 are considered. At lower downtime levels, insignificant differences in the hazard values between the five cases are observed across all the three archetype buildings. At higher downtime levels too, for the RC and steel moment frames, differences in the loss hazard values between the five cases are still negligible. For the wood shear wall, while there are some differences between the five cases, these differences are not strong.
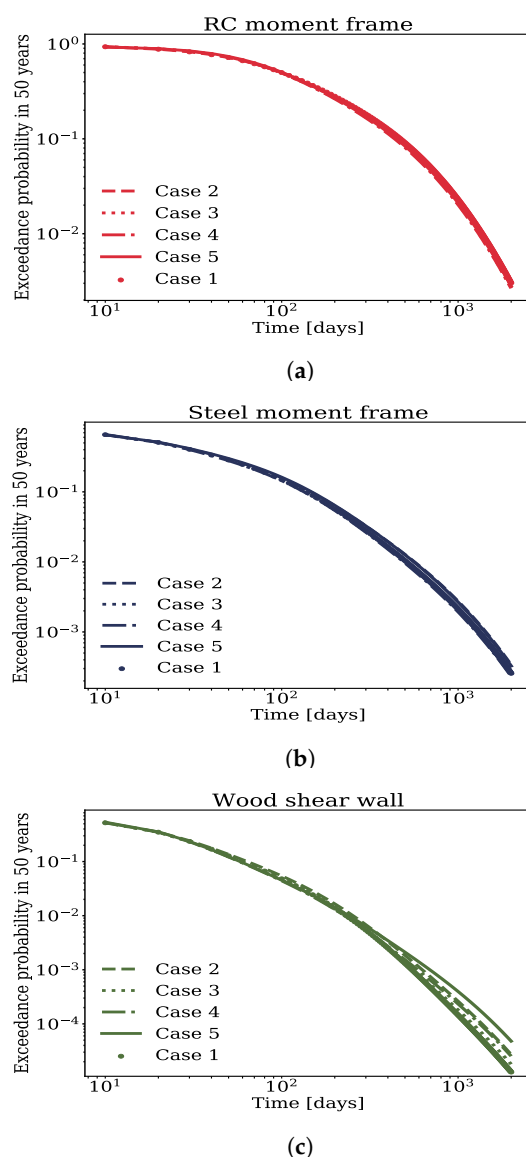
(a)

(b)

(c)

Figure 12. Downtime hazard the three considered archetypes. For a description of the cases for downtime hazard evaluation, refer to Table 1.

For all practical purposes and in general, we argue that the differences in the loss hazards between the five cases is not very significant. Deaggregation of the the loss hazard is performed to reveal more insights into the influence of the differences in fragilities on the loss hazard. The probability mass in any of the four damage states conditional upon the downtime level is given by:

$$P(DS = ds_i | T^* > t) = \frac{\int_{im} P(T^* > t | DS = ds_i) \ P(DS = ds_i | IM = im) \ |d\lambda(IM > im)|}{\lambda(T^* > t)} \quad (14)$$

Figure 13 presents the deaggregation probability masses for the five cases. The downtime level considered is 2000 days. Across the three archetype buildings, most of the probability mass is concentrated in the severe damage state. For this damage state, it was noted in Figure 9b,e that the differences between the fragilities are not significant for the RC and steel moment frames, respectively. For the wood shear wall (Figure 9h), difference between the Cases 1 and 5 were observed, and these differences subtly manifest in the deaggregation probabilities (Figure 13c) as well as the loss hazard (Figure 12c).
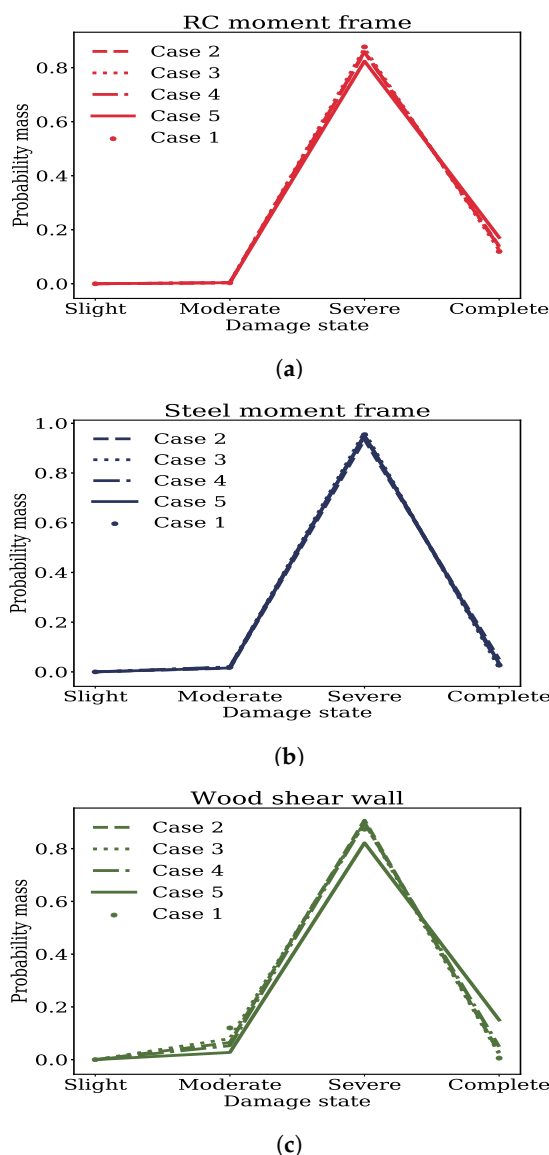
(a)

