# Chapter 4
# Variability and Uncertainty in Empirical Ground-Motion Prediction for Probabilistic Hazard and Risk Analyses

**Peter J. Stafford**

**Abstract**  The terms aleatory variability and epistemic uncertainty mean different things to people who routinely use them within the fields of seismic hazard and risk analysis. This state is not helped by the repetition of loosely framed generic definitions that actually inaccurate. The present paper takes a closer look at the components of total uncertainty that contribute to ground-motion modelling in hazard and risk applications. The sources and nature of uncertainty are discussed and it is shown that the common approach to deciding what should be included within hazard and risk integrals and what should be pushed into logic tree formulations warrants reconsideration. In addition, it is shown that current approaches to the generation of random fields of ground motions for spatial risk analyses are incorrect and a more appropriate framework is presented.

## 4.1  Introduction

Over the past few decades a very large number of empirical ground-motion models have been developed for use in seismic hazard and risk applications throughout the world, and these contributions to engineering seismology collectively represent a significant body of literature. However, if one were to peruse this literature it would, perhaps, not be obvious what the actual purpose of a ground-motion model is. A typical journal article presenting a new ground-motion model starts with a brief introduction, proceeds to outlining the dataset that was used, presents the functional form that is used for the regression analysis along with the results of this analysis, shows some residual plots and comparisons with existing models and then wraps up with some conclusions. In a small number of cases this pattern is broken by the authors giving some attention to the representation of the standard deviation of the model. Generally speaking, the emphasis is very much upon the development and

P.J. Stafford (✉)

Department of Civil & Environmental Engineering, Imperial College London, London, UK
e-mail: p.stafford@imperial.ac.uk

behaviour of the median predictions of these models and the treatment of the standard deviation (and its various components) is very minimal in comparison. If it is reasonable to suspect that this partitioning of effort in presenting the model reflects the degree of effort that went into developing the model then there are two important problems with this approach: (1) the parameters of the model for the median predictions are intrinsically linked to the parameters that represent the standard deviation – they cannot be decoupled; and (2) it is well known from applications of ground-motion models in hazard and risk applications that the standard deviation exerts at least as much influence as the median predictions for return periods of greatest interest.

The objective of the present article is to work against this trend by focussing almost entirely upon the uncertainty associated with ground-motion predictions. Note that what is actually meant by 'uncertainty' will be discussed in detail in subsequent sections, but the scope includes the commonly referred to components of aleatory variability and epistemic uncertainty. Furthermore, the important considerations that exist when one moves from seismic hazard analysis into seismic risk analysis will also be discussed.

As noted in the title of the article, the focus herein is upon empirical ground-motion models and discussion of the uncertainties associated with stochastic simulation-based models, or seismological models is not within the present scope. That said, some of the concepts that are dealt with herein are equally applicable to ground-motion models in a more general sense.

While at places in the article reference will be made to peak ground acceleration or spectral acceleration, the issues discussed here at not limited to these intensity measures. For the particular examples that are presented, although the extent of various effects will be tied to the choice of intensity measure, the emphasis is upon the underlying concept rather than the numerical results.

## 4.2   Objective of Ground-Motion Prediction

In both hazard and risk applications the objective is usually to determine how frequently a particular state is exceeded. For hazard, this state is commonly a level of an intensity measure at a site, while for risk applications the state could be related to a level demand on a structure, a level of damage induced by this demand, or the cost of this damage and its repair, among others. In order to arrive at estimates of these rates (or frequencies) of exceedance it is not currently possible to work with empirical data related to the state of interest as a result of insufficient empirical constraint. For example, if one wished to compute an estimate of the annual rate at which a level of peak ground acceleration is exceeded at a site then an option in an ideal world would be to assume that the seismogenic process is stationary and that what has happened in the past is representative of what might happen in the future. On this basis, counting the number of times the state was exceeded and dividing this by the temporal length of the observation period would provide an estimate of the

exceedance rate. Unfortunately, there is not a location on the planet for which this approach would yield reliable estimates for return periods of common interest.

To circumvent the above problem hazard and risk analyses break down the process of estimating rates of ground-motions into two steps: (1) estimate the rates of occurrence of particular earthquake events; and (2) estimate the rate of exceedance of a particular state of ground motion *given* this particular earthquake event. The important point to make here is that within hazard and risk applications the role of an empirical ground-motion model is to enable this second step in which the rate of exceedance of a particular ground-motion level is computed *for a given earthquake scenario*. The manner in which these earthquake scenarios are (or can be) characterised has a strong impact upon how the ground-motion models can be developed. For example, if the scenario can only be characterised by the magnitude of the event and its distance from the site then it is only meaningful to develop the ground-motion model as a function of these variables.

To make this point more clear, consider the discrete representation of the standard hazard integral for a site influenced by a single seismic source:

$$\lambda_{Y>y*} = \nu \sum_{k=1}^{K} \sum_{j=1}^{J} P\left[Y > y* | m_j, r_k\right] P\left[M = m_j, R = r_k\right] \tag{4.1}$$

where, $Y$ is a random variable representing the ground-motion measure of interest, $y*$ is a particular value of this measure, $\nu$ is the annual rate of occurrence of earthquakes that have magnitudes greater than some minimum value of interest, and $M$ and $R$ generically represent magnitude and distance, respectively. If we factor out the constant parameter $\nu$, then we have an equation in terms of probabilities and we can see that the objective is to find:

$$\begin{aligned} P[Y > y*] &= \frac{\lambda_{Y>y*}}{\nu} = \sum_{k=1}^{K} \sum_{j=1}^{J} P\left[Y > y* | m_j, r_k\right] P\left[M = m_j, R = r_k\right] \\ &= \int_{y*}^{\infty} f_Y(y) dy = \iint \int_{y*}^{\infty} f_{Y|m,r}(y|m,r) f_{M,R}(m,r) dm dr \end{aligned} \tag{4.2}$$

When we discuss the uncertainty associated with ground-motion models it is important to keep this embedding framework in mind. The framework shows that the role of a ground-motion model is to define the distribution $f_{Y|m,r}(y|m,r)$ of levels of motion that can occur for a given earthquake scenario, defined in this case by $m$ and $r$. The uncertainty that is ultimately of interest to us relates to the estimate of $P[Y > y*]$ and this depends upon the uncertainty in the ground-motion prediction as well as the uncertainty in the definition of the scenario itself.

For seismic hazard analysis, the ground-motion model alone is sufficient to provide the univariate distribution of the intensity measure for a given earthquake scenario. However, for seismic risk applications, a typical ground-motion model may need to be coupled with a model for spatial, and potentially spectral,

correlations in order to define a multivariate conditional distribution of motions at multiple locations (and response periods) over a region.

At a given site, both in hazard and risk applications, the conditional distribution of ground-motions (assuming spectral acceleration as the intensity measure) given a scenario is assumed to be lognormal and is defined as:

$$\ln Sa \sim N\left(\mu_{\ln Sa}, \sigma_{\ln Sa}^2\right) \tag{4.3}$$

where the moments of the distribution are specific to the scenario in question, *i.e.*, $\mu_{\ln Sa} \equiv \mu_{\ln Sa}(m, r, \ldots)$ and $\sigma_{\ln Sa} \equiv \sigma_{\ln Sa}(m, r, \ldots)$. The probability of exceeding a given level of motion for a scenario is therefore defined using the cumulative standard normal distribution $\Phi(z)$:

$$P\left[Sa > Sa^*|m, r, \ldots\right] = 1 - \Phi\left[\frac{\ln Sa^* - \mu_{\ln Sa}}{\sigma_{\ln Sa}}\right] \tag{4.4}$$

The logarithmic mean $\mu_{\ln Sa}$ and standard deviation $\sigma_{\ln Sa}$ for a scenario would differ for hazard and risk analyses as in the former case one deals with the marginal distribution of the motions conditioned upon the given the scenario while in the latter case one works with the conditional distribution of the motions, conditioned upon both the given scenario and the presence of a particular *event term* for the scenario. That is, in portfolio risk analysis one works at the level of inter-event variability and intra-event variability while for hazard analysis one uses the total variability.

An empirical ground-motion model must provide values of both the logarithmic mean $\mu_{\ln Sa}$ and the standard deviation $\sigma_{\ln Sa}$ in order to enable the probability calculations to be made and these values must be defined in terms of the predictor variables $M$ and $R$, among potentially others. Both components of the distribution directly influence the computed probabilities, but can exert greater or lesser influence upon the probability depending upon the particular value of $\ln Sa^*$.

## 4.3   Impact of Bias in Seismic Hazard and Risk

Equation (4.4) is useful to enable one to understand how the effects of bias in ground-motion models would influence the contributions to hazard and risk estimates. The computation of probabilities of exceedance is central to both cases. Imagine that we assume that any given ground-motion model is biased for a particular scenario in that the predicted median spectral accelerations differ from an unknown true value by a factor $\gamma_\mu$ and that the estimate of the aleatory variability also differs from the true value by a factor of $\gamma_\sigma$. To understand the impact of these biases upon the probability computations we can express Eq. (4.4) with explicit
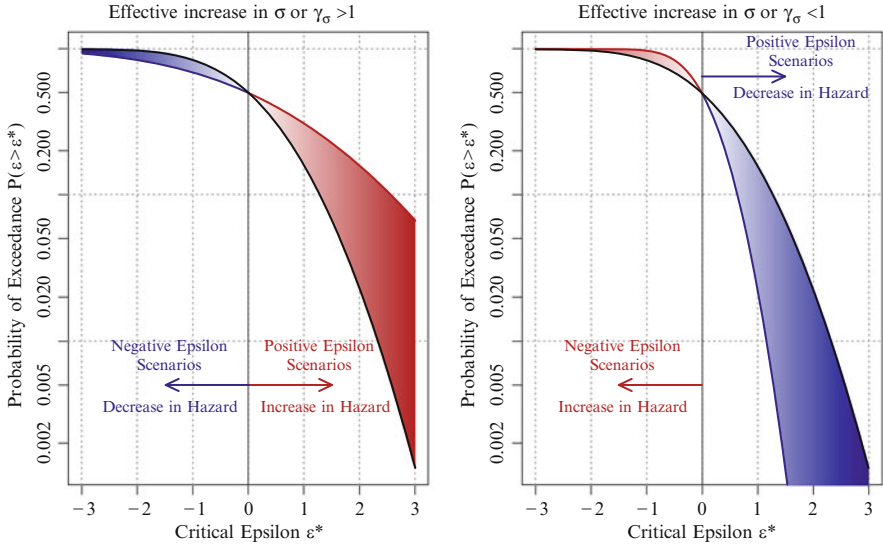
**Fig. 4.1** Illustration of the effect that a bias in the logarithmic standard deviation has upon the computation of probabilities of exceedance. The *left* panel corresponds to $\gamma_\sigma = 2$ while the *right* panel shows $\gamma_\sigma = 1/2$

inclusion of these bias factors as in Eq. (4.5). Now we recognise that the probability that we compute is an estimate and denote this as $\hat{P}$.

$$\hat{P}\left[Sa > Sa^* | m, r, \ldots\right] = 1 - \Phi\left[\frac{\ln Sa^* - \ln\gamma_\mu - \mu_{\ln Sa}}{\gamma_\sigma \sigma_{\ln Sa}}\right] \tag{4.5}$$

This situation is actually much closer to reality than Eq. (4.4). For many scenarios predictions of motions will be biased by some unknown degree and it is important to understand how sensitive our results are to these potential biases. The influence of the potential bias in the logarithmic standard deviation is shown in Fig. 4.1. The case shown here corresponds to an exaggerated example in which the bias factor is either $\gamma_\sigma = 2$ or $\gamma_\sigma = 1/2$.

