A New Procedure for Evaluating Ground Motion Models, with Application to Hydraulic-Fracture-Induced Seismicity in the UK

(Revised submission to BSSA)

Gemma Cremen, Maximilian J. Werner, and Brian Baptie

December 2019

Abstract

An essential component of seismic hazard analysis is the prediction of ground shaking (and its uncertainty), using ground motion prediction equations (GMPEs). This paper proposes a new method to evaluate (i.e. rank) the suitability of GMPEs for modelling ground motions in a given region. The method leverages a statistical tool from sensitivity analysis to quantitatively compare predictions of a GMPE with underlying observations. We demonstrate the performance of the proposed method relative to several other popular GMPE ranking procedures and highlight its advantages, which include its intuitive scoring system and its ability to account for the hierarchical structure of GMPEs. We use the proposed method to evaluate the applicability of several GMPEs for modelling ground motions induced by the commencement of UK shale gas exploration, with 195 recordings at distances (R) less than 10 km for 29 events with local magnitude (M_L) greater than 0 that relate to 2018/2019 hydraulic fracture operations at the Preston New Road shale gas site in Lancashire and 192 $R < 10$ km recordings for 48 $M_L > 0$ events induced - within the same geologic formation - by coal mining near New Ollerton, North Nottinghamshire. We examine: (1) the Akkar et al. (2014a) equations for European seismicity, (2) the Douglas et al. (2013) equation for geothermal induced seismicity, and (3) the Atkinson (2015) equation for eastern North America induced seismicity. We find the Douglas et al. (2013) equation to be the most suitable for almost all of the considered ground motion

intensity measures. We modify this equation by re-computing its coefficients in line with the observed data, to further improve its accuracy for future analyses of the seismic hazard of interest. This study both advances the state of the art in ground motion model evaluation and enhances understanding of the seismic hazard related to UK shale gas exploration.

1 Introduction

Ground motion prediction equations (GMPEs) are an essential component of seismic hazard analysis, used to predict ground shaking at a given distance from a particular magnitude event. It is therefore important that the GMPEs selected for inclusion in a given seismic hazard assessment are suitable for modelling the ground motions in the region of interest. A variety of methods have been proposed in the literature for evaluating (or ranking) GMPE suitability (e.g. Stewart et al., 2015). These include: (1) the analysis of residuals (i.e. differences between observations and corresponding predictions of the GMPE), which involves examining variations of the residuals with magnitude, distance, and site conditions (Scasserra et al., 2009), (2) the use of a likelihood-based score (Scherbaum et al., 2004; Stafford et al., 2008), which involves assessing the goodness-of-fit of the observations and the GMPE based on a likelihood parameter, and (3) the use of information theory (Scherbaum et al., 2009; Mak et al., 2017), which involves calculating log-likelihoods of observations for the GMPE. Interested readers are referred to Table 1 of Mak et al. (2017) for an excellent summary of the various methods that have been used in an extensive number of previous GMPE evaluation studies.

This paper proposes a new procedure for evaluating GMPEs. The method introduces a statistical tool from sensitivity analysis to quantify (score) the comparison between the cumulative distribution function (CDF) of residuals from a GMPE and the CDF expected if the equation correctly models the underlying observations. The proposed procedure offers a number of advantages over current evaluation methods (discussed in detail in a later section of the paper). For example, it correctly accounts for the hierarchical structure of GMPEs, i.e. the fact that they include correlation among ground motions from the same earthquake. It uses an intuitive scoring system, in which the optimum value is consistent; it does not depend on either the GMPE under evaluation or the observed data of interest. It also involves the calculation of residuals, which can act as a powerful visual tool to provide additional insight on how GMPEs compare with observations.

