A FRAMEWORK FOR VALIDATION OF SEISMIC RESPONSE ANALYSES USING SEISMOMETER ARRAY RECORDINGS

Brendon A Bradley

Department of Civil and Natural Resources Engineering, University of Canterbury, Private Bag 4800, Christchurch 8020, New Zealand, brendon.bradley@canterbury.ac.nz

ABSTRACT

A framework for the validation of computational models used to predict seismic response based on observations from seismometer arrays is presented. The framework explicitly accounts for the epistemic uncertainty related to the unknown characteristics of the 'site' (i.e. the problem under consideration) and constitutive model parameters. A mathematical framework which makes use of multiple prediction-observation pairs is used to improve the statistical significance of inferences regarding the accuracy and precision of the computational methodology and constitutive model. The benefits of such a formal validation framework include: (i) development of consistent methods for determination of constitutive model parameters; (ii) rigorous, objective and unbiased assessment of the validity of various constitutive models and computational methodologies for various problem types and ground motion intensities; and (iii) an improved understanding of the uncertainties in computational model assumptions, constitutive models and their parameters, relative to other seismic response uncertainties such as ground motion variability. Details regarding the implementation of such a framework to achieve the aforementioned benefits are also addressed.

INTRODUCTION

The continuing evolution toward the seismic design of engineered facilities based on their expected seismic performance places increasing emphasis on the use of computational models to predict the seismic response of such facilities. Despite our best efforts in the design and assessment of facilities to reduce their vulnerability to earthquake-induced hazards, the occurrence of every large earthquake provides new evidence of the complex phenomenon producing strong ground motions at the earth's surface, and weaknesses in these contemporary seismic design and/or assessment methods [1-3].

Quantitative data from seismometer arrays [e.g. 4, 5] represent one of the primary interactions between observations and computational simulation in earthquake engineering, with other interactions including: element testing, testing of subsystems, or testing of entire systems at full or reduced scales. Seismometer data offers several advantages over these other forms of quantitative data in that the instrumented facilities automatically have the correct in situ and boundary conditions which can be difficult, if not impossible, to reproduce in laboratory experiments. The reducing costs of deploying and maintaining seismometer arrays, as well as these perceived benefits are leading to a significant increase in the number, configuration and types of structures (both natural and man-made) being instrumented throughout seismically active areas of the world, e.g. [6-8].

This manuscript is devoted to the development of a framework in which seismic response models can be validated with seismic array recordings. Firstly, details regarding the concepts of verification, validation and prediction as applied to seismic response modelling are discussed. The conventional use of seismometer arrays in validation of seismic response modelling and its limitations are discussed. The details of the proposed framework, which addresses conventional limitations, are developed and its benefits for use in seismic response prediction are examined. Finally, procedural aspects regarding the implementation of the framework in order to realise its stated benefits are discussed.

VALIDATION IN SEISMIC RESPONSE MODELLING

Computational seismic response models are used to predict the response of engineered facilities in future seismic events. Verification and validation are the primary means by which confidence can be built as to the predictive capabilities of a computational model [9]. Verification is the assessment of the accuracy of the computational implementation of a conceptual model, while validation is concerned with the assessment of the degree to which the (computational implementation of the) conceptual model is representative of reality [9].

Conventionally, the validity of seismic response models is examined by primarily three means, which examine different aspects of system behaviour, as illustrated in Figure 1. Firstly, element tests are used to gain an understanding of fundamental material behaviour. Secondly, model subsystem tests offer insight into the interaction of the various components of a system which is not possible in element tests. Finally, system-level tests examine global system response and can be conducted in both laboratory and field environments.

In spite of the increasing capacity of laboratory facilities, there remain many seismic response problems for which system-level testing at full-scale or reasonably reduced-scales is not possible (e.g. high rise buildings, and large earth structures). Furthermore, while the laboratory-type nature of such tests allow a high degree of control in the preparation of the specimen and applied excitation, the primary limitation is the difficultly in representing the appropriate in situ and boundary conditions (as well as possible scale and load history effects). Conversely, seismic instrumentation of various engineered facilities provides a means to obtain instrumental records of the seismic response of full-scale specimens, with appropriate in situ and boundary conditions. An undesirable consequence of using seismic arrays for seismic response validation, however, is that the complexity of this real-world ‘specimen’ means that its characterisation is significantly more uncertain than a laboratory equivalent.

CONVENTIONAL COMPARISON OF SEISMIC RESPONSE MODELS WITH ARRAY OBSERVATIONS

Examples of the use of seismometer arrays to examine the capabilities of geotechnical site response analysis include Cubrinovski *et al*. [10]; Pecker [11]; Youd *et al*. [12]; Bernardie *et al.* [13]; Elgamal *et al.* [14]; Finn *et al.* [15]; and Kawano [16], among others. Figure 2 illustrates a schematic example of a seismic instrumentation array which can be used to validate a one-dimensional seismic response analysis. A computational seismic response model could be constructed, and the input excitation applied to the model can be obtained from one (or possibly more) of the seismometer recordings. Thus, in this case, the 'input' motion (i.e. that at the base of the one-dimensional computational model in Figure 3) is known explicitly, such that any difference between the computational prediction and seismometer observations is due to the computational model (including how the input motion is applied as a boundary condition). This feature is a prerequisite to enable seismic arrays to be used for seismic response validation.

The conventional use of seismometer arrays for validation of seismic response computational models, as exemplified by the aforementioned references, can be regarded as deterministic in the sense that no uncertainties in the seismic response model are considered. It was previously mentioned that one of the consequences of constructing a computational model of a site which exists in reality (rather than a model which is created under laboratory-type conditions) is that it is not possible to fully characterise the physical and mechanical properties of the site. Hence there exists significant uncertainty in the characterisation of the problem under consideration. This consequently results in uncertainty in the parameters of the constitutive models required for computational analyses. A consequence of the failure to account for these uncertainties in the computational model is that it cannot be determined if a good agreement between a single prediction and observation is due to a capable computational model or is in fact due to 'cancellation' of errors that result from the unknown site characterisation and inconsistencies in the computational model.

As also exemplified by the aforementioned references, the comparison between the results of the computational model and the seismometer observations is commonly performed by the same personnel who create the computational model itself. Thus, in such cases the so-called 'prediction' is potentially in fact merely 'post-test consistency' [9]. True confidence in the predictive capability of a computational model can only be obtained when model development and simulation is kept independent from validation via comparison with observational data [9].

The following section presents a framework which, amongst other things, explicitly accounts for uncertainties which are conventionally not explicitly considered. This framework consequently allows inference as to the true predictive capability of the computation model. A subsequent section elaborates upon the procedural and implementation aspects of the proposed framework in order to achieve its stated benefits.