What sort of bias could one expect to be reasonable for a given ground-motion model? This is a very difficult question to answer in any definitive way, but one way to get a feel for this is to compare the predictions of both median logarithmic motions and logarithmic standard deviations for two generations of modern ground-motion models. In particular, the very recent release of the models from the second phase of the PEER NGA project (NGA West 2) provides one with the ability to compare the predictions from the NGA West 1 and NGA West 2 studies.

Figures 4.2 and 4.3 show these estimates of the possible extent of bias for the ground-motion models of Campbell and Bozorgnia (2008, 2014) and Chiou and Youngs (2008, 2014). It should be noted that the point here is not that these models are necessarily biased, but that it is reasonable to assume that the 2014 versions are
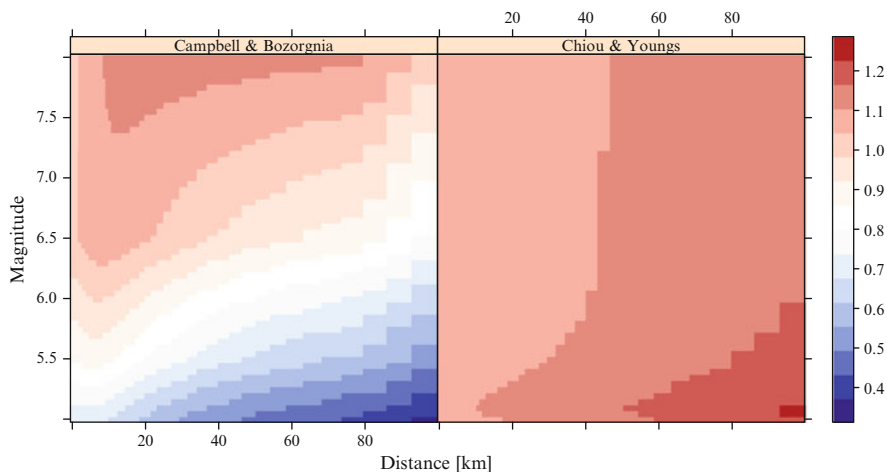
**Fig. 4.2** Example *bias factors* computed as the ratios between predictions of two generations of models from the same developers. The *left* panel shows ratios between the medians, $Sa(T = 0.01s)$, of Campbell and Bozorgnia (2014, 2008) – 2014:2008, while the *right* panel is for Chiou and Youngs (2014, 2008) – 2014:2008

less biased than their 2008 counterparts. Therefore, the typical extent of bias that has existed through the use of the 2008 NGA models over the past few years can be characterised through plots like those shown in Figs. 4.2 and 4.3. However, in order to see how these differences in predicted moments translate into differences in hazard estimates the following section develops hazard results for a simple academic example.

### 4.3.1 Probabilistic Seismic Hazard Analysis

A probabilistic seismic hazard analysis is conducted using the ground-motion models of Campbell and Bozorgnia (2008, 2014) as well as those of Chiou and Youngs (2008, 2014). The computations are conducted for a hypothetical case of a site located in the centre of a circular source. The seismicity is described by a doubly-bounded exponential distribution with a *b*-value of unity and minimum and maximum magnitudes of 5 and 8 respectively. The maximum distance considered in the hazard integrations is 100 km. For this exercise, the depths to the top of the ruptures for events of all magnitudes are assumed to be the same and it is also assumed that the strike is perpendicular to the line between the site and the closest point on the ruptures. All ruptures are assumed to be for strike-slip events and the site itself is characterised by an average shear-wave velocity over the uppermost 30 m of 350 m/s. Note that these assumptions are equivalent to ignoring finite source dimensions and working with a point-source representation. For the
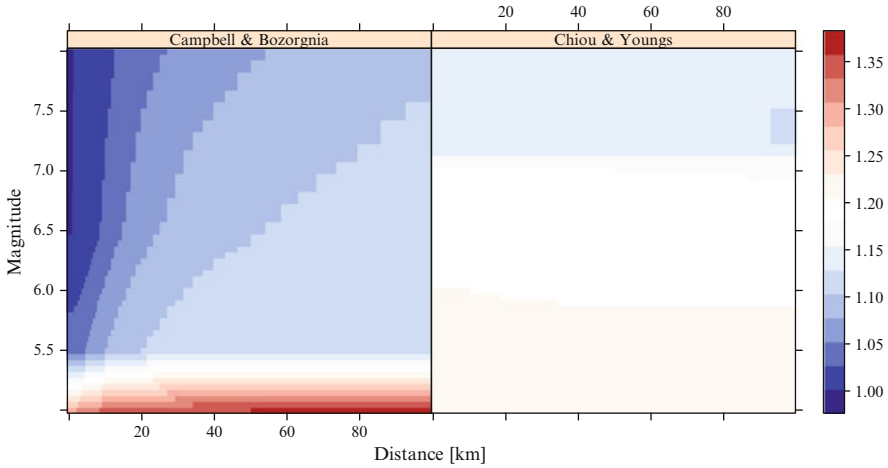
**Fig. 4.3** Example *bias factors* for the logarithmic standard deviations. The *left* panel shows ratios between the $\sigma_{\ln Sa}$ predictions of Campbell and Bozorgnia (2014, 2008) – 2014:2008, while the *right* panel shows the ratios for Chiou and Youngs (2014, 2008) – 2014:2008. The standard deviations are for a period of 0.01 s

purposes of this exercise, this departure from a more realistic representation does not influence the point that is being made.

Hazard curves for spectral acceleration at a response period of 0.01 s are computed through the use of the standard hazard integral in Eq. (4.6).

$$\lambda_{Y>y*} = \sum_{i=1} \nu_i \iint P\big[Y > y*\big|m,r\big] f_{M,R}(m,r) dm dr \qquad (4.6)$$

For this particular exercise we have just one source ($i = 1$) and will also appreciate that $\nu_i$ simply scales the hazard curve linearly and so using $\nu_1 = 1$ enables us to convert the annual rates of exceedance $\lambda_{Y>y*}$ directly into annual probabilities of exceedance.

Hazard curves computed according to this equation are shown in Fig. 4.4. The curves show that for long return periods the hazard curves predicted by both models of Campbell and Bozorgnia are very similar while at short return periods there are significant differences between the two versions of their model. From consideration of Figs. 4.2 and 4.3 we can see that the biggest differences between the two versions of the Campbell and Bozorgnia model for the scenarios of relevance to this exercise ($T = 0.01$ seconds and $V_{S,30} = 350$ m/s) are at small magnitudes between roughly $M_w 5.0$ and $M_w 5.5$ where the new model predicts significantly smaller median motions but also has a much larger standard deviation for these scenarios. As will be shown shortly, both of these effects lead to a reduction in the hazard estimates for these short return periods.

In contrast, the two versions of the Chiou and Youngs model compare favourably for the short return periods but then exhibit significant differences as
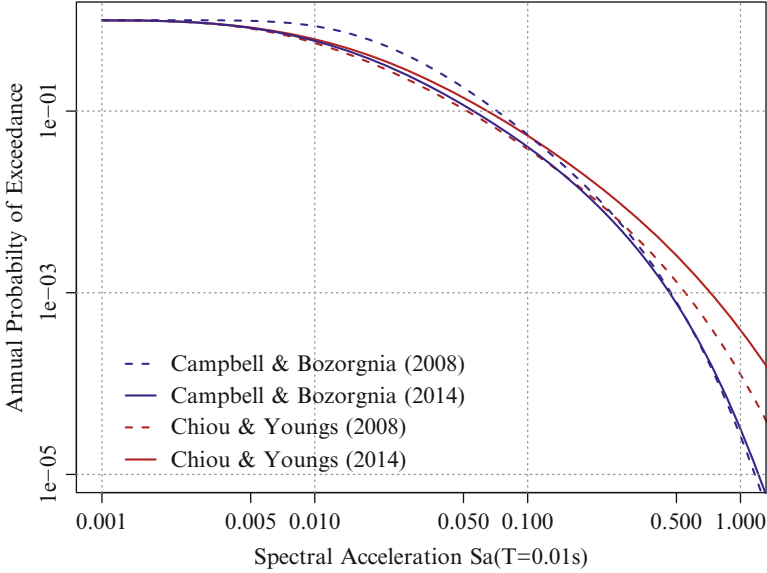
**Fig. 4.4** Hazard curves computed for the ground-motion models of Campbell and Bozorgnia (2008, 2014) and Chiou and Youngs (2008, 2014)

one moves to longer return periods. Again making use of Figs. 4.2 and 4.3 we can see that the latest version of their model provides a relatively consistent, yet mild ($\gamma_\mu \approx 1.0 - 1.1$), increase in motions over the full magnitude-distance space considered here and that we have a 15–20 % increase in the standard deviation over this full magnitude-distance space. Again, from the developments that follow, we should expect to observe the differences between the hazard curves at these longer return periods.

We have just seen how *bias factors* for the logarithmic mean $\gamma_\mu$ and logarithmic standard deviation $\gamma_\sigma$ can influence the computation of estimates of the probability of exceedance for a given scenario. The hazard integral in Eq. (4.6) is simply a weighted sum over all relevant scenarios as can be seen from the approximation (that this ceases to be an approximation in the limit as $\Delta m, \Delta r \to 0$):

$$\lambda_{Y>y^*} \approx \sum_{i=1} v_i \sum_j \sum_k P\left[Y > y^* | m_j, r_k\right] f_{M,R}\left(m_j, r_k\right) \Delta m \Delta r \qquad (4.7)$$

If we now accept that when using a ground-motion model we will only obtain an estimate of the annual rate of exceedance we can write:

$$\hat{\lambda}_{Y>y^*} \approx \sum_{i=1} v_i \sum_j \sum_k \hat{P}\left[Y > y^* | m_j, r_k\right] f_{M,R}\left(m_j, r_k\right) \Delta m \Delta r \qquad (4.8)$$

where now this expression is a function of the bias factors for both the logarithmic

motions for every scenario. One can consider the effects of systematic bias from the ground motion model expressed through factors modifying the conditional mean and standard deviation for a scenario. The biases in this case hold equally for all scenarios (although this can be relaxed). At least for the standard deviation, this assumption is not bad given the distributions shown in Fig. 4.3.