We use the proposed GMPE evaluation procedure to help improve understanding of the seismic hazard associated with shale gas exploration in the UK, where such industrial activity is relatively new; the first well to specifically test for UK shale gas was drilled in 2010 (Selley, 2012), and the first recorded instance of seismicity induced by hydraulic fracturing in the UK occurred in 2011 (Clarke et al., 2014). We specifically focus on the Preston New Road (PNR) shale gas site near Blackpool in Lancashire (Clarke et al., 2019), where the British Geological Survey (BGS) surface array detected 57 seismic events in 2018 and 121 seismic events in 2019 (up to 27th August), related to hydraulic fracture operations. While the magnitudes of the PNR events are significantly lower than those considered in conventional seismic hazard analyses, it is still useful to assess whether the associated shaking has the potential to be felt.

We test a number of pre-existing GMPEs for suitability to modelling the ground motions induced by UK shale gas exploration: (1) the Akkar et al. (2014a) equations, developed for European seismicity, (2) the Douglas et al. (2013) equation, developed for induced seismicity in geothermal areas, and (3) the Atkinson (2015) equation, developed for induced seismicity in eastern North America. Evaluation of the GMPEs is specifically carried out for peak ground velocity (PGV), peak ground acceleration (PGA), and 5% damped spectral accelerations at periods of 0.05s, 0.1s, and 0.2s ($SA_{0.05}$, $SA_{0.1}$, and $SA_{0.2}$ respectively). We then adjust the coefficients of the most suitable GMPE, to create a model specific to the seismicity of interest so that it can be used for future related hazard analyses (see **Developing a Modified GMPE** for details).

This paper is structured as follows. In **Proposed GMPE Evaluation Procedure**, we introduce the proposed GMPE evaluation procedure, demonstrate its performance relative to other evaluation methods, and describe its advantages as well as its limitations. In **Evaluating GMPEs for Modelling UK Shale Gas Seismicity**, we use the proposed procedure to evaluate the suitability of the aforementioned GMPEs for modelling ground shaking related to UK shale gas exploration. In **Developing a Modified GMPE**, we modify the most applicable GMPE to better suit the UK data, and compare the adjusted model to the previously examined GMPEs.

2 Proposed GMPE Evaluation Procedure

Ground motion prediction equations typically take the following mathematical form:

$$\log(im_{obs,i,j}) = \log(im_{GMPE,i,j}) + z_{E,i}\sigma_E + z_{A,i,j}\sigma_A \tag{1}$$

where - for the jth recording of the ith event - $\log(im_{obs,i,j})$ is the logarithm of the observed ground motion measure, $\log(im_{GMPE,i,j})$ is the logarithm of the median estimate of the ground motion measure given certain predictor variables (e.g. magnitude and distance) and model parameters, $z_{E,i}$ is the normalised inter-event residual (common to all recordings of the ith event), $z_{A,i,j}$ is the normalised intra-event residual, and σ_E and σ_A are the inter-event and intra-event standard deviations, respectively. $z_{E,i}$ can be estimated using (Abrahamson and Youngs, 1992):

$$z_{E,i} = \frac{\sigma_E \times \sum_{j=1}^{n_i}\left[\log(im_{obs,i,j}) - \log(im_{GMPE,i,j})\right]}{n_i\sigma_E^2 + \sigma_A^2} \tag{2}$$

where n_i is the number of recordings for the ith event. $z_{A,i,j}$ can then be calculated from:

$$z_{A,i,j} = \frac{\log(im_{obs,i,j}) - \log(im_{GMPE,i,j}) - z_{E,i}\sigma_E}{\sigma_A} \tag{3}$$

The format of equation 1 implies that both $z_{E,i}$ and $z_{A,i,j}$ should follow a standard normal distribution (i.e. mean=0, standard deviation =1) if the GMPE correctly models the observed data; this forms the basis of our proposed evaluation procedure. We use the Euclidean metric distance (EMD) between the cumulative distribution function (CDF) of the standard normal distribution and that of the maximum likelihood normal distribution for each type of normalised residual, to score models. This metric has previously been used to quantify uncertainty importance for sensitivity analyses (Chun et al., 2000). It may be calculated as follows:

$$EMD_x = \sqrt{(\mu_x - \mu_o)^2 + (\sigma_x - \sigma_o)^2} = \sqrt{\mu_x^2 + (\sigma_x - 1)^2} \tag{4}$$