A FORMAL FRAMEWORK FOR SEISMIC RESPONSE VALIDATION

Uncertainty classification

There are numerous significant uncertainties in any seismic response problem, and these must be considered in a robust validation framework. Here, such seismic response uncertainties are differentiated into four classes (Figure 3): (i) site characterisation uncertainties; (ii) constitutive model parameter uncertainties; (iii) constitutive model uncertainties; and (iv) model methodology uncertainties. An example of type (i) uncertainty is the unknown value of the shear modulus of surficial soils obtained from seismic site surveys (Figure 3a). That is, type (i) uncertainties are the uncertainty in quantities which can be directly measured for the problem of interest. There is also potential uncertainty (type (ii)) in the parameters of the constitutive models used in the seismic response analysis, which are often obtained from empirical correlations with site characterisation data. Note that some constitutive model parameters are directly measured (i.e. the shear modulus of soil) and therefore have no type (ii) uncertainty. Constitutive models themselves are often empirically constructed based on observations, or theoretically derived based on simplifiying assumptions. Hence, there exists uncertainty (type (iii)) resulting from the use of a specific constitutive model in a seismic response analysis (Figure 3c). Finally, the adopted computational model methodology and domain always represents a simplification of the 'real world' problem, and therefore also contains uncertainty (type (iv)). Common simplifications include the use of one or two dimensional analyses (Figure 3d), neglection of soil-structure interaction (in structural analyses), assumed boundary conditions, damping formulations, and ground motion input (i.e. wave propagation assumptions), among others.

It should be mentioned that uncertainties in addition to those mentioned in the previous paragraph exist, both in the computation model and seismic array recordings. For example, in the computational model there are uncertainties due to the discrete solution of the continuum problem, computer precision/round-off etc. These are uncertainties which can be quantified (and minimised) by proper verification. There are also uncertainties in the seismometer array recordings (i.e. the validation experiments), e.g. instrument accuracy and calibration. In general (i.e. in the case of adequate verification and 'nominally' functioning seismometer array equipment) these uncertainties are believed to be small relative to the site and constitutive model uncertainties and are therefore not explicitly considered in the framework herein.

Consideration of site characterisation and constitutive model parameter uncertainties

Consider initially that the computational model methodology and constitutive relations are an exact representation of the physical problem of interest. Therefore the only uncertainties in the seismic response predicted by the computational model are related to the exact characterisation of the mechanical, physical and geometric properties of the materials in the seismic response problem. This includes both uncertainties in the measured values of mechanical and physical properties and the uncertain relationships between measured properties and the parameters of constitutive relationships (i.e. type (i) and (ii) uncertainties in Figure 3).

When type (i) and (ii) uncertainties are considered in the computational model then the resulting seismic response, measured by one of more engineering demand parameters (EDPs), will have a distribution (with each EDP having a different value for each possible realisation of the uncertain parameters). Figure 4a illustrates this uncertainty in the form of a probability density function of a predicted EDP from the computational model. Figure 4a also illustrates the unique value of the seismic response quantity, $edp_{i,j,k}$, as measured from the seismometer array.

The probability density function (pdf) of the prediction of a particular demand measure, EDP_i, (e.g. peak displacement at the surface) for a single observation k, at a single site j, $f_{EDP_{i,j,k}}$, shown in Figure 4a gives the likelihood that a particular value of EDP_i is observed based on the computational model. A comparison of $f_{EDP_{i,j,k}}$ with the actual observation, $edp_{i,j,k}$, cannot however be used to clearly determine the capability of the computational model methodology and constitutive model for seismic response prediction. This is because an observation, $edp_{i,j,k}$, which is notably different than the mode of $f_{EDP_{i,j,k}}$ is merely an observation that, while less likely to be observed than the mode itself, is possible nonetheless. To make more robust inferences on the predictive capability of a computational model multiple data (i.e. multiple observation-prediction pairs) are needed.

Consideration of multiple observations and sites

Consideration of the type (i) and (ii) uncertainties for a specific problem leads, firstly, to the realisation that these uncertainties are epistemic, that is, with improved site characterisation and better constitutive model parameter correlations, these uncertainties can be reduced. Secondly, because the uncertainties are specific to the site of interest they are therefore independent from site-to-site. Finally, multiple seismic response predictions at a single site are dependent because of the unique (albeit unknown) site under consideration. Keeping in mind the second and third points above it is desired to use multiple comparisons between seismic response model predictions and seismometer observations at multiple sites to improve the statistical significance of inferences regarding the capability of a computational model to predict a certain type of seismic response.

Consider the uncertain prediction from the computational model in terms of the cumulative density function (CDF) shown in Figure 4b (rather than the pdf in Figure 4a). Using this CDF the actual seismometer observation (the k^{th} observation at site j of EDP_i), $edp_{i,j,k}$, corresponds to a value $F_{EDP_{i,j,k}}\left(edp_{i,j,k}\right)$. The normalised residual of the seismometer observation for $edp_{i,j,k}$ relative to the computational model prediction can then be computed from:

$$z_{i,j,k} = \Phi^{-1}\left[F_{EDP_{i,j,k}}\left(edp_{i,j,k}\right)\right] \tag{1}$$

where $\Phi^{-1}[\;]$ is the inverse normal cumulative density function.

Based on its definition, $z_{i,j,k}$ represents a random observation from a standard normal distribution. In order to account for the dependence between multiple observations at a single site this normalised (total) residual is expressed as:

$$z_{i,j,k} = a + \eta_{i,j} + \varepsilon_{i,j,k} \tag{2}$$

where a is a constant; $\eta_{i,j}$ is the inter-site residual for EDP_i and site j, and $\varepsilon_{i,j,k}$ is the intra-site residual for the k^{th} observation of EDP_i at site j. It is assumed that $\eta_{i,j}$ and $\varepsilon_{i,j,k}$ are independent and are characterised by a normal distribution with zero means and variances σ_S^2 and σ_O^2, respectively. Since the sum of two normal distributions is also normal, then the formulation in (2) is clearly compatible with the fact that $z_{i,j,k}$ has (based on the stated assumptions) a standard normal distribution.

Figure 5 illustrates schematically how multiple sites and observations can be considered and reiterates the adopted notation. For each different site considered, which are classed as being of a similar type (addressed more rigorously in the following sections), a computational model is developed using the adopted computational model methodology and constitutive model under consideration. For each of the k observations at site j the observed value of EDP_i, $edp_{i,j,k}$ is compared with the predicted distribution using the computational model, $f_{EDP_{i,j,k}}$, to determine the (total) normalised residual, $z_{i,j,k}$.

The use of Equation (2) with multiple prediction-observation data represents a linear mixed effects model for repeated measures data [17]. Because such models are common in engineering and applied sciences, there exist good numerical algorithms and software for their solution (e.g. [18]). Upon conducting regression to determine the unknown parameters in Equation (2) (i.e. a, σ_S^2 and σ_O^2), the mean and variance of the regression model of the normalised residuals, are given by:

$$\hat{\mu}_Z = a \tag{3}$$

$$\hat{\sigma}_Z^2 = \sigma_S^2 + \sigma_O^2 \tag{4}$$

where $\hat{\mu}_Z$ is the point-estimate of the mean of Z; and $\hat{\sigma}_Z^2$ is the point-estimate of the variance of Z.

Based on the aforementioned assumption that the computational methodology and constitutive model are exact (hence only type (i) and (ii) uncertainties in Figure 3 are present), each $z_{i,j,k}$ represents a random variable from a standard normal distribution. Hence, comparison of the mean and variance of the regression model for Z with that of a standard normal distribution (which has a mean of 0 and variance of 1) can be used to examine the bias and precision of the computational methodology and constitutive model.