Therefore, for each considered combination of $m_j$ and $r_k$ we can define our estimate of the probability of exceeding $y*$ from Eq. (4.5). Note that the bias in the median ground motion is represented by a factor $\gamma_\mu$ multiplying the median motion $\hat{S}a = \gamma_\mu Sa$. This translates to an additive contribution to the logarithmic mean leading to $\mu_{\ln Sa} + \ln\gamma_\mu$ representing the biased median motion.

To understand how such systematic biases could influence hazard estimates we can compute the partial derivatives with respect to these bias factors, considering one source of bias at a time.

$$\frac{\partial \hat{\lambda}}{\partial \gamma_\mu} \approx \sum_{i=1} \nu_i \sum_j \sum_k \frac{\partial}{\partial \gamma_\mu}\left[1 - \Phi\left(\frac{\ln y* - \ln\gamma_\mu - \mu}{\sigma}\right)\right] f_{M,R}(m_j, r_k)\Delta m \Delta r \quad (4.9)$$

and

$$\frac{\partial \hat{\lambda}}{\partial \gamma_\sigma} \approx \sum_{i=1} \nu_i \sum_j \sum_k \frac{\partial}{\partial \gamma_\sigma}\left[1 - \Phi\left(\frac{\ln y* - \mu}{\gamma_\sigma \sigma}\right)\right] f_{M,R}(m_j, r_k)\Delta m \Delta r \quad (4.10)$$

which can be shown to be equivalent to:

$$\frac{\partial \hat{\lambda}}{\partial \gamma_\mu} = \sum_{i=1} \nu_i \iint \frac{1}{\gamma_\mu \sigma \sqrt{2\pi}} \exp\left[-\frac{(\ln\gamma_\mu + \mu - \ln y*)^2}{2\sigma^2}\right] f_{M,R}(m, r)dm dr \quad (4.11)$$

and

$$\frac{\partial \hat{\lambda}}{\partial \gamma_\sigma} = \sum_{i=1} \nu_i \iint \frac{\ln y* - \mu}{\gamma_\sigma^2 \sigma \sqrt{2\pi}} \exp\left[-\frac{(\mu - \ln y*)^2}{2\gamma_\sigma^2 \sigma^2}\right] f_{M,R}(m, r)dm dr \quad (4.12)$$

When these expressions are evaluated for the hypothetical scenario that we have considered we obtain partial derivatives as shown in Fig. 4.5. The curves in this figure show that the sensitivity of the hazard curve to changes in the mean predictions for the scenarios is most significant when there is relatively weak influence from the standard deviation. That is, when the hazard curve is dominated by contributions with epsilon values near zero then biases in the mean predictions matter most strongly.

The scaling of the partial derivatives with respect to the bias in the standard deviation is more interesting, and reflects the schematic result previously shown in Fig. 4.1. We see that we have positive gradients for the larger spectral accelerations
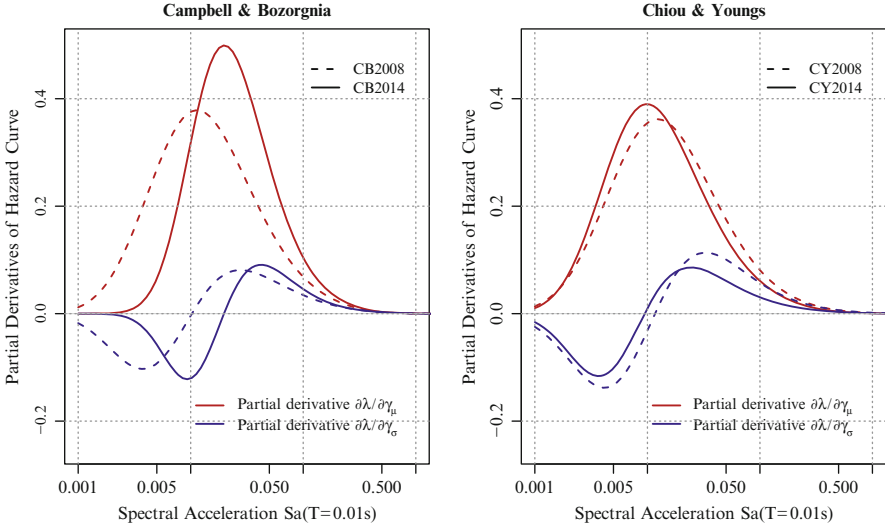
**Fig. 4.5** Partial derivatives of the hazard curves with respect to the bias factors $\gamma_\mu$ and $\gamma_\sigma$

while we have negative gradients for weak motions. These ranges effectively represent the positive and negative epsilon ranges that were shown explicitly in the previous section. However, in this case we must recognise that when considering the derivative of the hazard curve that we have many different contributions for epsilon values corresponding to a given target level of the intensity measure $y*$ and that the curves shown in Fig. 4.5 reflect a weighted average of the individual curves that have the form shown in Fig. 4.1.

The utility of the partial derivative curves shown in Fig. 4.5 is that they enable one to appreciate over which range of intensity measures (and hence return periods) changes to either the median motion or logarithmic standard deviation will have the greatest impact upon the shape of the hazard curves. Note that with respect to the typical hazard curves shown in Fig. 4.4, these derivatives should be considered as being in some sense *orthogonal* to the hazard curves. That is, they are not representing the slope of the hazard curve (which is closely related to the annual rate of occurrence of a given level of ground-motion), but rather saying that for any given level of motion, how sensitive is the annual rate of exceedance to a change in the logarithmic mean and standard deviation. It is clear from Fig. 4.4 that a change in the standard deviation itself has a strong impact upon the actual nature of the hazard curve at long return periods, whereas the sensitivity indicated in Fig. 4.5 is low for the corresponding large motions. However, it should be born in mind that these partial derivatives are $\partial\hat{\lambda}/\partial\gamma_i$ rather than, say, $\partial\ln\hat{\lambda}/\partial\gamma_i$ and that the apparently low sensitivity implied by Fig. 4.6 should be viewed in terms of the fact that small changes $\Delta\hat{\lambda}$ are actually very significant when the value of $\hat{\lambda}$ itself is very small over this range.
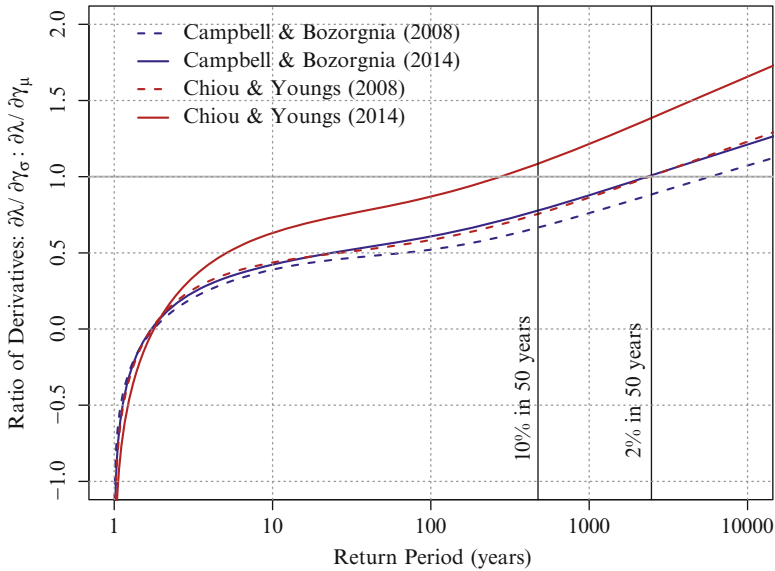
**Fig. 4.6** Ratios of the partial derivatives with respect to the logarithmic standard deviation and mean. *Vertical lines* are shown to indicate the commonly encountered 475 and 2,475 year return periods

Another way of making use of these partial derivatives is to compare the relative sensitivity of the hazard curve to changes in the logarithmic mean and standard deviation. This relative sensitivity can be computed by taking the ratio of the partial derivatives with respect to both the standard deviation and the mean and then seeing the range of return periods (or target levels of the intensity measure) for which one or the other partial derivative dominates. Ratios of this type are computed for this hypothetical scenario and are shown in Fig. 4.6. When ratios greater than one are encountered the implication is that the hazard curves are more sensitive to changes in the standard deviation than they are to changes in the mean. As can be seen from Fig. 4.6, this situation arises as the return period increases. However, for the example shown here (which is fairly typical of active crustal regions in terms of the magnitude-frequency distribution assumed) the influence of the standard deviation tends to be at least as important as the median, if not dominant, at return periods of typical engineering interest (on the order of 475 years or longer).

The example just presented has highlighted that ground-motion models must provide estimates of both the logarithmic mean and standard deviation for any given scenario, and that in many cases the ability to estimate the standard deviation is at least as important as the estimate of the mean. Historically, however, the development of ground-motion models has focussed overwhelmingly upon the scaling of median predictions, with many people (including some ground-motion model developers) still viewing the standard deviation as being some form of error in the prediction of the median rather than being an important parameter of the

ground-motion distribution that is being predicted. The results presented for this example here show that ground-motion model developers should shift the balance of attention more towards the estimation of the standard deviation than what has historically occurred.

### 4.3.2 Probabilistic Seismic Risk Analysis

When one moves to seismic risk analyses the treatment of the aleatory variability can differ significantly. In the case that a risk analysis is performed for a single structure the considerations of the previous section remain valid. However, for portfolio risk assessment it becomes important to account for the various correlations that exist with ground-motion fields for a given earthquake scenario. These correlations are required for developing the conditional ground-motion fields that correspond to a multivariate normal distribution.

The multivariate normal distribution represents the conditional random field of relative ground-motion levels (quantified through normalised intra-event residuals) conditioned upon the occurrence of an earthquake and the fact that this event will generate seismic waves with a source strength that may vary from the expected strength. The result of this source deviation is that all locations that register this ground-motion will have originally had this particular level of source strength. This event-to-event variation that systematically influences all sites is represented in ground-motion models by the inter-event variability, while the conditional variation of motions at a given site is given by the intra-event variability.

For portfolio risk analysis it is therefore important to decompose the total aleatory variability in ground-motions into a component that reflects the source strength (the inter-event variability) and a component that reflects the site-specific aleatory variability (the intra-event variability). It should also be noted in passing that this is not strictly equivalent to the variance decomposition that is performed using mixed effects models in regression analysis.

When one considers ground-motion models that have been developed over recent years it is possible to appreciate that some significant changes have occurred to the value of the total aleatory variability that is used in hazard analysis, but also to the decomposition of this total into the inter-event and intra-event components. For portfolio risk analysis, this decomposition matters. To demonstrate why this is the case, Fig. 4.7 compares conditional ground-motion fields that have been simulated for the 2011 Christchurch Earthquake in New Zealand. In each case shown, the inter-event variability is assumed to be a particular fraction of the total variability and this fraction is allowed to range from 0 to 100 %. As one moves from a low to a high fraction it is clear that the within event spatial variation of the ground-motions reduces.