where x refers to the normalised inter- or intra- event residuals, μ_x and σ_x are the maximum likelihood estimates of the mean and standard deviation, respectively, for the normalised residuals, and μ_o and σ_o are respectively the mean and standard deviation of the standard normal distribution. Note that equation 4 assumes the distribution of normalised residuals to be symmetric; the generic equation for the distance between any two CDFs is presented in equation 2 of Chun et al. (2000). The assumption of symmetric data is reasonable, since it is also the fundamental underlying assumption of a GMPE (from equation 1) and it is always valid (for sufficient sample sizes), based on the Central Limit Theorem (Kwak and Kim, 2017).

A graphical representation of EMD_x is provided in Figure 1. The final score for the proposed evaluation procedure (EMD_{total}) is a combination of the inter- and intra- event Euclidean metric distances, as follows:

$$EMD_{total} = \sqrt{EMD_{inter}^2 + EMD_{intra}^2} \tag{5}$$

The smaller the score, the closer the residuals are to the ideal distribution and the better the model. The format of equation 5 assumes that the errors from both types of residual are additive, independent, and equally important, which is directly consistent with the treatment of inter- and intra-event variabilities within GMPEs (e.g. Ornthammarath et al., 2011).

The proposed EMD scoring method is not to be confused with the Euclidean distance-based ranking (EDR) procedure proposed by Kale and Akkar (2013) , which is fundamentally different in its methodology. The EDR approach measures the Euclidean distance directly between an observed ground motion amplitude and the corresponding probability distribution of predictions from a GMPE, whereas the EMD method first calculates normalised residuals based on the median prediction of a GMPE; then measures the Euclidean distance between the probability distribution of residuals and the standard distribution expected for a perfect prediction by the model (which is independent of the GMPE in question). The proposed EMD score has a number of advantages over the EDR score: (1) the EMD score is proper, (2) residuals are a natural by-product of calcuating the EMD score, which can provide additional useful insight on the performance of a GMPE, and (3) the EMD score accounts for model hierarchy in GMPEs by considering inter- and intra-event variability separately. Further discussion on these advantages is left to

Advantages and Limitations of the Proposed Procedure, where the benefits of the EMD approach over other popular GMPE ranking methods are also explained.

2.1 Extension to Non-Constant Inter- and Intra-Event Standard Deviations

It is important to note that equations 2 and 3 are only valid if the inter- and intra-event standard deviations of a GMPE are constant (homoskedastic) across all values of the predictor variables (Stafford, 2015), which is not always the case (e.g. Ambraseys et al., 2005; Akkar and Bommer, 2007; Chiou and Youngs, 2014). The normalised inter-event residual vector for scenario-dependent inter- and intra-event standard deviations may be formulated as (Laird, 2004):

$$\mathbf{z_{E,i}} = \mathbf{D^{0.5}Z_i'\Sigma_i^{-1}}[\log(\mathbf{im_{obs,i}}) - \log(\mathbf{im_{GMPE,i}})] \tag{6}$$

where, for the *ith* earthquake, $\mathbf{D}$ is the inter-event covariance matrix and $\mathbf{Z_i}$ describes the linear relation of random effects (note that $\mathbf{Z_i'}$ denotes the transpose of $\mathbf{Z_i}$). $\log(\mathbf{im_{obs,i}})$ and $\log(\mathbf{im_{GMPE,i}})$ are vectors (in logarithmic scale) of the observed and median estimates of the ground motion measures, respectively. $\mathbf{\Sigma_i}$ can be calculated as follows:

$$\mathbf{\Sigma_i} = \mathbf{R_i} + \mathbf{Z_i D Z_i'} \tag{7}$$

where $\mathbf{R_i}$ is the intra-event covariance matrix for the ith earthquake, and both $\mathbf{Z_i}$ and $\mathbf{D}$ are as described for equation 6. The normalised intra-event residual vector for scenario-dependent inter- and intra-event standard deviations is then calculated as:

$$\mathbf{z_{A,i}} = \mathbf{R_i^{-0.5}}[\log(\mathbf{im_{obs,i}}) - \log(\mathbf{im_{GMPE,i}}) - \mathbf{D^{0.5}z_{E,i}}] \tag{8}$$

where all other variables are as defined previously. $\mathbf{z_{E,i}}$ and $\mathbf{z_{A,i}}$ should follow standard multivariate normal distributions if the GMPE is a correct model for the observations. Extension of the EMD_{total} metric to quantify the distance between the maximum likelihood multivariate normal distribution of the residuals and the standard multivariate normal distribution could be achieved using tools from optimal transport theory (e.g. Villani, 2008). However, since all the GMPEs to be evaluated in this study have homoskedastic standard deviations, further discussion on the ranking of GMPEs with scenario-dependent inter- and intra-event variabilities is left for future work.

2.2 Advantages and Limitations of the Proposed Procedure

To demonstrate the relative performance of the proposed procedure, we use the synthetic datasets of Mak et al. (2017), i.e. we assume there are four earthquakes with event terms η_i, $i \in \{1, 2, 3, 4\}$ that uniformly sample the distribution $\mathcal{N}(0, \sigma_b)$ and the N_i records for each earthquake uniformly sample the distribution $\mathcal{N}(\eta_i, \sigma_w)$. Thus, the jth residual for the ith event is calculated as:

$$r_{i,j} = Q_w(y_j) + \eta_i \tag{9}$$

where $y_j = \frac{2j-1}{2N_i}$, $j \in \{1, 2, ..., N_i\}$, $\eta_i = Q_b(x_i)$ with $x_i = \frac{2i-1}{8}$, and $Q_c(.)$ is the quantile function for $\mathcal{N}(0, \sigma_c)$. We also make use of Examples 1-3 of Mak et al. (2017), which compare the performance of different scores in various scenarios. Example 1 examines the variability of scores for perfect models across similar cases: (1) $N_i = \{20, 5, 5, 20\}$ and (2) $N_i = \{5, 20, 20, 5\}$ with $\sigma_b = 0.35$ and $\sigma_w = 0.5$. Example 2 examines the performance of scores using $N_i = \{10, 10, 10, 50\}$, $\sigma_b = 0.35$, $\sigma_w = 0.5$, and a biased model. Example 3 examines the ability of scores to distinguish between the correct and incorrect model partitioning of total uncertainty into inter-event and intra-event uncertainties for Case 1 of Example 1.

While these examples are useful for highlighting the pros and cons of the proposed procedure, it is important to emphasise their hypothetical (unrealistic) nature. Imbalances in the number of recordings per earthquake are over-exaggerated, as mentioned in Mak et al. (2017). In addition, sample sizes are exceptionally small (for instance, we note that Bommer et al. (2010) recommends at least 10 earthquakes per unit of magnitude and at least 100 records per 100 km to adequately constrain a GMPE) and an accurate evaluation of GMPEs may not be the only challenge to overcome when analysing such a limited amount of data. Actual datasets of this scale have led to difficulties in successfully calculating inter- and intra-event residuals (Bourne et al., 2015), for instance.

2.2.1 Advantages

The proposed evaluation procedure provides numerous benefits over similar methods previously proposed in the literature. The proposed score has three main advantages over both the $\overline{LLH}$ score proposed by