While, Equations (3) and (4) give only the point-estimates of mean and variance of the prediction residuals confidence intervals for these parameters for a given significance level can easily be obtained from bootstrap sampling. Here, the (1-α)% confidence interval for λ_Z is denoted as $\left[\lambda_{\alpha/2}, \lambda_{1-\alpha/2}\right]$, where $\lambda = \mu$ and σ^2 for the mean and variance, respectively. The critical value of the confidence interval, α_{crit}, for μ_Z and σ_Z^2 (i.e. the largest α value which includes the theoretical value of the parameter) can then be used as a measure of the degree of bias and precision, respectively, of the computational model methodology and constitutive model.

FRAMEWORK APPLICATION

Hypothetical observations

Figure 6 illustrates possible situations which may arise when comparing the predicted distribution of *Z* (i.e. using Equations (2)-(4)), with the theoretical standard normal distribution for a particular seismic response problem. Figure 6a illustrates the case in which the mean and variance of *Z* are very similar to the standard normal distribution. It can be seen that the 90% confidence interval of μ_Z easily encompasses the theoretical value of zero (i.e. α_{crit} >> 0.05), and hence the bias of the computational model methodology and constitutive model for the sites considered is relatively small. Figure 6b illustrates a situation where the computational model systematically over predicts the response for some EDP_i, resulting in residuals which are predominantly negative. This over-prediction bias is significant as can be seen from the 90% confidence interval for μ_Z not including the theoretical value of zero (i.e. α_{crit} < 0.05). Figure 6c illustrates a situation in which there is little bias in the computational model (similar to Figure 6a), but that the variance of Z, σ_Z^2, is significantly larger than that of the theoretical value of 1 (i.e. α_{crit} < 0.05), indicating that the computational model is imprecise. Figure 6c, in particular, represents a case in which additional uncertainty (either, type (iii) constitutive model; or type (iv) model methodology) needs to be included in the computational prediction.

Until now it has been explicitly assumed that the model methodology and constitutive model were exact (such that only type (i) and (ii) uncertainties were present). What has also being assumed in the interpretation of Figure 6 in the previous paragraph is that the quantitative models which define the type (i) and (ii) uncertainties are also correct. That is, with reference to Figure 3 it has been assumed that the probabilistic models (in particular the uncertainty in the models) for the shear value velocity with depth and the constitutive model parameter with SPT blowcount are correct. If, for example, the probabilistic model for the shear wave velocity contains significantly more uncertainty than actually exists, then the probability density function of the predicted demand, $f_{EDP_{i,j,k}}$, will also have an increased uncertainty, and consequently the normalised residuals will have a smaller variability (i.e. it is less likely to observe an normalised residual which is significantly different than zero). In such cases it may be possible to observe the situation depicted in Figure 6d in which σ_Z^2 is significantly less than one.

One further situation of particular importance, similar to the discussion above is where the type (i) and (ii) uncertainties are too large (i.e. similar to Figure 6d), but that the computational model methodology and constitutive model are also imprecise (i.e. similar to Figure 6c), such that the combined observation is a distribution of normalised residuals with a variance, σ_Z^2, similar to one (i.e. Figure 6a). In such a case an analyst may therefore wrongly conclude that both computational model methodology and constitutive model are precise, and that the type (i) and (ii) uncertainties are appropriately considered. While such a situation is not-ideal (and can be avoided using a multi-tiered hierarchy as to be discussed), what is effectively happening is that model methodology and constitutive model (type (iii) and (iv)) uncertainties are being re-parameterised as parameter (type (i) and (ii)) uncertainty (e.g. [19]). This, in fact, has the benefit that multiple models are not required to account for type (iii) and (iv) uncertainties, including how to handle inter-model dependence [19].

Consideration of alternative constitutive models

Because constitutive models used in seismic response analyses are typically empirically constructed based on direct observations (e.g. [20, 21]), or theoretically derived based on various assumptions (e.g. [22, 23]), then it is unlikely that a single constitutive model is perfectly representative of an engineering material. Consequently, this imperfection leads to uncertainty in the prediction of the seismic response of such a material (i.e. type (iii) uncertainty in Figure 3). As noted in the previous section, this constitutive model uncertainty can be accounted for by: (i) considering multiple constitutive models in the seismic response analysis; or (ii) re-parameterising constitutive model uncertainty in the form of constitutive model parameter uncertainty (i.e. type (ii) uncertainty in Figure 3).

The formal consideration of constitutive model uncertainty in seismic response analysis is a relatively undeveloped area in comparison with the emphasis it is given in similar topics such as seismic hazard analysis (PSHA) [24, 25]. Notwithstanding this, it is well recognised that significant differences in computational model predictions can be obtained using various commonly adopted constitutive models for certain problems [26].

In contemporary PSHA, the determination of a model hierarchy or relative belief in alternative models is difficult as such a determination is often subjectively based on expert opinion (e.g. [27, 28]). The use of expert opinion is often also further complicated by the fact that such expert personnel are often also the developers of the various alternative models, therefore potentially compromising their assessment of model hierarchy. Furthermore, the use of subjective expert opinion also leads to the tendency to include too many models which are often: (i) highly dependent, resulting in redundancy; or (ii) not plausible, resulting in a potentially significant over-estimation of model uncertainty [27, 28].

The proposed seismic response analysis validation framework offers the opportunity to quantify a hierarchy of constitutive model validity based on the observed bias and precision of the alternative models, and therefore avoid problems associated with a significant reliance on expert opinion. Figure 7a schematically illustrates the distribution of the normalised residuals for a given computational model methodology, but using three different constitutive models. It can be seen that the use of constitutive models 1 and 2 leads to a small over-prediction and under-prediction bias, respectively, as indicated by the small negative and positive mean values of the normalised residuals, respectively. It is also noted that the use of constitutive models 1 and 2 leads to an appropriate level of prediction precision (as indicated by the similarity in the variance of the normalised residuals relative to the theoretical standard normal distribution). On the other hand, the use of constitutive model 3 leads to a large over-prediction bias, as indicated by the mean value of the normalised residuals being significantly different than zero. It addition, the variance of the normalised residuals obtained using constitutive model 3 is significantly larger than one, indicating that the use of constitutive model 3 also leads to significant prediction imprecision. Hence on the basis of Figure 7a, an analyst could comfortably reject the use of constitutive model 3 (for a seismic response problem which is 'within' those encompassed by the array recordings providing the observed normalised residuals), and consider only constitutive models 1 and 2 when accounting for constitutive model (type (iii)) uncertainty. Furthermore, the critical values, α_{crit}, for the mean and variance for each of the alternative constitutive models considered can be used to develop a hierarchy of model belief, although a specific method to do so is not presented here.

Consideration of multiple EDPs

The presented discussion in the previous sections was limited to the consideration of a single measure of seismic demand, EDP_i. However, the same framework can obviously be applied to a vector of engineering demand parameters, $\boldsymbol{EDP}$ (e.g. peak surface acceleration, displacement, arias intensity etc.). Obviously the selection of which EDP_i's are considered in this vector will differ from problem-to-problem but should include all those that are conventionally used to assess seismic response for the problem considered. It is important to note that it should not be expected that all measures of seismic response for a particular problem type can be predicted with a similar level of accuracy and precision. For example, in the prediction of the seismic response of pile founded structures located near quay walls [29, 30] it was clearly identified that prediction of the maximum lateral displacement of the quay wall was significantly more difficult than prediction of the maximum lateral displacement of the pile founded structure. In this regard it is valuable to note that the dimensionless nature of the normalised residuals allows for a clear comparison between the accuracy and precision with which various measures of seismic response can be predicted.