For portfolio risk assessment, these differences in the spatial variation are important as the extreme levels of loss correspond to cases in which spatial regions of high-intensity ground-motion couple with regions of high vulnerability and
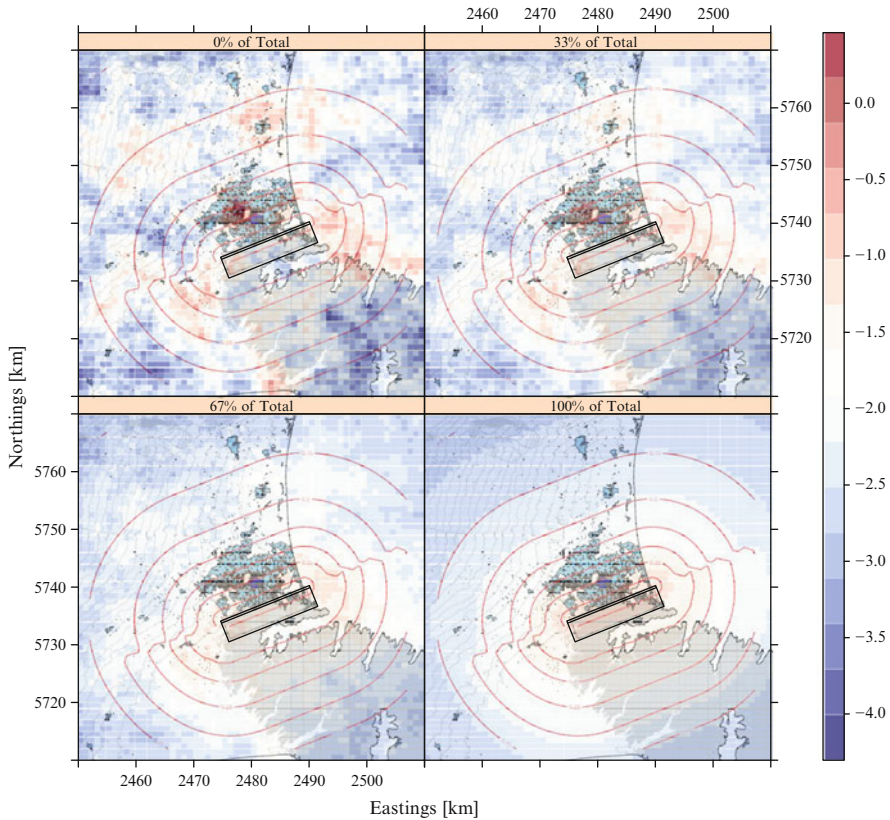
**Fig. 4.7** Impact upon the nature of ground-motion fields generated assuming that the inter-event variability is a given fraction of the total aleatory variability. The ground-motion fields shown are possible fields consistent with a repeat of the Christchurch earthquake

exposure. The upper left panel of Fig. 4.7 shows a clear example of this where a patch of high intensity is located in a region of high exposure.

In addition to ensuring that the total aleatory variability is well-estimated, it is therefore also very important (for portfolio risk analysis) to ensure that the partitioning of the total variability between inter- and intra-event components is done correctly.

## 4.4 Components of Uncertainty

The overall uncertainty in ground-motion prediction is often decomposed into components of Aleatory Variability and Epistemic Uncertainty. In the vast majority of applications *only* these two components are considered and they are defined in

such as way that the aleatory variability is supposed to represent inherent randomness in nature while epistemic uncertainties represent contributions resulting from our lack of knowledge. The distinction is made for more than semantic reasons and the way that each of these components is treated within hazard and risk analysis differ. Using probabilistic seismic hazard analysis as an example, the aleatory variability is directly accounted for within the *hazard integral* while epistemic uncertainty is *accounted for* or *captured* through the use of logic trees.

However, when one constructs a logic tree the approach is to consider alternative hypotheses regarding a particular effect, or component, within the analysis. Each alternative is then assigned a weight that has been interpreted differently by various researchers and practitioners, but is ultimately treated as a probability. No alternative hypotheses are considered for effects that we do not know to be relevant. That is, the representation of epistemic uncertainty in a logic tree only reflects our uncertainty regarding the components of the model that we think are relevant. If we happen to be missing an important physical effect then we will never think to include it within our tree and this degree of ignorance is never reflected in our estimate of epistemic uncertainty.

It is therefore clear that there is a component of the overall uncertainty in our analyses that is not currently accounted for. This component is referred to as *Ontological Uncertainty* (Elms 2004) and represents the unknown unknowns from the famous quote of Donald Rumsfeld.

These generic components of uncertainty are shown schematically in Fig. 4.8. The actual numbers that are shown in this figure are entirely fictitious and the objective is not to define this partitioning. Rather, the purpose of this figure is to illustrate the following:

- What we currently refer to as being aleatory variability is not all aleatory variability and instead contains a significant component of epistemic uncertainty (which is why it reduces from the present to the near future)
- The fact that ontological uncertainty exists means that we cannot assign a numerical value to epistemic uncertainty
- The passage of time allows certain components to be reduced

In the fields of seismic hazard and risk it is common for criticism to be made of projects due to the improper handling of aleatory variability and epistemic uncertainty by the analysts. However, the distinction between these components is not always clear and this is at least in part a result of loose definitions of the terms as well as a lack of understanding about the underlying motivation for the decomposition.

As discussed at length by Der Kiureghian and Ditlevsen (2009), what is aleatory or epistemic can depend upon the type of analysis that is being conducted. The important point that Der Kiureghian and Ditlevsen (2009) stress is that the categorisation of an uncertainty as either aleatory of epistemic is largely at the discretion of the analyst and depends upon what is being modelled. The uncertainties themselves are generally not properties of the parameter in question.
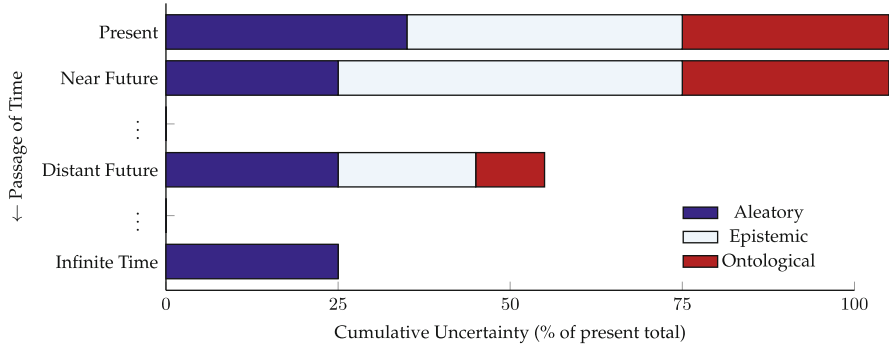
**Fig. 4.8** Components of the total uncertainty in ground motion prediction, and their evolution in time. The percentage values shown are entirely fictitious

### 4.4.1 Nature of Uncertainty

Following the more complete discussion provided by Der Kiureghian and Ditlevsen (2009), consider the physical process that results in the generation of a ground motion $y$ for a particular scenario. The underlying basic variables that parameterise this physical process can be written as $\mathbf{X}$.

Now consider a perfect deterministic ground-motion model (*i.e.*, one that makes predictions with no error) that provides a mathematical description of the physical link between these basic variables and the observed motion. In the case that we knew the exact values of all basic variables for a given scenario we would write such a model as:

$$y = g\left(\mathbf{x}, \theta_g\right) \qquad (4.13)$$

where, here $\theta_g$ are the parameters or coefficients of the model. Note that the above model must account for all relevant physical effects related to the generation of $y$. In practice, we cannot come close to accounting for all relevant effects and so rather than working with the full set $\mathbf{X}$, we instead work with a reduced set $\mathbf{X}_k$ (representing the known random variables) and accept that the effect of the unknown basic variables $\mathbf{X}_u$ will manifest as differences between our now approximate model $\hat{g}$ and the observations. Furthermore, as we are working with an observed value of $y$ (which we assume to be known without error) we also need to recognise that we will have an associated observed instance of $\mathbf{X}_k$ that is not perfectly known $\mathbf{x}_k$. Our formulation is then written as:

$$y = \hat{g}\left(\hat{\mathbf{x}}_k, \hat{\theta}_g\right) + \varepsilon \qquad (4.14)$$

What is important to note here is that the residual error $\varepsilon$ is the result of three distinct components:

- The effect of unobserved, or not considered, variables $\mathbf{X}_u$
- The imperfection of our mathematical model, both in terms of its functional form and the estimation of its parameters $\hat{\theta}_g$
- The uncertainties associated with estimated known variables $\hat{\mathbf{x}}_k$

The imperfection referred to in the second point above means that the residual error $\varepsilon$ does not necessarily have a zero mean (as is the case for regression analysis). The reason being that the application of imperfect physics does not mean that our simplified model will be unbiased – both when applied to an entire ground-motion database, but especially when applied to a particular scenario. Therefore, it could be possible to break down the errors in prediction into components representing bias, $\Delta(,\hat{\mathbf{x}},\hat{\theta}_g)$, and variability, $\varepsilon'$:

$$\varepsilon \rightarrow \Delta(\hat{\mathbf{x}},\hat{\theta}_g) + \varepsilon' \tag{4.15}$$

In the context seismic hazard and risk analysis, one would ordinarily regard the variability represented by $\varepsilon$ as being aleatory variability and interpret this as being inherent randomness in ground motions arising from the physical process of ground-motion generation. However, based upon the formulation just presented one must ask whether any actual inherent randomness exists, or whether we are just seeing the influence of the unexplained parameters $\mathbf{x}_u$. That is, should our starting point have been:

$$y = g(\mathbf{x},\theta_g) + \varepsilon_A \tag{4.16}$$

where here the $\varepsilon_A$ represents intrinsic randomness associated with ground motions.

When one considers this problem one must first think about what type of randomness we are dealing with. Usually when people define aleatory variability they make an analogy with the rolling of a die, but often they are unwittingly referring to one particular type of randomness. There are broadly three classes of randomness:

- **Apparent Randomness**: This is the result of viewing a complex deterministic process from a simplified viewpoint.
- **Chaotic Randomness**: This randomness arises from nonlinear systems that evolve from a particular state in a manner that depends very strongly upon that state. Responses obtained from very slightly different starting conditions can be markedly different from each other, and our inability to perfectly characterise a particular state means that the system response is unpredictable.
- **Inherent Randomness**: This randomness is an intrinsic part of reality. Quantum mechanics arguably provides the most pertinent example of inherent randomness.

Note that there is also a subtle distinction that can be made between systems that are deterministic, yet unpredictable, and systems that possess genuine randomness.