(b)

(c)

Figure 13. Deaggregation of the loss hazard with respect to the damage states for a downtime level of 2000 days. For a description of the cases for downtime hazard evaluation, refer to Table 1.

In relation to the influence of differences in the fragility functions on the loss hazard, two main conclusions can be drawn: (1) Although considerable differences between the fragilities were observed for the Complete damage state, their influence on the downtime hazard is subtle, in general; (2) This is due to the concentration of most of the probability mass in the lower damage states, given a downtime level. For damage states lower than the Complete state, overall marginal differences between the fragilities were noted earlier.

6. Summary and Conclusions

Cloud analysis is a popular method for the seismic fragility and risk assessment. This method mostly uses linear regression which assumes that the: median relationship between seismic response and intensity is linear in a log-log space; standard deviation is constant across different intensity levels; and distribution of response residuals follows a Gaussian distribution. This paper has evaluated the impacts of systematically alleviating these assumptions on the fragility functions and seismic risk, evaluated through the demand and loss hazard (downtime is the decision variable). Gaussian kernel techniques were employed as an alternative to linear regression for analyzing the seismic response data.

Three common archetype buildings were considered which include an RC frame, a steel frame, and a wood shear wall. A generic set of 380 ground motions composed mostly of large magnitude and small distance records is used. The following are the highlights with regard to the application of Gaussian kernel techniques:

- For characterizing the relation between response and intensity, local linear regression with LS-CV bandwidth criterion was considered since it made more physical predictions of response within the intensity bounds of interest.
- Variation of standard deviation with the intensity was captured using local constant regression with AIC bandwidth criterion since it predicted constant estimates of standard deviation beyond the data bounds and this is more conservative.
- Distribution of residuals were characterized using a Gaussian kernel density given the residuals closest to the input intensity level. These closest residuals were selected using the K Nearest Neighbor algorithm with a conservative value for K to avoid over-fitting.

In relation to the impacts on fragility functions and risk, the following general observations were made:

- The compounded effects of alleviating the assumptions made by linear regression on fragilities is more significant than any of those individually.
- Alleviating the linear regression assumptions impact the Complete damage state to a significant extent and the Severe damage state to a marginal extent but not any of the lower damage states.
- For all practical purposes, linear regression assumptions seem to have lesser impacts on the loss hazard, even for large downtime levels. Deaggregation of the loss hazard at a large downtime level revealed that these subtle impacts are due to the concentration of probability mass mostly in the Severe damage state. For this damage state, it was noted that the assumptions made by linear regression had marginal impacts on the fragility functions.

If the goal is of a seismic analysis is to infer the loss hazard using the PBEE framework, linear regression may still be adequate. However, if the goal is to infer the fragility functions and demand hazard accurately, the assumptions made by linear regression can have some impacts on the higher damage states and larger PID levels, respectively. Finally, more widespread adoption of advanced statistical approaches, such as the ones presented in this paper, are need for efficient and accurate characterization of the seismic responses of buildings and the subsequent prediction of seismic risk. However, the practical application of state of the art research methods is often slow and bridging this gap is a crucial component for the future.

Institutional Review Board Statement: Not applicable to this paper.

Informed Consent Statement: Not applicable to this paper.