The consideration of multiple EDP_i's (at the same point and/or spatially different locations of the computational model) is also important in examining whether a computational model is able to capture the key deformation and possible failure mechanisms which occur in a particular observation. For example, it may be possible to accurately predict (in a quantitative sense) the peak strain at the surface of a soil deposit using an equivalent-linear site response model, even if the actual response is highly non-linear. However, it would be expected that the prediction from such a model contain significantly larger error in the prediction of other measures (i.e. peak acceleration, cumulative energy etc.) of the surface ground motion. Conversely, it is also possible that a computational model may predict the correct deformation mechanism, but over- or under-predict all the EDPs considered. While such a case is unlikely to be observed frequently if a sufficient number of prediction-observation pairs and an appropriate vector of EDPs are considered, it emphasises the fact that the proposed procedure is intended as a complement to, and not a replacement for, a detailed (deterministic) examination of the computational model prediction.

PREDICTION CAPABILITY FOR DIFFERENT PROBLEM TYPES

In the previous section it was discussed how multiple prediction-observation pairs (and consequently multiple normalised residuals) for ‘similar’ sites can be used to improve the statistical significance of computational model validation and consequently the assessment of computational model prediction capability. Sites are considered to be ‘similar’ if the physical systems for which the seismic response is to be predicted, lend themselves to be computed via ‘similar’ computational models. The computational models are deemed to be ‘similar’ if they adopt the same constitutive models or computational model methodology. For example, a set of similar sites maybe all of the down-array soil sites in Japan which are instrumented as part of the KiK-Net project [5, 6]. The computational model methodology adopted for such sites in this case could be one-dimensional wave propagation models to predict the seismic response of surficial soil deposits. Furthermore, all the one-dimensional models could use the same constitutive model (with the constitutive model parameters obviously being a function of the particular soils encountered). In this case, the results which would be observed in the form of Figure 5-Figure 7 would provide validation for the use of one-dimensional models of site response using the particular constitutive model.

Clearly, the number of prediction-observation pairs which could be used to obtain normalised residuals in this example would be large [5, 6], and based on the observed distribution of the normalised residuals (i.e. Figure 5) it may be desirable to understand the predictive capability of the computational model within these such sites as a function of one or more ‘problem characterisation parameters’. Such problem characterisation parameters could be the intensity of the input ground motion observed at the base of the down-hole array, or the aspect ratio (depth divided by width) of the sedimentary basin, among others. In order to examine such trends, the distributional properties of $z_{i,j,k}$ can be considered as dependent on the various problem characterisation parameters, ***X***, of interest. When this dependence on ***X*** is considered, the mixed-effects regression of the normalised residuals can be expressed as:

$$z_{i,j,k} = f\left(\boldsymbol{X}\right) + \eta_{i,j} + \varepsilon_{i,j,k} \tag{5}$$

where $f\left(\boldsymbol{X}\right)$ represents some pre-determined function (with unknown parameters obtained from regression) of the site characterisation parameters; and $\eta_{i,j}$ and $\varepsilon_{i,j,k}$, as before, are the inter-site and intra-site residuals with zero means, but now with variances $\sigma_S^2(\boldsymbol{X})$ and $\sigma_O^2\left(\boldsymbol{X}\right)$, respectively (i.e. variances that are potentially a function of ***X***). The point-estimates of the mean and variance of the model of the residuals in Equation (5) is hence given by:

$$\hat{\mu}_z\left(\boldsymbol{X}\right) = f\left(\boldsymbol{X}\right) \tag{6}$$

$$\hat{\sigma}_z^2\left(\boldsymbol{X}\right) = \sigma_S^2\left(\boldsymbol{X}\right) + \sigma_O^2\left(\boldsymbol{X}\right) \tag{7}$$

As previously mentioned the confidence intervals of $\mu_z\left(\boldsymbol{X}\right)$ and $\sigma_z^2\left(\boldsymbol{X}\right)$ can be used as indicators of the bias and precision of the computational methodology and constitutive model as a function of the site characterisation parameters, ***X***. Figure 8 illustrates two possible cases which may be observed. With reference to the aforementioned site response example, Figure 8a may represent that as the intensity of the input ground motion (the problem characterisation parameter) increases the computational model systematically over-predicts (leading to negative normalised residuals) the observed peak accelerations at the ground surface. On the other hand, Figure 8b may illustrate that an increasing basin aspect ratio (the problem characterisation parameter), which leads to an increasing effect of two-dimensional wave propagation, causes a significant increase in the imprecision of the computational model (Figure 8b illustrates that while imprecise, the computational model does not illustrate significant bias, although this could be present also). It should finally be noted that Equations (5)-(7) can be used both for problem characterisation variables which are continuous (as in Figure 8) or discrete (e.g. soil classification such as sand, clay, and sand with fines is one possible discrete variable in computational site response modelling).

Clearly, examination of the normalised residuals as a function of multiple variables within a single problem class offers the potential of establishing a 'multi-tiered' validation hierarchy [9]. Such a validation hierarchy is important to enable confidence in various aspects of a complex computational model to be developed. For example, extensively validating a non-linear inelastic constitutive model for seismic response analysis for ground motions of small intensity (i.e. in which the response is essentially linear elastic) can be used to gain confidence in the other aspects of the seismic response model (i.e. the computational model methodology and the probabilistic models for type (i) and (ii) uncertainties which effect the computational model for such linear elastic response). As computational seismic response models may contain more than a single material, and therefore potentially require multiple constitutive models, the use of a multi-tiered validation hierarchy can be used to develop confidence in each of the various constitutive models, and therefore attribute any potential bias and imprecision identified observed from the normalised residuals to specific features of the computational model.

The proposed validation framework has so far being presented in a general sense, but its application for specific problems requires potentially additional considerations. As the complexity of the seismic response analysis increases the importance of an 'expert analyst' becomes pivotal in the development of a computational model based on interpretation of the physical problem. Therefore the ‘expertise’ of an analyst represents a further uncertainty in the seismic response analysis. Although no attempt is made here to consider this in the proposed framework, it follows that those computational models which require the least input from analysts (e.g. total stress equivalent-linear one-dimensional site response) are more directly amenable to utilization of the proposed framework. The proposed framework is still applicable however to complex computational models, but care is needed to ensure that the obtained results are not devoid of the correct use and interpretation of the model which an expert user offers.

PROCEDURAL ASPECTS OF VALIDATION TESTING

Application of the proposed validation framework is relatively straightforward in that open-source software for the mixed effects regression model (i.e. Equations (2) and (5)) is available [18], and hence the key tasks are the development of the computational models and access to the seismometer array data. While use of this framework by individual computational analysts is still a step forward compared with the conventional use of seismometer recordings for validating seismic response computational models, such an approach still has two majority shortcomings: (i) the credibility of the comparison is not guaranteed, because computational model developers act as testers; and (ii) a comparison of alternative models may not be possible due to the use of different metrics for comparison (i.e. different EDP’s used). It is emphasised by Schorlemmer and Gerstenberger [31] that similar shortcomings have severely affected progress in earthquake likelihood forecasting-related research.