In addition, some (including historically Einstein) argue that systems that possess 'genuine randomness' are actually driven by deterministic processes and variables that we simply are not aware of. In this case, these systems would be subsumed within the one or more of the other categories of apparent or chaotic randomness. However, at least within the context of quantum mechanics, Bell's theorem demonstrates that the randomness that is observed at such scales is in fact inherent randomness and not the result of apparent randomness.

For ground-motion modelling, what is generally referred to as aleatory variability is at least a combination of both apparent randomness and chaotic randomness and could possibly also include an element of inherent randomness – but there is no hard evidence for this at this point. The important implication of this point is that the component associated with apparent randomness is actually an epistemic uncertainty that can be reduced through the use of more sophisticated models. The following two sections provide examples of apparent and chaotic randomness.

### 4.4.2   Apparent Randomness – Simplified Models

Imagine momentarily that it is reasonable to assume that ground-motions arise from deterministic processes but that we are unable to model all of these processes. We are therefore required to work with simplified models when making predictions. To demonstrate how this results in apparent variability consider a series of simplified models for the prediction of peak ground acceleration (here denoted by $y$) as a function of moment magnitude $\mathbf{M}$ and rupture distance $R$:

Model 0

$$\ln y = \beta_0 + \beta_1 \mathbf{M} \tag{4.17}$$

Model 1

$$\ln y = \beta_0 + \beta_1 \mathbf{M} + \beta_2 \ln\sqrt{R^2 + \beta_3^2} \tag{4.18}$$

Model 2

$$\ln y = \beta_0 + \beta_1 \mathbf{M} + \beta_2 \ln\sqrt{R^2 + \beta_3^2} + \beta_4 \ln V_{S,30} \tag{4.19}$$

Model 3

$$\ln y = \beta_0 + \beta_1 \mathbf{M} + \beta_{1a}(\mathbf{M} - 6.5)^2 + [\beta_2 + \beta_{2a}(\mathbf{M} - 6.5)]\ln\sqrt{R^2 + \beta_3^2}$$
$$+ \beta_4 \ln V_{S,30} \tag{4.20}$$

Model 4

$$\ln y = \beta_0 + \beta_1 \mathbf{M} + \beta_{1a}(\mathbf{M} - 6.5)^2 + [\beta_2 + \beta_{2a}(\mathbf{M} - 6.5)]\ln\sqrt{R^2 + \beta_3^2} \\ + \beta_4 \ln V_{S,30} + \beta_5 F_{nm} + \beta_6 F_{rv}$$ (4.21)

Models 5 and 6

$$\ln y = \beta_0 + \beta_1 \mathbf{M} + \beta_{1a}(\mathbf{M} - 6.5)^2 + [\beta_2 + \beta_{2a}(\mathbf{M} - 6.5)]\ln\sqrt{R^2 + \beta_3^2} \\ + \beta_4 \ln V_{S,30} + \beta_5 F_{nm} + \beta_6 F_{rv} + \beta_7 F_{as}$$ (4.22)

where we see that the first of these models is overly simplified, but that by the time we reach Models 5 and 6, we are accounting for the main features of modern models. The difference between Models 5 and 6 is not in the functional form, but in how the coefficients are estimated. Models 1–5 use standard mixed effects regression with one random effect for event effects. However, Model 6 includes this random effect, but also distinguishes between these random effects depending upon whether we have mainshocks or aftershocks and also partitions the intra-event variance into components for mainshocks and aftershocks. The dataset consists of 2,406 records from the NGA database.

Figure 4.9 shows estimates of apparent randomness for each of these models, assuming that Model 6 is 'correct'. That is, the figure shows the difference between the total standard deviation of Model $i$ and Model 6 and because we assume the latter model is correct, this difference in variance can be attributed to apparent randomness. The figure shows that the inclusion of distance scaling and distinguishing between mainshocks and aftershocks has a very large impact, but that other additions in complexity provide a limited reduction in apparent randomness. The important point here is that this apparent randomness is actually epistemic uncertainty – not aleatory as is commonly assumed.

### 4.4.3 Chaotic Randomness – Bouc-Wen Example

Chaotic randomness is likely to be a less-familiar concept than apparent randomness given that the latter is far more aligned with our normal definition of epistemic uncertainty. To explain chaotic randomness in the limited space available here is a genuine challenge, but I will attempt this through the use of an example based heavily upon the work of Li and Meng (2007). The example concerns the response of a nonlinear oscillator and is not specifically a ground-motion example. However, this type of model has been used previously for characterising the effects of nonlinear site response. I consider the nonlinear Bouc-Wen single-degree-of-freedom system characterised by the following equation:
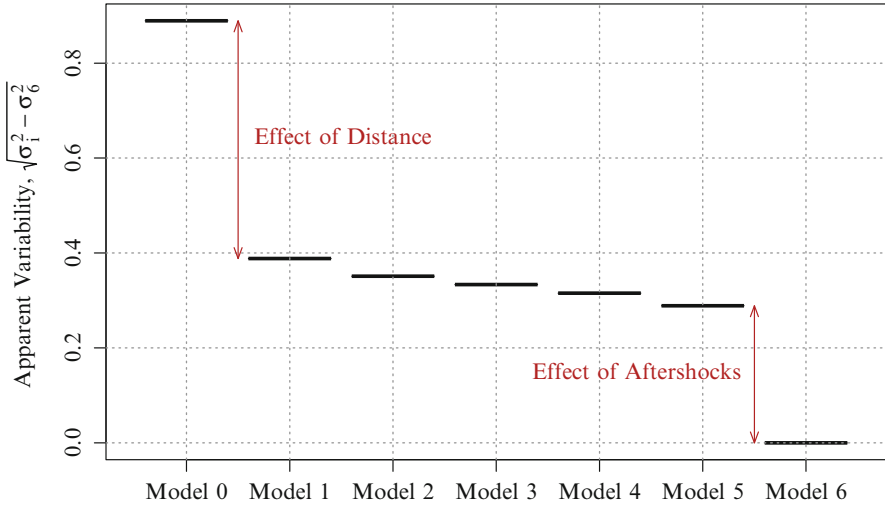
**Fig. 4.9** Variation of apparent randomness associated with models of increasing complexity

$$\ddot{u} + 2\zeta\omega_0\dot{u} + \alpha\omega_0^2 u + (1 - \alpha)\omega_0^2 z = B \sin{(\Omega t)} \qquad (4.23)$$

where the nonlinear hysteretic response is defined by:

$$\dot{z} = A\dot{u} - \gamma|\dot{u}||z||z|^{n-1} - \beta\dot{u}|z|^n \qquad (4.24)$$

This model is extremely flexible and can be parameterised so that it can be applied in many cases of practical interest, but in the examples that follow we will assume that we have a system that exhibits hardening when responding in a nonlinear manner (see Fig. 4.10).

Now, if we subject this system to a harmonic excitation we can observe a response at relatively low amplitudes that resembles that in Fig. 4.11. Here we show the displacement response, the velocity response, the trajectory of the response in the phase space ($u - \dot{u}$ space) and the nonlinear restoring force. In all cases the line colour shifts from light blue, through light grey and towards a dark red as time passes. In all panels we can see the influence of the initial transient response before the system settles down to a steady-state. In particular, we can see that we reach a limit-cycle in the phase space in the lower left panel.

For Fig. 4.11 the harmonic amplitude is $B = 5$ and we would find that if we were to repeat the analysis for a loading with an amplitude slightly different to this value that our response characteristics would also only be slightly different. For systems in this low excitation regime we have predictable behaviour in that the effect of small changes to the amplitude can be anticipated.

However, consider now a plot of the maximum absolute displacement and maximum absolute velocity against the harmonic amplitude shown in Fig. 4.12. Note that the response values shown in this figure correspond to what are essentially
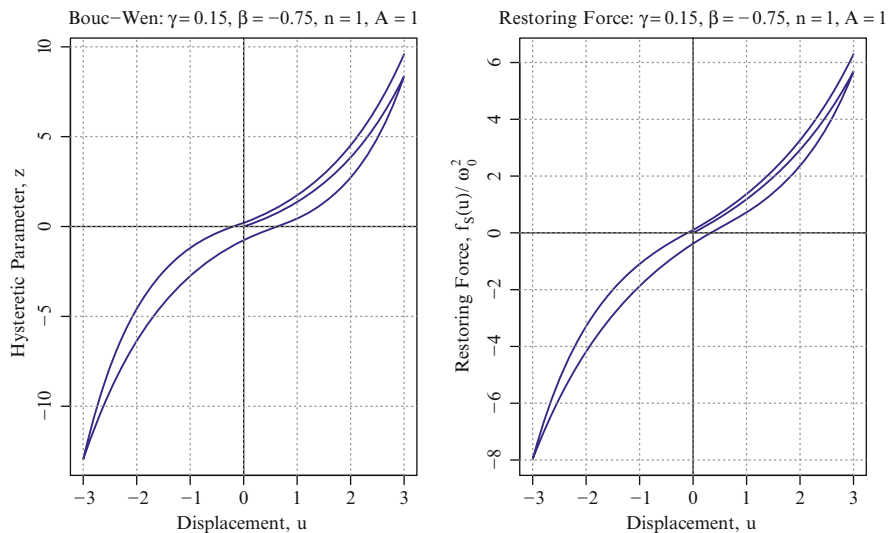
**Fig. 4.10** Dependence of the hysteretic parameter $z$ (*left*), and the normalised restoring force $f_S$ $(u, \dot{u}, z)$ (*right*) on the displacement for the example system considered

*steady-state* conditions. For this sort of system we expect that the transient terms will decay according to $\exp(-\zeta \omega_0 t)$ and for these examples we have set $\zeta = 0.05$ and $\omega_0 = 1.0$ and we only look at the system response after 200 s have passed in order to compute the maximum displacement and velocity shown in Fig. 4.12. We would expect that the transient terms would have decayed to less than $0.5 \times 10^{-5}$ of their initial amplitudes at the times of interest.

Figure 4.12 shows some potentially surprising behaviour for those not familiar with nonlinear dynamics and chaos. We can see that for low harmonic amplitudes we have a relatively smoothly varying maximum response and that system response is essentially predictable here. However, this is not to say that the response does not become more complex. For example, consider the upper row of Fig. 4.13 that shows the response for $B = 15$. Here we can see that the system tends towards some stable state and that we have a stable limit-cycle in the phase space. However, it has a degree of periodicity that corresponds to a loading/unloading phase for negative restoring forces.