Data Availability Statement: Data will be made available upon request.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Bruneau, M.; Chang, S.; Eguchi, R.; Lee, G.; O'Rourke, T.; Reinhorn, A.; Shinozuka, M.; Tierney, K.; Wallace, W.; Von Winterfeldt, D. A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthq. Spectra* **2003**, *4*, 733–752. [CrossRef]
2. Cimellaro, G.; Reinhorn, A.; Bruneau, M. Framework for analytical quantification of disaster resilience. *Eng. Struct.* **2010**, *32*, 3639–3649. [CrossRef]
3. Burton, H.; Deierlein, G.; Lallemant, D.; Lin, T. Framework for Incorporating Probabilistic Building Performance in the Assessment of Community Seismic Resilience. *J. Struct. Eng.* **2016**, *142*, C4015007. [CrossRef]
4. Dhulipala, S.; Flint, M. Series of semi-Markov processes to model infrastructure resilience under multihazards. *Reliab. Eng. Syst. Saf.* **2020**, *193*, 106659. [CrossRef]

5. Kia, M.; Banazadeh, M.; Bayat, M. Rapid seismic vulnerability assessment by new regression-based demand and collapse models for steel moment frames. *Earthq. Struct.* **2018**, *14*, 203–214.

6. Kia, M.; Banazadeh, M.; Bayat, M. Rapid seismic loss assessment using new probabilistic demand and consequence models. *Bull. Earthq. Eng.* **2019**, *17*, 3545–3572. [CrossRef]

7. Vamvatsikos, D.; Cornell, A.C. Incremental dynamic analysis. *Earthq. Eng. Struct. Dyn.* **2002**, *31*, 491–514. [CrossRef]

8. Jalayer, F.; Cornell, C. Alternative non-linear demand estimation methods for probability-based seismic assessments. *Earthq. Eng. Struct. Dyn.* **2009**, *38*, 951–972. [CrossRef]

9. Baker, J.W. Efficient Analytical Fragility Function Fitting Using Dynamic Structural Analysis. *Earthq. Spectra* **2015**, *31*, 579–599. [CrossRef]

10. Luco, N.; Cornell, C.A. Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. *Earthq. Spectra* **2007**, *23*, 357–392. [CrossRef]

11. Jalayer, F. Direct Probabilistic Seismic Analysis: Implementing Non-Linear Dynamic Assessments. Ph.D. Thesis, Stanford University, Menlo Park, CA, USA, 2003.

12. Luco, N.; Bazzurro, P. Does amplitude scaling of ground motion records result in biased nonlinear structural drift responses? *Earthq. Eng. Struct. Dyn.* **2007**, *36*, 1813–1835. [CrossRef]

13. Dávalos, H.; Miranda, E. Evaluation of bias on the probability of collapse from amplitude scaling using spectral-shape-matched records. *Earthq. Eng. Struct. Dyn.* **2019**, *48*, 970–986. [CrossRef]

14. Tubaldi, E.; Freddi, F.; Barbato, M. Probabilistic seismic demand model for pounding risk assessment. *Earthq. Eng. Struct. Dyn.* **2016**, *45*, 1743–1758. [CrossRef]

15. Modica, A.; Stafford, P. Vector fragility surfaces for reinforced concrete frames in Europe. *Bull. Earthq. Eng.* **2014**, *12*, 1725–1753. [CrossRef]

16. Chu, M.; Katerina, K.; Sudret, B. Seismic fragility curves for structures using non-parametric representations. *Front. Struct. Civ. Eng.* **2017**, *11*, 169–186.

17. Moehle, J.; Deierlein, G.G. A framework methodology for performance-based earthquake engineering. In Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver, BC, Canada, 1–6 August 2004.

18. Trifunac, M.; Ivanović, S.; Todorovska, M. *Instrumented 7-Story Reinforced Concrete Building in Van Nuys, California: Description of the Damage from the 1994 Northridge Earthquake and Strong Motion Data*; Technical Report; University of Southern California: Los Angeles, CA, USA, 1999.

19. Krawinkler, H. *Van Nuys Hotel Building Testbed Report: Exercising Seismic Performance Assessment*; Technical Report; Pacific Earthquake Engineering Research (PEER) Center: Berkeley, CA, USA, 2005.

20. Mazzoni, S.; McKenna, F.; Scott, M.H.; Fenves, G.L. *Open System for Earthquake Engineering Simulation (OpenSEES)*; Technical Report; Pacific Earthquake Engineering Research (PEER) Center: Berkeley, CA, USA, 2006.