Scherbaum et al. (2009) and the the EDR score proposed by Kale and Akkar (2013). (1) The proposed score is proper (Lindley, 1991), since the best (lowest) score is achieved when the GMPE perfectly fits the observed data, i.e. when the CDFs of the residuals exactly match that of the standard normal distribution. On the other hand, the $\overline{LLH}$ score can favour a biased model if the number of recordings is unbalanced between earthquakes (Mak et al., 2017) and the EDR score favours a smaller predicted uncertainty value, regardless of what the true uncertainty is, when the predicted mean is close to correct (Mak et al., 2014). We demonstrate this benefit of the proposed score using the data of Example 2 in Mak et al. (2017). Unlike the $\overline{LLH}$ score, the proposed procedure correctly assigns a better score to the unbiased model (EMD_{total} for the correct model is 0.25 and EMD_{total} for the biased model is 0.48). If we halve both the inter-event (σ_b) and intra-event (σ_w) standard deviations of the correct model while keeping the observations unchanged, the proposed score disimproves from 0.25 to 1.04, whereas the EDR score incorrectly improves from 0.72 to 0.62. (2) Residuals are also calculated as part of obtaining the EDR score, which can provide additional insight on whether a GMPE is high or low relative to the observed data of interest (e.g. Bradley, 2013). (3) Through the separate consideration of intra- and inter-event residuals, the proposed procedure correctly accounts for the hierarchical nature of ground motion prediction equations, whereas the $\overline{LLH}$ score and the EDR score do not distinguish between these two types of variability.

The EMD approach has a number of advantages over the $\ell\ell(\mathbf{p}, \mathbf{V}, \mathbf{q})$ score proposed by Mak et al. (2017), which (to the best of our knowledge) is the only other score that incorporates model hierarchy. Firstly, the proposed score is more intuitive than the $\ell\ell(\mathbf{p}, \mathbf{V}, \mathbf{q})$ score, since its best possible value is always 0 whereas the $\ell\ell(\mathbf{p}, \mathbf{V}, \mathbf{q})$ score has a variable optimum value that depends on the length of the given dataset (via the '$N \log(2\pi)$' term) and the variance of the model to be evaluated (via the '$\log |\mathbf{V}| + (\mathbf{q} - \mathbf{p})'\mathbf{V}^{-1}(\mathbf{q} - \mathbf{p})$' terms). The variability of the optimum $\ell\ell(\mathbf{p}, \mathbf{V}, \mathbf{q})$ score is highlighted in Example 2 and Example 3 of Mak et al. (2017); the score for the correct model in Example 2 (where there are 4 earthquakes and 80 records) is 61.2, whereas the score for the correct model in Example 3 (where there are 4 earthquakes and 50 records) is 38.8. (The EMD_{total} score for both models is 0.25). Secondly, the proposed procedure is significantly less computationally expensive than that proposed by Mak et al. (2017) (at least when evaluating GMPEs with homoskedastic standard deviations); it only requires space for $\sum_{i=1}^{n_e} n_{r_i} + n_e$ residuals (where n_e is the number of earthquakes and n_{r_i} is the number of records for the ith earthquake) plus the maximum likelihood estimates of the residual distributions, whereas the procedure of Mak et al. (2017) necessitates the storage of $\sum_{i=1}^{n_e} n_{r_i}^2$ non-zero elements for the $\mathbf{V}$ matrix alone. To demonstrate the practical significance of the difference in computational requirements between the two methods, we use a hypothetical dataset with 100 earthquakes and 100 records per earthquake - which roughly corresponds in record number to half the size of the NGA West-2 ground motion database (Ancheta et al., 2014) - and assume that we are evaluating a GMPE with homoskedastic standard deviation. For double precision in the MATLAB environment, the procedure of Mak et al. (2017) will require 800 MB of storage for the $\mathbf{V}$ matrix, whereas the necessary data for the proposed procedure can be stored in a vector of less than 0.1 MB in size. The computational advantage of the proposed procedure will become even more apparent as future evaluations of models involve increasing amounts of recorded data.

The proposed score also has an advantage over goodness-of-fit measures proposed for evaluating GMPEs - such as the Kolmogorov-Smirnov test and the mean test p-value (Scherbaum et al., 2004) - since it does not include the use of classical statistical hypothesis testing, which can be limited in ability to measure the importance of a result (Wasserstein and Lazar, 2016).