This complexity continues to increase as the harmonic amplitude increases as can be seen in the middle row of Fig. 4.13 where we again have stable steady-state response, but also have another periodic component of unloading/reloading for both positive and negative restoring forces. While these figures show increased complexity as we move along the harmonic amplitude axis of Fig. 4.12, the system response remains stable and predictable in that we know that small changes in the value of $B$ continues to map into small qualitative and quantitative changes to the response. However, Fig. 4.12 shows that once the harmonic amplitude reaches values of roughly $B = 53$ we suddenly have a qualitatively different behaviour. The
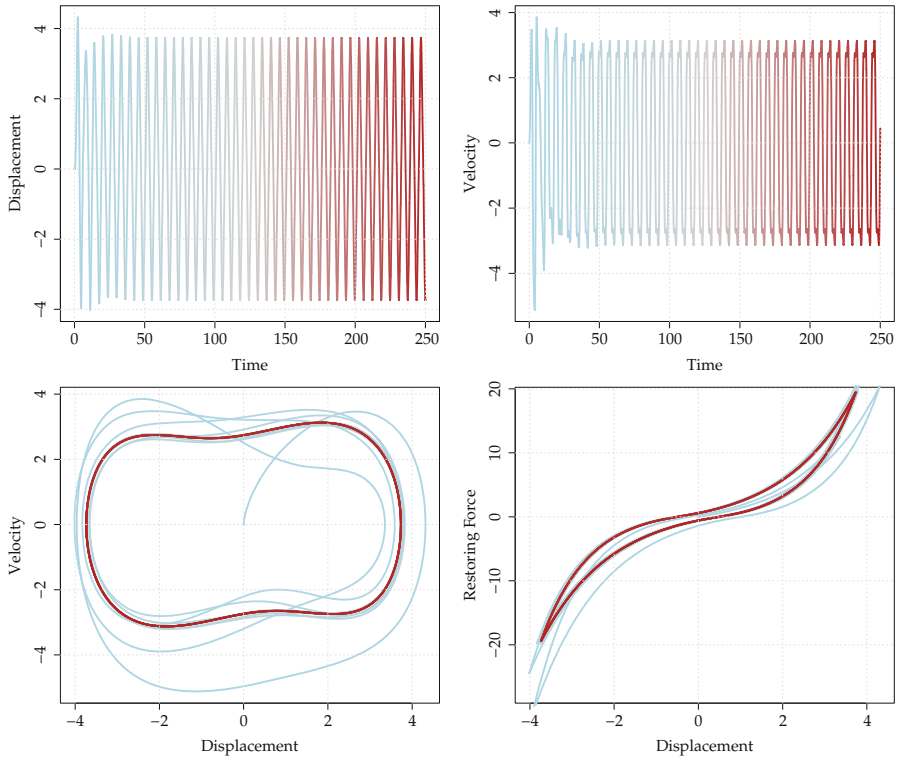
**Fig. 4.11** Response of the nonlinear system for a harmonic amplitude of $B = 5$. *Upper left* panel shows the displacement time-history; *upper right* panels shows the velocity time history; *lower right* panel shows the response trajectory in phase space; and *lower right* panel shows the hysteretic response

system response now becomes extremely sensitive to the particular value of the amplitude that we consider. The reason for this can be seen in the bottom row of Fig. 4.13 in which it is clear that we never reach a stable steady state. What is remarkable in this regime is that we can observe drastically different responses for very small changes in amplitude of the forcing function. For example, when we move from $B = 65.0$ to $B = 65.1$ we have transition back into a situation in which we have a stable limit cycle (even if it is a complex cycle).

This lesson here is that for highly nonlinear processes there exist response regimes where the particular response trajectory and system state depends very strongly upon a prior state of the system. There are almost certainly aspects of the ground-motion generation process that can be described in this manner. Although these can be deterministic processes, as it is impossible to accurately define the state of the system the best we can do is to characterise the observed chaotic randomness. Note that although this is technically epistemic uncertainty, we have no choice but to treat this as aleatory variability as it is genuinely irreducible.
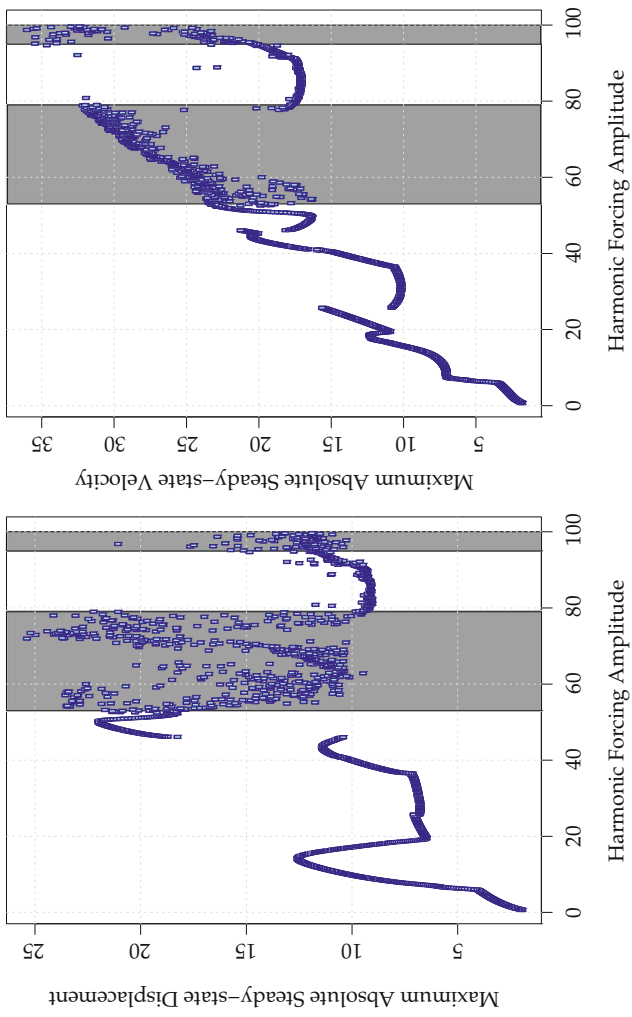
**Fig. 4.12** Maximum absolute steady-state displacement (*left*) and velocity (*right*) response against the harmonic forcing amplitude *B*
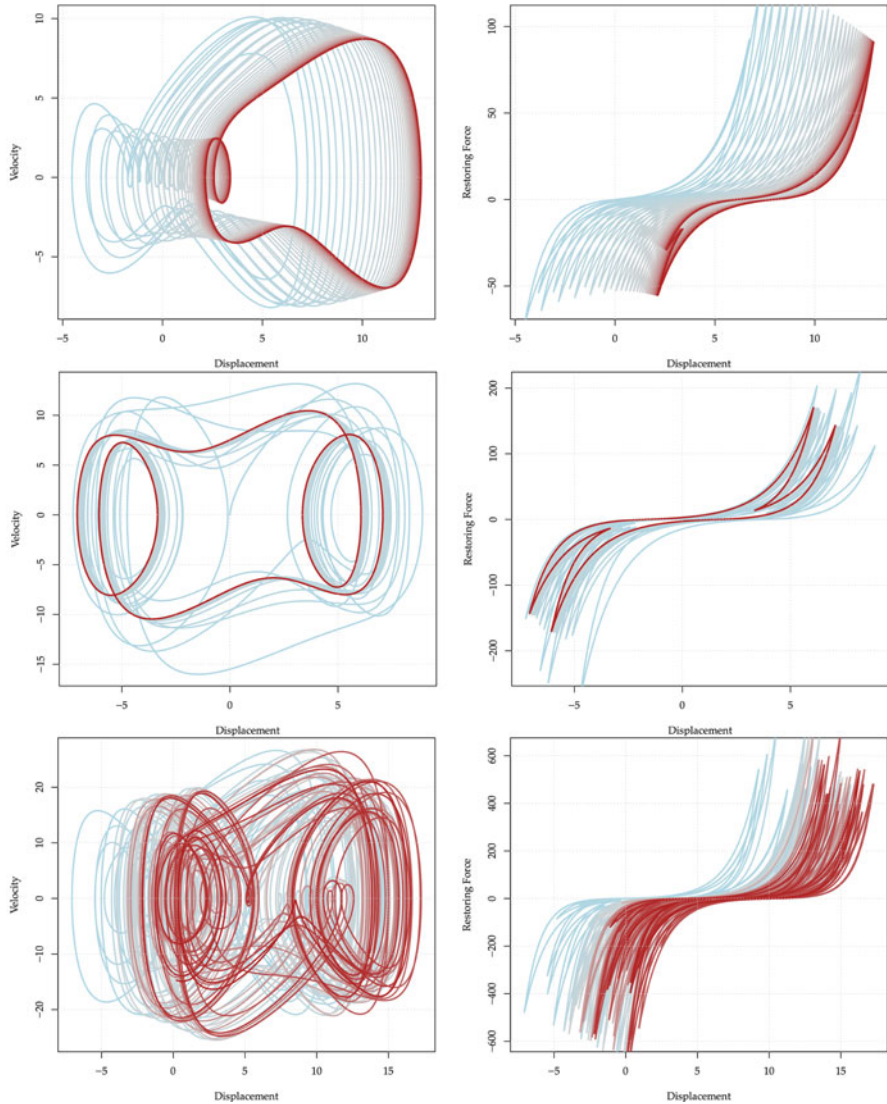
**Fig. 4.13** Response of the nonlinear system for a harmonic amplitude of $B = 15$ (*top*), $B = 35$ (*middle*), and $B = 65$ (*bottom*). Panels on the *left* show the response trajectory in phase space; and panels on the *right* show the hysteretic response

## 4.4.4   Randomness Represented by Ground-Motion Models

The standard deviation that is obtained during the development of a ground-motion model definitely contains elements of epistemic uncertainty that can be regarded as apparent randomness, epistemic uncertainty that is the result of imperfect metadata,

and variability that arises from the ergodic assumption. It is also almost certain that the standard deviation reflects a degree of chaotic randomness and possibly also includes some genuine randomness and it is only these components that are actually, or practically, irreducible. Therefore, it is clear that the standard deviation of a ground-motion model does not reflect aleatory variability as it is commonly defined – as being 'inherent variability'.

If the practical implications of making the distinction between aleatory and epistemic are to dictate what goes into the hazard integral and what goes into the logic tree then one might take the stance that of these contributors to the standard deviation just listed we should look to remove the effects of the ergodic assumption (which is attempted in practice), we should minimise the effects of metadata uncertainty (which is not done in practice), and we should increase the sophistication of our models so that the apparent randomness is reduced (which some would argue has been happening in recent years, *vis-à-vis* the NGA projects).

An example of the influence of metadata uncertainty can be seen in the upper left panel of Fig. 4.14 in which the variation in model predictions is shown when uncertainties in magnitude and shear-wave velocity are considered in the regression analysis. The boxplots in this figure show the standard deviations of the predictions for each record in the NGA dataset when used in a regression analysis with Models 1–6 that were previously presented. The uncertainty that is shown here should be regarded as a lower bound to the actual uncertainty associated with meta-data for real ground-motion models. The estimates of this variable uncertainty are obtained by sampling values of magnitude and average shear-wave velocity for each event and site assuming a (truncated) normal and lognormal distribution respectively. This simulation process enables a hypothetical dataset to be constructed upon which a regression analysis is performed. The points shown in the figure then represent the standard deviation of median predictions from each developed regression model.