In order to avoid the shortcomings mentioned above an appropriate organisational setup and a corresponding ‘set of rules’ are needed in order to achieve the full potential of the proposed validation framework. A ‘testing centre’ cyber infrastructure, in which a multitude of different computational seismic response models (i.e. different computational model methodologies, and constitutive models for different sites with seismometer arrays) are internally installed makes it possible to achieve truly prospective (i.e. real-time), reproducible and unbiased testing. A similar type of testing centre has recently being established for testing earthquake likelihood models [31], which offers the four key benefits of: (i) transparency - all the inputs and outputs provided by the multitude of computational models can be tracked; (ii) a controlled-environment - all computational models are provided with the identical input motion for each observation, and computational model development is kept independent from testing against observations; (iii) comparability - all computational models for a given problem class will be tested not only against observations, but also against each other; and (iv) reproducibility - computational models are stored internally so testing can be rerun at a later date.

IMPLICATIONS FOR UNCERTAINTY ASSESSMENT IN SEISMIC RESPONSE MODELLING

Uncertainties in earthquake occurrence and consequent ground motion intensity have been considered using probabilistic seismic hazard analysis for over four decades [32], while explicit treatment of uncertainties in the seismic response of the built environment has only begun to be widely considered within the last decade [33]. The majority of works explicitly addressing uncertainties in the response of the built environment have constrained their focus to uncertainties in the seismic demand resulting from so-called record-to-record variability (e.g. [34, 35]), while refraining from considering uncertainty in the computational model of the problem under consideration. This is clearly a result of the fact that consideration of record-to-record randomness via the use of multiple input ground motions is simple from both a conceptual and implementation viewpoint (although the manner in which such ground motions are selected is often inconsistent [36]). On the other hand, the consideration of computational model uncertainty is complicated by the fact that it can vary significantly between different computational model methodologies, constitutive models, and problems under consideration. To this end, it is believed that the proposed validation framework can provide significant progress toward the consideration of uncertainties in computational seismic response models. Firstly, use of the proposed validation framework requires the development of probabilistic models for uncertainty in the physical, material and geometrical properties of the problem under consideration based on field investigations (i.e. type (i) uncertainty in Figure 3). Secondly, the validation framework also requires the development of probabilistic models for the relationships between measured properties at the site under consideration and the parameters of the constitutive models which will be used within the computational model (i.e. type (ii) uncertainty in Figure 3). Thus, the validation framework requires as a prerequisite that uncertainties in the parameters of constitutive models can be determined due to uncertainties in the site and constitutive model parameter correlations. Thirdly, the validation framework can be used to assess the predictive capability of, and hence uncertainty in using, computational models developed based upon certain computational model methodologies and constitutive models (i.e type (iii) and (iv) uncertainties in Figure 3). Therefore the framework offers a quantitative means in which the magnitude of these four types of computational model uncertainties can be estimated for different problems. Such uncertainties can therefore be compared with other uncertainties, such as record-to-record randomness. Furthermore, an appreciation of the locations at which the major uncertainties in the seismic response problem are apparent will enable more efficient allocation of resources in order to improve seismic performance estimation.

CONCLUSIONS

This manuscript has presented a framework for the validation of computational models used to predict seismic response based on observations from seismometer arrays. The framework explicitly accounts for the epistemic uncertainties related to the unknown characteristics of the 'site' (i.e. the problem under consideration) and constitutive model parameters. The statistical significance of inferences regarding the accuracy and precision of the computational modelling methodologies and constitutive models is enhanced via the use of a mixed-effects model in which multiple prediction-observation pairs are considered. The benefits of the formal validation framework include: (i) development of consistent methods for determination of constitutive model parameters; (ii) rigorous, objective and unbiased assessment of the validity of various constitutive models and computational methodologies for various problem types and ground motion intensities; and (iii) an improved understanding of the uncertainties in computational model assumptions, constitutive models and their parameters, relative to other seismic response uncertainties such as ground motion variability. It is further proposed that the validation framework be implemented in a 'testing centre' cyber-infrastructure in order to be provide results in a controlled environment which are transparent, comparable, and reproducible, and hence provide the greatest benefit for validation of computational seismic response models used within the earthquake engineering community.

REFERENCES

Japanese Geotechnical Society. Special issue on geotechnical aspects of the January 17, 1995 Hyogoken-Nambu earthquake. *Soils and foundations* 1998.

Gates W and Morden M. *Lessons from Inspection, Evaluation, Repair and Construction*, SAC Joint Venture: Sacramento, 1995. pp.

Hanks TC and Brady AG. *The Loma Prieta earthquake, ground motion, and damage in Oakland, Treasure Island, and San Francisco*, 1991, 2019-2047pp.

Brady AG. *Strong-motion accelerographs: Early history*, 2009, 1121-1134pp.

National Institute For Earthquake and Disaster Prevention. *Kyoshin Network (K-Net)* 2010, pp.
Kinoshita S. Kyoshin Net (K-NET). *Seismological Research Letters* 1998; **69**(4): 309–332.
GeoNet. *GeoNet*, 2010, pp.
United States Geological Survey. *USGS National strong motion program (NSMP)*, 2010, pp.
Oberkampf WL, Trucano TG and Hirsch C. *Verification, validation, and predictive capability in computational engineering and physics*, in *Foundations for Verification and Validation in the 21st Century Workshop*: Laurel, Maryland, 2002, 74pp.
Cubrinovski M, Ishihara K and Furukawazono K. *Analysis of two case histories on liquefaction of reclaimed deposits*, in *12th World Conference on Earthquake Engineering* Auckland, New Zealand, 2000, pp.
Pecker A. Validation of small strain properties from recorded weak seismic motions. *Soil Dynamics and Earthquake Engineering* 1995; **14**(6): 399-408.
Youd TL, Steidl JH and Nigbor RL. Lessons learned and need for instrumented liquefaction sites. *Soil Dynamics and Earthquake Engineering* 2004; **24**(9-10): 639-646.
Bernardie S, Foerster E and Modaressi H. Non-linear site response simulations in Chang-Hwa region during the 1999 Chi-Chi earthquake, Taiwan. *Soil Dynamics and Earthquake Engineering* 2006; **26**(11): 1038-1048.
Elgamal AW, Zeghal M, Parra E, Gunturi R, Tang HT and Stepp JC. Identification and modeling of earthquake ground response -- I. Site amplification. *Soil Dynamics and Earthquake Engineering* 1996; **15**(8): 499-522.
Finn WDL, Ventura CE and Wu G. Analysis of ground motions at Treasure Island site during the 1989 Loma Prieta earthquake. *Soil Dynamics and Earthquake Engineering* 1993; **12**(7): 383-390.
Kawano M, Asano K, Dohi H and Matsuda S. Verification of predicted non-linear site response during the 1995 Hyogo-Ken Nanbu earthquake. *Soil Dynamics and Earthquake Engineering* 2000; **20**(5-8): 493-507.
Lindstrom MJ and Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics* 1990; **46**(3): 673-687.
Pinheiro J, Bates DM, DebRoy S, Sarkar D and the R Core team. *nlme: linear and nonlinear mixed effects models*, 2008. pp.
Page MT and Carlson JM. Methodologies for earthquake hazard assessment: Model uncertainty and the WGCEP-2002 forecast. *Bulletin of the Seismological Society of America* 2006; **96**(5): 1624-1633.
Otani S. *SAKE, a computer program for inelastic response of R/C frames to earthquakes*, University of Illinois at Urbana-Champaign, 1974. pp.
Ibarra LF and Krawinkler H. *Global collapse of frame structures under seismic excitations*, University of California at Berkeley: Berkeley, CA, 2005. pp.
Cubrinovski M and Ishihara K. State concept and modified elastoplasticity for sand modelling. *Soils and foundations* 1998; **38**(4): 213-225.
Yang Z, Elgamal AW and Parra E. Computational model for cyclic mobility and associated shear deformation. *Journal of Geotechnical and Geoenvironmental Engineering* 2003; **129**(12): 1119-1127.
Kulkarni RB, Youngs RR and Coppersmith KJ. *Assessment of confidence intervals for results of seismic hazard analysis*, in *8th World Conference on Earthquake Engineering,*: San Francisco, CA, 1984, 263–270pp.
Bommer JJ, Scherbaum F, Bungum H, Cotton F, Sabetta F and Abrahamson N. On the use of logic trees for ground-motion prediction equations in seismic hazard assessment. *Bulletin of the Seismological Society of America* 2005; **95**(2): 377–389.