2.2.2 Sample Size Constraints

To assess the reliability of the proposed procedure for modest sample sizes, we compute scores for the small datasets examined in Examples 1-3 of Mak et al. (2017), which contain 50-80 records across 4 earthquakes. It can be observed from Table 1 that the proposed procedure correctly scores the models in Example 2 (as discussed in **Advantages**) but it does not perform as expected for Examples 1 and 3; the scores are not equivalent for both (correct) cases in Example 1, and the model with the smallest σ_b is incorrectly deemed to be the best in Example 3. The incorrect scoring in Examples 1 and 3 is due to the inaccurate estimation of inter-event residuals by equation 2, which only represents the best predictor of random effects given the set of available observations (Jiang, 2007). For instance in Example 1, inter-event residuals for Case 1 are estimated to be $\{-1.04, -0.23, 0.23, 1.04\}$ and for Case 2 are estimated to be $\{-0.82, -0.29, 0.29, 0.82\}$, whereas the true inter-event residuals for both cases (simulated according to equation 9) are $\{-1.15, -0.32, 0.32, 1.15\}$.

The incorrect scoring by the proposed procedure can also be minorly attributed to the use of maximum likelihood estimation for obtaining the means and standard deviations of the normalised residuals, which is well known to have reduced accuracy for small sample sizes (e.g. Lee and Song, 2004). We note that many other popular GMPE evaluation scores - such as that of Mak et al. (2017) as well as the Scherbaum et al. (2009) $\overline{LLH}$ score - involve maximum likelihood estimates and are therefore also somewhat impacted by small sample sizes (Beauval et al., 2012).

To understand the sample sizes necessary for the proposed evaluation procedure to perform correctly in Examples 1 and 3, we calculate the scores for datasets with an increasing number of events and recordings (Figure 2). Increasing 'Earthquake Number' involves adding earthquakes to the centre (Case 1 in Example 1 and Example 3) or outside (Case 2 in Example 1) of a dataset, with the same number of records as the nearest events in the set. Increasing 'Record Number Scaling' involves multiplying the number of records per earthquake by a factor. For example, an 'Earthquake Number' of 10 and a 'Record Number Scaling' of 2 for Case 1 in Example 1 yields the dataset $N_i = \{40, 10, 10, 10, 10, 10, 10, 10, 10, 40\}$, and for Case 2 in Example 1 yields the dataset $N_i = \{10, 10, 10, 10, 40, 40, 10, 10, 10, 10\}$. Residuals are still calculated according to equation 9, with the denominator of x_i replaced by $2 \times$ Earthquake Number.

Figure 2a plots the absolute difference between the EMD_{total} values for Case 1 and Case 2 in Example 1, and Figure 2b plots the difference between the EMD_{total} values for the correct model and the model with deflated σ_b. It can be seen in Figure 2a that the absolute difference in EMD_{total} values for both cases in Example 1 will reduce to 0.01 if the number of records for each earthquake is scaled by 22 (1100 total records per case), or if the number of earthquakes is increased to 30 and the number of records per earthquake is scaled by 9 (1620 total records per case), for example. The proposed procedure will score the correct model better than the model with smallest σ_b in Example 3 if the number of earthquakes is increased to 10 and the number of records per earthquake is scaled by 4 (320 total records), or if the number of earthquakes is increased to 30 and the number of records for each earthquake is scaled by 3 (540 total records), for example (Figure 2b). It can be concluded that the number of earthquakes and recordings necessary for the proposed evaluation procedure to perform reliably for Examples 1 and 3 is notably larger than that examined by Mak et al. (2017), however we again emphasise that these examples are far from those expected in real-life applications.

3 Evaluating GMPEs for Modelling UK Shale Gas Seismicity

We use the proposed GMPE evaluation procedure to help improve understanding of the seismic hazard related to shale gas exploration in the UK, where such industrial activity is relatively new. We focus on 2018 and 2019 seismic events associated with the PNR shale gas site near Blackpool in Lancashire (Figure 4a), which are the only well-recorded series of shale gas-related events that have occurred in the UK. We also use a high quality dataset of ground motion recordings from events that were induced by coal mining near New Ollerton (NO) in North Nottinghamshire (Figure 4b; Verdon et al., 2017), as these earthquakes had very similar magnitudes and depths to those of the PNR sequence (Figures 4c and 4d), they occurred in the same geological formation (Butcher et al., 2017), and were found to have comparable ground motion amplitudes to those of the 2018 PNR events for most of the intensity measures of interest (Cremen et al., 2019).