Figure 4.14 also shows how an increase in model complexity is accompanied by an increase in parametric uncertainty for the models presented previously. It should be noted that these estimates of parametric uncertainty are also likely to be near lower bounds given that the functional forms used for this exercise are relatively simple and that the dataset is relatively large (consisting of 2,406 records from the NGA database). The upper right panel of Fig. 4.14 shows this increasing parametric uncertainty for the dataset used to develop the models, but the lower panel shows the magnitude dependence of this parametric uncertainty when predictions are made for earthquake scenarios that are not necessarily covered by the empirical data. In this particular case, the magnitude dependence is shown when motions are computed for a distance of just 1 km and a shear-wave velocity of 316 m/s is used. It can be appreciated from this lower panel that the parametric uncertainty is a function of both the model complexity but also of the particular functional form adopted. The parametric uncertainty here is estimated by computing the covariance matrix of the regression coefficients and then sampling from the multivariate normal distribution implied by this covariance matrix. The simulated coefficients
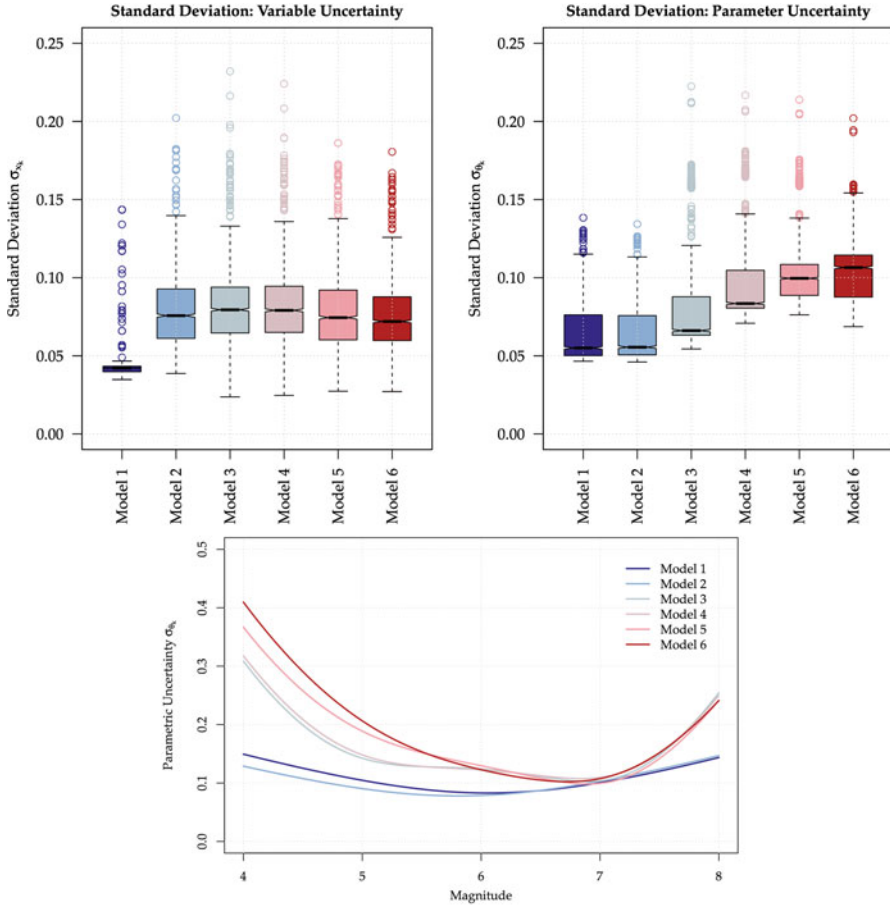
**Fig. 4.14** Influence of meta-data uncertainty (*upper left*), increase in parametric uncertainty with increasing complexity of models (*upper right*), and the dependence of parametric uncertainty upon magnitude (*bottom*)

are then used to generate predictions for each recording and the points shown in this panel represent the standard deviation of these predictions for every record.

Rather than finally looking to increase the complexity of the functional forms that are used for ground-motion predictions, herein I propose that we look at this problem in a different light and refer back to Eq. (4.2) in which we say explicitly that what matters for hazard and risk is the overall estimate of ground-motion exceedance and that this is the result of two components (not just the ground-motion model). We should forget about trying to push the concept that only aleatory variability should go into the hazard integral and rather take the viewpoint that our *optimal model* (which is a model of the ground motion *distribution* – not median predictions) should go into the hazard integral and that our uncertainties should then be reflected in the logic tree. The reason why we should forget about only pushing

aleatory variability into the hazard integral is that from a quantitative ground-motion perspective we are still not close to understanding what is actually aleatory and irreducible.

The proposed alternative of defining an *optimal model* is stated in the light of minimising the uncertainty in the estimate of the probability of exceedance of ground-motions. This uncertainty comes from two components: (1) our ability to accurately define the probability of occurrence of earthquake scenarios; and (2) our ability to make robust predictions of the conditional ground-motion distribution. Therefore, while a more complex model will act to reduce the apparent variability, if this same model requires the specification of a number of independent variables that are poorly constrained in practice then the overall uncertainty will be large. In such cases, one can obtain a lower level of overall uncertainty in the prediction of ground-motion exceedance by using a less complex ground-motion model. A practical example of this trade-off is related to the requirement to define the depth distribution of earthquake events. For most hazard analyses this depth distribution is poorly constrained and the inclusion of depth-dependent terms in ground-motion models only provides a very small decrease in the apparent variability.

Figure 4.15 presents a schematic illustration of the trade-offs between apparent randomness (the epistemic uncertainty that is often regarded as aleatory variability) and parametric uncertainty (the epistemic uncertainty that is usually ignored) that exist just on the ground-motion modelling side. The upper left panel of this figure shows, as we have seen previously, that the apparent randomness decreases as we increase the complexity of our model. However, the panel also shows that this reduction saturates once we reach the point where we have chaotic randomness, inherent randomness, or a combination of these irreducible components. The upper right panel, on the other hand, shows that as this model complexity increases we also observe an increase in parametric uncertainty. The optimal model must balance these two contributors to the overall uncertainty as shown in the lower left panel. On this basis, one can identify an optimal model when only ground-motion modelling is considered. When hazard or risk is considered then the parametric uncertainty shown here should reflect both the uncertainty in the model parameters (governed by functional form complexity, and data constraints) and the uncertainty associated with the characterisation of the scenario (*i.e.*, the independent variables) and its likelihood.

The bottom right panel of Fig. 4.15 shows how one can justify an increased complexity in the functional form when the parametric uncertainty is reduced, as in this case the optimal complexity shifts to the right. To my knowledge, these sorts of considerations have never been explicitly made during the development of more complex ground-motion models. Although, in some ways, the quantitative inspection of residual trends and of parameter *p*-values is an indirect way of assessing if increased complexity is justified by the data.

Recent years have seen the increased use of external constraint during ground-motion model development. In particular, numerical simulations are now commonly undertaken in order to constrain nonlinear site response scaling, large
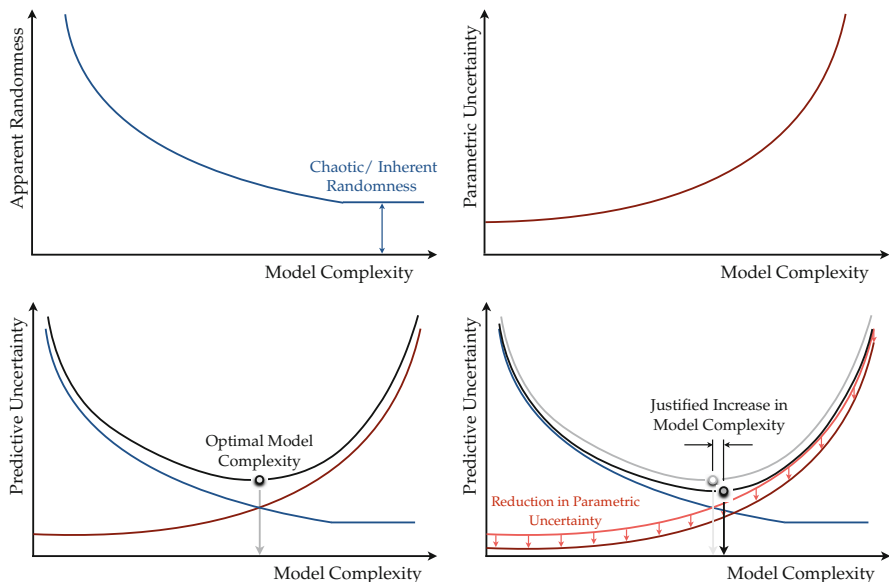
**Fig. 4.15** Schematic illustration of the trade-off that exists between the reduction in apparent randomness (*upper left*) and the increase in parametric uncertainty (*upper right*). The optimal model in this context balances the two components (*lower left*) and an increase in complexity is justified when parametric uncertainty is reduced (*lower right*)

magnitude scaling, and near field effects. Some of the most recent models that have been presented have very elaborate functional forms and the model developers have justified this additional complexity on the basis of the added functional complexity being externally constrained. In the context of Fig. 4.15, the implication is that the model developers are suggesting that the red curves do not behave in this manner, but rather that they saturate at some point as all of the increasing complexity does not contribute to parametric uncertainty. On one hand, the model developers are correct in that the application of external constraints does not increase the estimate of the parametric uncertainty *from the regression analysis on the free parameters*. However, on the other hand, in order to properly characterise the parametric uncertainty the uncertainty associated with the models used to provide the external constraint must also be accounted for. In reality this additional parametric uncertainty is actually larger than what would be obtained from a regression analysis because the numerical models used for these constraints are normally very complex and involve a large number of poorly constrained parameters. Therefore, it is not clear that the added complexity provided through the use of external constraints is actually justified.

## 4.5 Discrete Random Fields for Spatial Risk Analysis

The coverage thus far has been primarily focussed upon issues that arise most commonly within hazard analysis, but that are also relevant to risk analysis. However, in this final section the attention is turned squarely to a particular issue associated with the generation of ground-motion fields for use in earthquake loss estimation for spatially-distributed portfolios. This presentation is based upon the work of Vanmarcke (1983) and has only previously been employed by Stafford (2012).

The normal approach that is taken when performing risk analyses over large spatial regions is to subdivide the region of interest into geographic cells (often based upon geopolitical boundaries, such as districts, or postcodes). The generation of ground-motion fields is then made by sampling from a multivariate normal distribution that reflects the joint intra-event variability of epsilon values across a finite number of sites equal to the number of geographic cells. The multivariate normal distribution for epsilon values is correctly assumed to have a zero mean vector, but the covariance matrix of the epsilon values is computed using a combination of the point-to-point distances between the centroids of the cells (weighted geographically, or by exposure) and a model for spatial correlation between two points (such as that of Jayaram and Baker 2009). The problem with this approach is that the spatial discretisation of the ground-motion field has been ignored. The correct way to deal with this problem is to discretise the random field to account for the nature of the field over each geographic cell and to define a covariance matrix for the average ground-motions over the cells. This average level of ground-motion over the cell is a far more meaningful value to pass into fragility curves than a single point estimate.