Arulanandan K and Scott RF. *Verification of numerical procedures for the analysis of soil liquefaction problems*. A.A. Balkema: Rotterdam, 1993, pp.
Stepp JC, Wong I, Whitney J, Quittmeyer R, Abrahamson NA, Toro GR, Youngs RR, Coppersmith KJ, Savy J and Sullivan T. Probabilistic seismic hazard analyses for ground motions and fault displacement at Yucca Mountain, Nevada. *Earthquake Spectra* 2001; **17**(1): 113-151.
Bommer JJ and Scherbaum F. The use and misuse of logic trees in probabilistic seismic hazard analysis. *Earthquake Spectra* 2008; **24**(4): 997-1009.
Cubrinovski M, Uzuoka R, Sugita H, Tokimatsu K, Sato M, Ishihara K, Tsukamoto Y and Kamata T. Prediction of pile response to lateral spreading by 3-D soil-water coupled dynamic analysis: shaking in the direction of ground flow. *Soil Dynamics and Earthquake Engineering* 2008; **28**(6): 421-435.
Uzuoka R, Cubrinovski M, Sugita H, Sato M, Tokimatsu K, Sento N, Kazama M, Zhang F, Yashima A and Oka F. Prediction of pile response to lateral spreading by 3-D soil–water coupled dynamic analysis: shaking in the direction perpendicular to ground flow. *Soil Dynamics and Earthquake Engineering* 2008; **28**(6): 436-452.
Schorlemmer D and Gerstenberger MC. RELM Testing Center. *Seismological Research Letters* 2007; **78**(1): 30-36.
McGuire RK. Probabilistic seismic hazard analysis: early history. *Earthquake Engineering and Structural Dynamics* 2008; **37**329-338.
Cornell CA and Krawinkler H. Progress and challenges in seismic performance assessment. *PEER Center News* 2000; **3**(2): 1-3.
Shome N, Cornell CA, Bazzurro P and Carballo JE. Earthquakes, records, and nonlinear responses. *Earthquake Spectra* 1998; **14**(3): 469-500.
Bradley BA, Cubrinovski M, Dhakal RP and MacRae GA. Intensity measures for the seismic response of pile foundations. *Soil Dynamics and Earthquake Engineering* 2009; **29**(6): 1046-1058.
Bradley BA. A generalised conditional intensity measure approach and holistic ground motion selection. *Earthquake Engineering and Structural Dynamics* 2009; (published online).

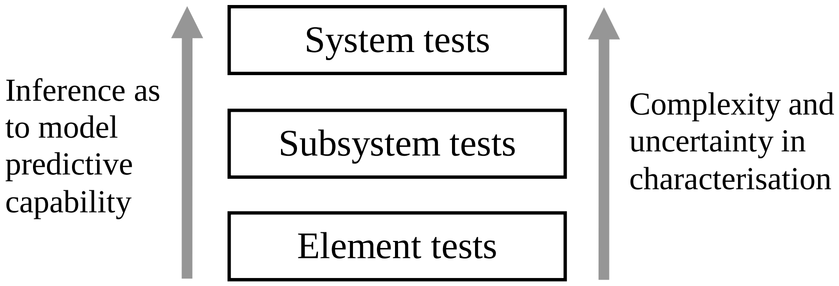

Figure 1: The primary test classes used in assessing computation seismic response model validity in earthquake engineering.

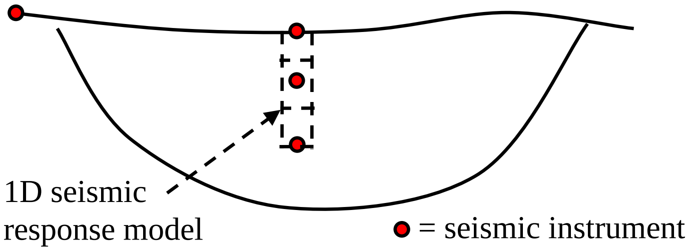

Figure 2: Illustration of a site response example in which seismometer arrays can be used to provide validation of seismic site response analyses.

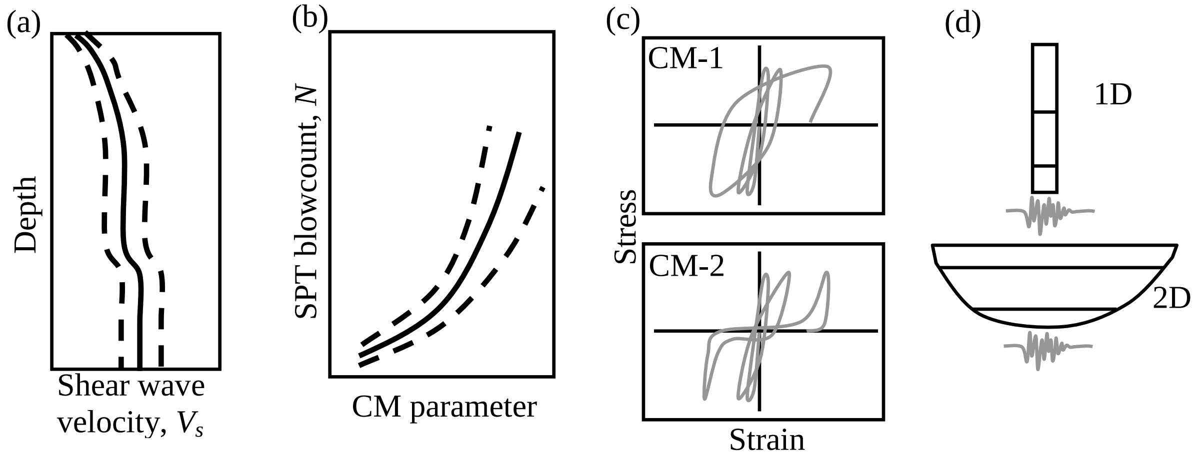

Figure 3: Examples of the four different types of uncertainties in the case of a geotechnical seismic site response analysis: (a) site characterisation uncertainty; (b) constitutive model (CM) parameter uncertainty; (c) constitutive model uncertainty; and (d) computational model methodology uncertainty.