21. Kalkan, E. Available online: www.quakelogic.net/research (accessed on 15 August 2019).

22. Paspuleti, C. Seismic Analysis of an Older Reinforced Concrete Frame. Master's Thesis, University of Washington, Washington, DC, USA, 2002.

23. Lignos, D. Sideway Collapse of Deteriorating Structural Systems under Seismic Excitations. Ph.D. Thesis, Stanford University, Menlo Park, CA, USA, 2008.

24. Eads, L. Seismic Collapse Risk Assessment of Buildings: Effects of Intensity Measure Selection and Computational Approach. Ph.D. Thesis, Stanford University, Menlo Park, CA, USA, 2013.

25. Ibarra, L.F.; Medina, R.A.; Krawinkler, H. Hysteretic models that incorporate strength and stiffness deterioration. *Earthq. Eng. Struct. Dyn.* **2005**, *34*, 1489–1511. [CrossRef]

26. Jayamon, J.R. Seismic Performance Assessment of Wood-Frame Shear Wall Structures. Ph.D. Thesis, Virginia Tech, Blacksburg, VA, USA, 2017.

27. Folz, B.; Filiatrault, A. *SAWS—Version 1.0, A Computer Program for the Seismic Analysis of Woodframe Structures*; Technical Report; Consortium of Universities for Research in Earthquake Engineering: Richmond, CA, USA, 2001.

28. Ancheta, T.; Darragh, R.; Stewart, J.; Seyhan, E.; Silva, W.; Chiou, B.; Wooddell, K.; Graves, R.; Kottke, A.; Boore, D.; et al. NGA-West2 database. *Earthq. Spectra* **2014**, *30*, 989–1005. [CrossRef]

29. Sehhati, R.; Rodriguez-Marek, A.; ElGawady, M.; Cofer, W. Effects of near-fault ground motions and equivalent pulses on multi-story structures. *Eng. Struct.* **2011**, *33*, 767–779. [CrossRef]

30. Archila, M.; Ventura, C.E.; Liam Finn, W.D. New insights on effects of directionality and duration of near-field ground motions on seismic response of tall buildings. *Struct. Des. Tall Spec. Build.* **2017**, *26*, e1363. [CrossRef]

31. Dhulipala, S.; Rodriguez-Marek, A.; Ranganathan, S.; Flint, M. A site-consistent method to quantify sufficiency of alternative IMs in relation to PSDA. *Earthq. Eng. Struct. Dyn.* **2018**, *47*, 377–396. [CrossRef]

32. Dhulipala, S. Bayesian Methods for Intensity Measure and Ground Motion Selection in Performance-Based Earthquake Engineering. Ph.D. Thesis, Virginia Tech, Blacksburg, VA, USA, 2019.

33. Racine, J. Nonparametric econometrics: A primer. *Found. Trends Econom.* **2008**, *3*, 1–88. [CrossRef]

34. Noh, H.; Lallemant, D.; Kiremidjian, A. Development of empirical and analytical fragility functions using kernel smoothing methods. *Earthq. Eng. Struct. Dyn.* **2015**, *44*, 1163–1180. [CrossRef]

35. Loader, C. Bandwidth selection: Classical or plug-in? *Ann. Stat.* **1999**, *27*, 415–438. [CrossRef]

36. Al Mamun, A.; Saatcioglu, M. Seismic fragility analysis of pre-1975 conventional concrete frame buildings in Canada. *Can. J. Civ. Eng.* **2018**, *45*, 728–738. [CrossRef]

37. Muller, H.; Stadtmuller, U. Variable bandwidth kernel estimators of regression curves. *Annal. Stat.* **1987**, *15*, 182–201. [CrossRef]

38. Hassanat, A.; Abbadi, M.; Altarawneh, G.; Alhasanat, A. Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach. *arXiv* **2014**, arXiv:1409.0919.

39. FEMA. *Multi-Hazard Loss Estimation Methodology Earthquake Model (HAZUS-MH MR5): Technical Manual*; Technical Report; Federal Emergency Management Agency: Washington, DC, USA, 2010.

40. Field, E.; Jordan, T.; Cornell, C. OpenSHA: A developing community-modeling environment for seismic hazard analysis. *Seismol. Res. Lett.* **2003**, *74*, 406–419. [CrossRef]

41. Boore, D.M.; Atkinson, G.M. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthq. Spectra* **2008**, *24*, 99–138. [CrossRef]