Fortunately, the approach for discretisation of a two-dimensional random field is well established (Vanmarcke 1983). The continuous field is denoted by $\ln y(\mathbf{x})$ where $y$ is the ground motion and $\mathbf{x}$ now denotes a spatial position. The logarithmic motion at a point can be represented as a linear function of the random variable $\varepsilon(\mathbf{x})$. Hence, the expected value of the ground motion field at a given point is defined by Eq. (4.25), where $\mu_{\ln y}$ is the median ground motion, and $\eta$ is an event term.

$$E[\ln y(\mathbf{x})] = \mu_{\ln y} + \eta + E[\varepsilon(\mathbf{x})] \tag{4.25}$$

Therefore, in order to analyse the random field of ground motions, attention need only be given to the random field of epsilon values. Once this field is defined it may be linearly transformed into a representation of the random field of spectral ordinates.

In order to generate ground-motion fields that account for the spatial discretisation, under the assumption of joint normality, we require three components:

- An expression for the average mean logarithmic motion over a geographic cell
- An expression for the variance of motions over a geographic cell
- An expression for the correlation of average motions from cell-to-cell

For the following demonstration, assume that the overall region for which we are conducting the risk analysis is discretised into a regular grid aligned with the N-S and E-W directions. This grid has a spacing (or dimension) in the E-W direction of $D_1$ and a spacing in the N-S direction of $D_2$. Note that while the presentation that follows concerns this regular grid, Vanmarcke (1983) shows how to extend this treatment to irregularly shaped regions (useful for regions defined by postcodes or suburbs, *etc.*).

Within each grid cell one may define the local average of the field by integrating the field and dividing by the area of the cell $(A = D_1 D_2)$.

$$\ln y_A = \frac{1}{A} \int_A \ln y(\mathbf{x}) d\mathbf{x} \tag{4.26}$$

Now, whereas the variance of the ground motions for a single point in the field, given an event term, is equal to $\sigma^2$, the variance of the local average $\ln y_A$ must be reduced as a result of the averaging. Vanmarcke (1983) shows that this reduction can be expressed as in Eq. (4.27).

$$\sigma_A^2 = \gamma(D_1, D_2)\sigma^2 \quad \rightarrow \quad \gamma(D_1, D_2) = \frac{1}{D_1 D_2}\int_{-D_2}^{D_2}\int_{-D_1}^{D_1}\left(1 - \frac{|\delta_1|}{D_1}\right)\left(1 - \frac{|\delta_2|}{D_2}\right)$$
$$\rho(\delta_1, \delta_2)\mathrm{d}\delta_1\mathrm{d}\delta_2 \tag{4.27}$$

In Eq. (4.27), the correlation between two *points* within the region is denoted by $\rho(\delta_1, \delta_2)$, in which $\delta_1$ and $\delta_2$ are orthogonal co-ordinates defining the relative positions of two points within a cell. In practice, this function is normally defined as in Eq. (4.28) in which $b$ is a function of response period.

$$\rho(\delta_1, \delta_2) = \exp\left(-\frac{3\sqrt{\delta_1^2 + \delta_2^2}}{b}\right) \tag{4.28}$$

The reduction in variance associated with the averaging of the random field is demonstrated in Fig. 4.16 in which values of $\gamma(D_1, D_2)$ are shown for varying values of the cell dimension and three different values of the range parameter $b$. For this example the cells are assumed to be square.

With the expressions for the spatial average and the reduced variance now given, the final ingredient that is required is the expression for the correlation between the average motions over two cells (rather than between two points). This is provided in Eq. (4.29), with the meaning of the distances $D_{1k}$ and $D_{2l}$ shown in Fig. 4.17.
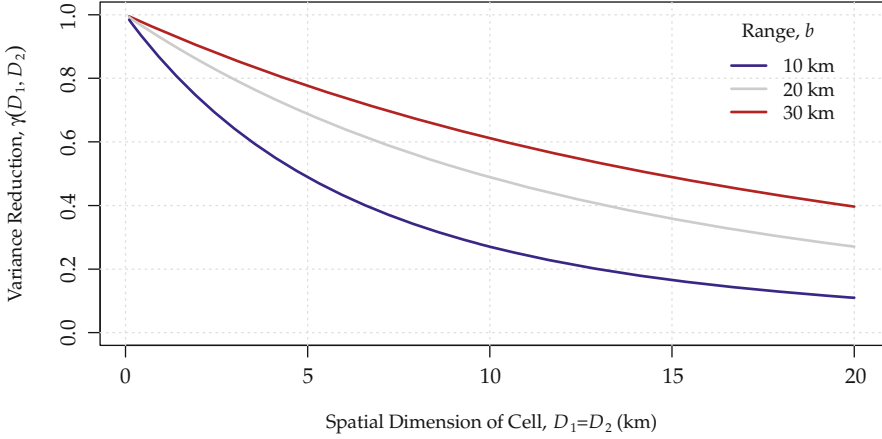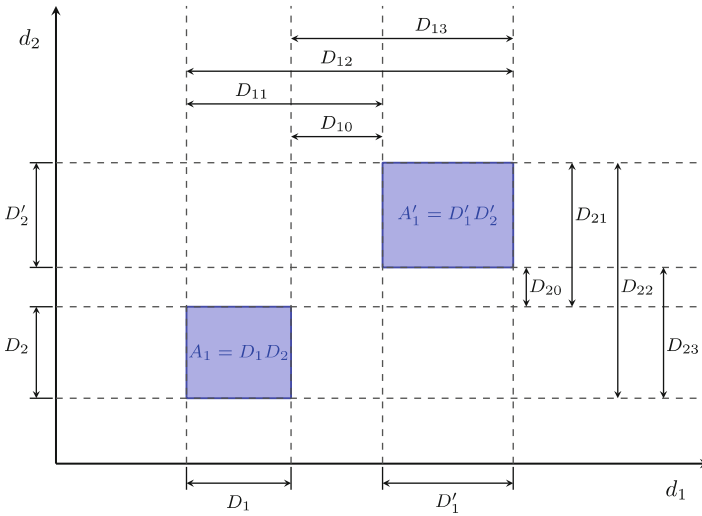
**Fig. 4.16** Variance function for a regular square grid



**Fig. 4.17** Definition of geometry used in Eq. (4.29) (Redrawn from Vanmarcke (1983))

$$\rho\left(\ln y_{A_1}, \ln y_{A_2}\right) = \frac{\sigma^2}{4A_1 A_2 \sigma_{A_1} \sigma_{A_2}} \sum_{k=0}^{3} \sum_{l=0}^{3} (-1)^k (-1)^l (D_{1k} D_{2l})^2 \gamma(D_{1k}, D_{2l}) \quad (4.29)$$

The correlations that are generated using this approach are shown in Fig. 4.18 both in terms of the correlation against separation distance of the cell centroids and in terms of the correlation against the separation measured in numbers of cells. Figure 4.18 shows that the correlation values can be significantly higher than the corresponding point-estimate values (which lie close to the case for the smallest
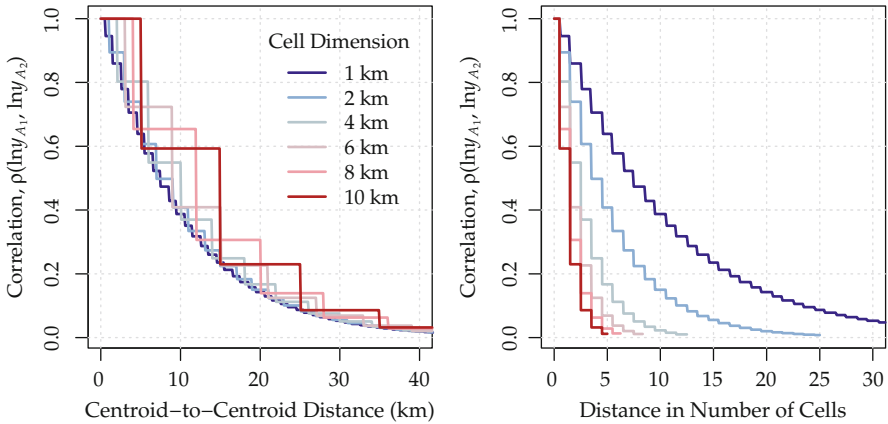
**Fig. 4.18** Example correlations computed using Eq. (4.29) for square cells of differing dimension

dimension shown). However, the actual covariances do not differ as significantly due to the fact that these higher correlations must be combined with the reduced variances.

## 4.6  Conclusions

Empirical ground-motion modelling is in a relatively mature state, but the historical emphasis has been biased towards median predictions with the result that the characterisation of ground-motion variability has been somewhat neglected. This paper emphasises the importance of the variance of the ground-motion distribution and quantifies the sensitivity of hazard results to this variance. The partitioning of total uncertainty in ground-motion modelling among the components of aleatory and epistemic uncertainty is also revisited and a proposal is made to relax the definitions that are often blindly advocated, but not properly understood. A new approach for selecting an optimal model complexity is proposed. Finally, a new framework for generating correlated discrete random fields is presented.

# References

Campbell KW, Bozorgnia Y (2008) NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5 %-damped linear-elastic response spectra for periods ranging from 0.01 to 10.0 s. Earthq Spectra 24:139–171

Campbell KW, Bozorgnia Y (2014) NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5%-damped linear acceleration response spectra. Earthq Spectra. http://dx.doi.org/10.1193/062913EQS175M

Chiou BSJ, Youngs RR (2008) An NGA model for the average horizontal component of peak ground motion and response spectra. Earthq Spectra 24:173–215

Chiou BSJ, Youngs RR (2014) Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. Earthq Spectra. http://dx.doi.org/10.1193/072813EQS219M

Der Kiureghian A, Ditlevsen O (2009) Aleatory or epistemic? Does it matter? Struct Saf 31:105–112

Elms DG (2004) Structural safety – issues and progress. Prog Struct Eng Mat 6:116–126

Jayaram N, Baker JW (2009) Correlation model for spatially distributed ground-motion intensities. Earthq Eng Struct D 38:1687–1708

Li H, Meng G (2007) Nonlinear dynamics of a SDOF oscillator with Bouc-Wen hysteresis. Chaos Soliton Fract 34:337–343

Stafford PJ (2012) Evaluation of the structural performance in the immediate aftermath of an earthquake: a case study of the 2011 Christchurch earthquake. Int J Forensic Eng 1(1):58–77

Vanmarcke E (1983) Random fields, analysis and synthesis. The MIT Press, Cambridge, MA