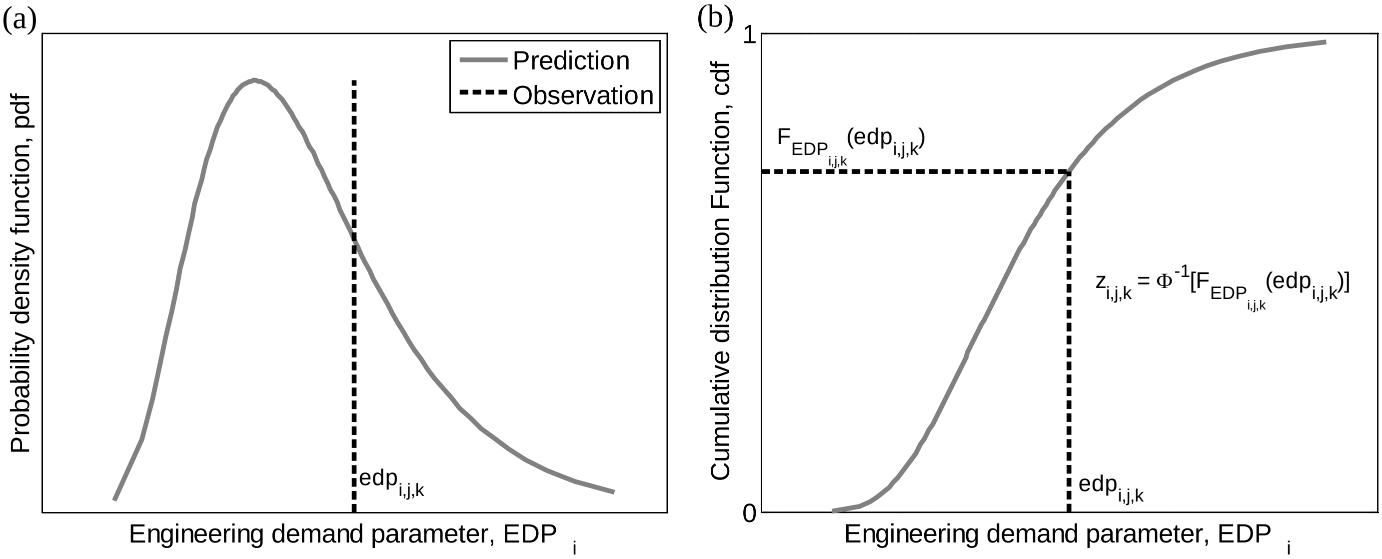

Figure 4: (a) Schematic comparison between prediction probability density function and observation; (b) computation of normalised residual based on cumulative prediction distribution and observation.

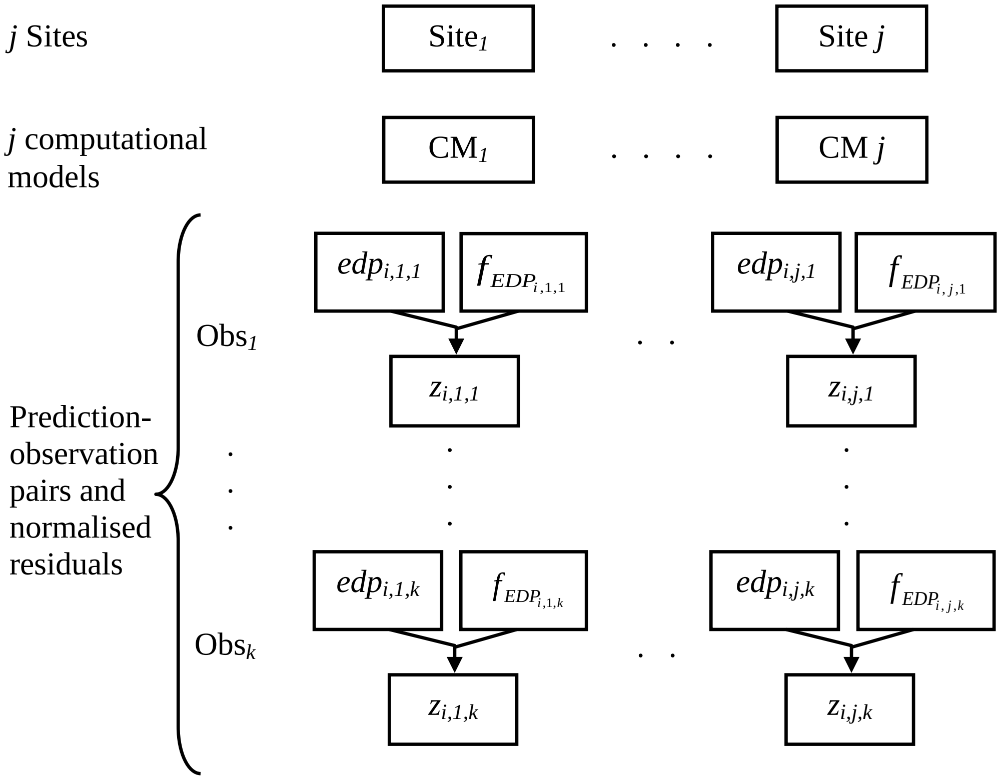

Figure 5: Illustration of the use of multiple observations and multiple sites to obtain prediction-observation pairs for use in validation of seismic response models. Site$_j$ = the jth site; CM$_j$ = the computational model of site j; $edp_{i,j,k}$ = the k^{th} observation of EDP$_i$ at site j; $z_{i,j,k}$ = the (total) normalised residual for the k^{th} observation of EDP_i at site j. $f_{EDP_{i,j,k}}$ = probability density function of the prediction of EDP_i for the k^{th} observation at site j.

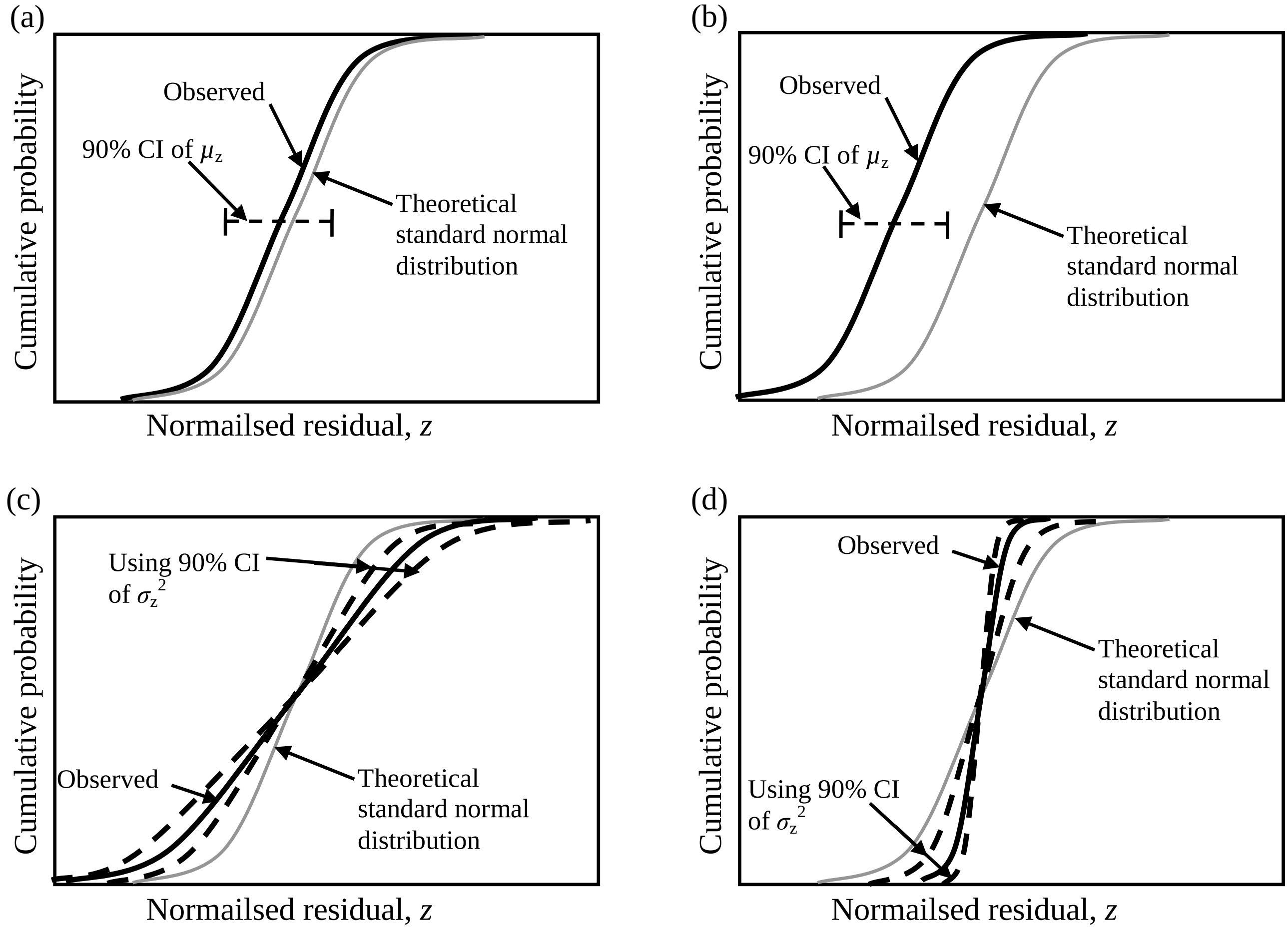

Figure 6: Illustration of the resulting distribution of the normalised residuals compared to the theoretical standard normal distribution (and statistical error bounds) in the cases of: (a) insignificant bias; (b) significant bias; (c) imprecision; and (d) 'over-precision'.

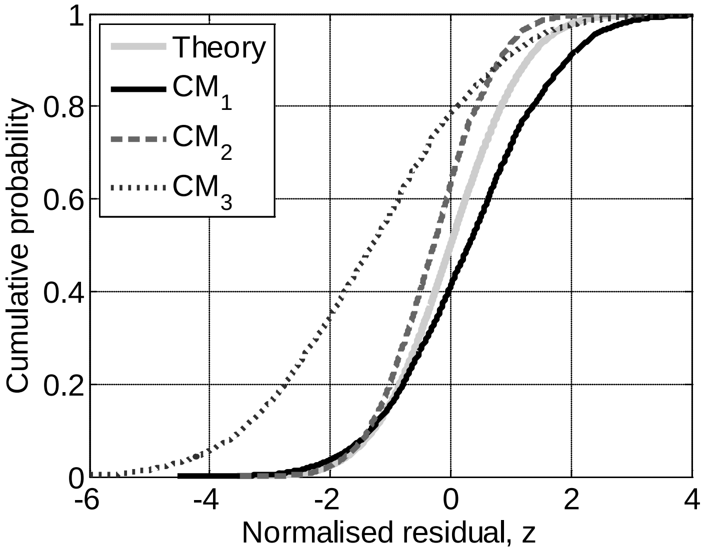

Figure 7: Use of the validation framework to assess the capability of various constitutive models (CM) based on the distribution of the normalised residuals.

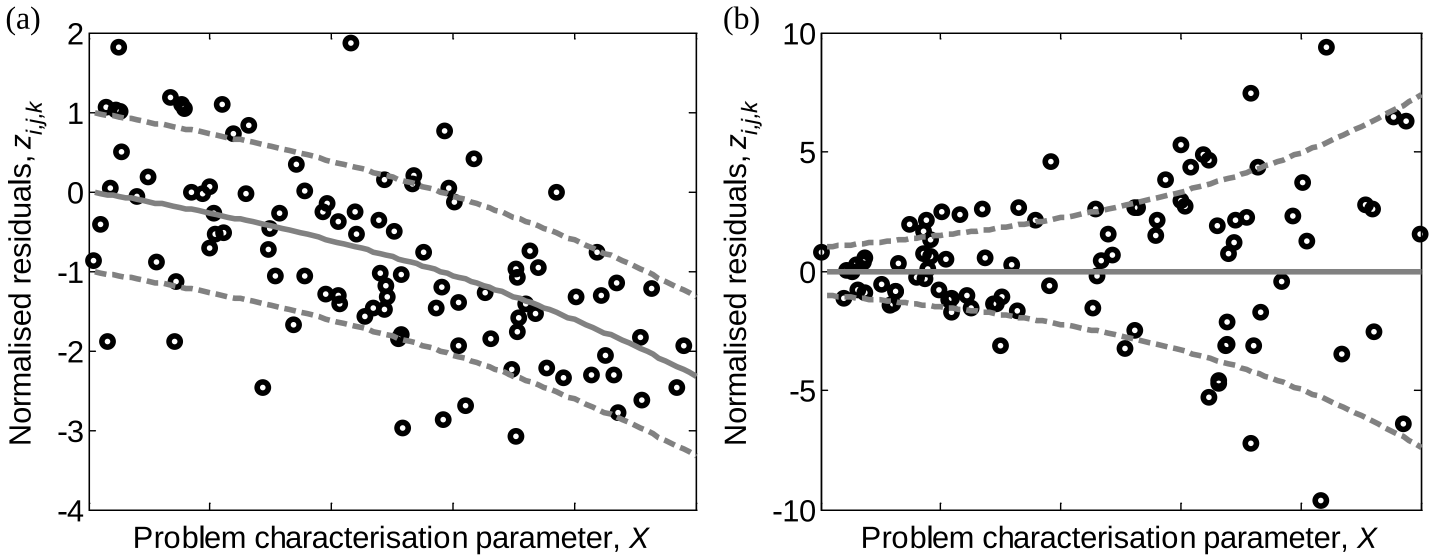

Figure 8: Possible observed trends in the normalised residuals with a particular parameter characterising the problem: (a) observed model bias with increasing *X*; and (b) increasing model imprecision with increasing *X*.