3.1 GMPEs Examined

We evaluate the suitability of various GMPEs for modelling the ground motions of interest: (1) Akkar et al. (2014a, hereafter ASB14), (2) Douglas et al. (2013, hereafter D13) and (3) Atkinson (2015, hereafter A15). ASB14 was chosen because they were used for planning purposes in preliminary shale gas-related PNR hazard calculations (Arup, 2014). D13 and A15 were chosen for their application to induced seismicity.

ASB14 are a series of GMPEs developed for European and Middle East crustal seismicity that were derived using a subset of the Reference Database for Seismic Ground-Motion in Europe (RESORCE) (Akkar et al., 2014b) . They are applicable for moment magnitudes (M_w) greater than 4 and distances less than 200 km. The equations use either point-source (i.e. epicentral and hypocentral distance) or finite-fault (surface projection of rupture distance) distance metrics. Events are sufficiently small such that rupture distance is not important in this study, so we only use the point-source equations (henceforth referred to as $\text{ASB14}_{\text{hypo}}$ and $\text{ASB14}_{\text{epi}}$).

D13 are a series of GMPEs developed for geothermal induced seismicity that were derived using data from induced and natural seismicity in Basel (Switzerland), Campi Flegrei (Italy), Geysers (United States), Hengill (Iceland), Roswinkel and Vorendaal (the Netherlands), and Soultz-sous-Forets (France). They are applicable for M_w greater than 1 and distances less than 50 km. All equations except one are site corrected to a reference rock condition (V_{s30} = 1100 m/s). This condition is significantly different to that observed at sites in this study (V_{s30} = 280 m/s, as explained in **Data Used**), so we only use the equation that represents an unknown site condition in this case. (Note that we could make our data compatible with the site corrected condition by obtaining site-specific estimates of amplification and attenuation, but this is outside the scope of the current study.)

A15 is a GMPE developed for induced seismicity in eastern North America that was derived using a subset of the Next Generation Attenuation-West 2 (NGA-West 2) database (Ancheta et al., 2014). It is applicable for magnitudes between 3 and 6 and distances less than approximately 50 km. The equation is site corrected to a reference soft rock condition (V_{s30} = 760 m/s). However, it can be conveniently adjusted to another site condition by inputting the appropriate V_{s30} value to the empirical site correction model of Seyhan and Stewart (2014), which was calibrated using the same database. We use this model to site correct our data.

3.2 Data Used

We only examine data recorded at distances less than 10 km from events with local magnitude (M_L) > 0 in this study, since smaller magnitude events and farther locations (for the magnitude range considered in this study) will have extremely low levels of shaking that will not be felt. 29 Preston New Road (PNR) events fit the magnitude criterion, for which there are 76 recordings available within 10 km from nine Guralp 3-ESP broadband seismometers deployed by the BGS near the site. A further 119 recordings are available for 2018 events from eight seismic instruments (two Kinematics Shallow Borehole Episensor 2 broadband accelerometers and six Geospace Technologies SNG 3C GS-ONE LF geophones) used for monitoring by the shale gas exploration operator at the site, Cuadrilla Resources Ltd. We retrieve the event phase data and the raw waveforms of the BGS instruments from the BGS seismic database, and the raw waveforms of the operator's instruments from the UK Oil and Gas Authority (see Data and Resources). We consider 48 New

Ollerton (NO) earthquakes greater than 0 M_L, for which there are 192 recordings available within 10 km from four Guralp 3-ESP broadband seismometers installed by the BGS. Waveforms and phase data for the earthquakes are accessed using the BGS seismic database. A histogram of the complete database is provided in Figure 3.

We convert waveforms from dimensions of digital counts to velocity or acceleration using the procedure of Haney et al. (2012) (for broadband seismometers), assuming a causal third-order high-pass Butterworth filter with frequency 3 Hz, a causal fifth-order low-pass Butterworth filter with frequency 20 Hz, and an oversampling rate of 5. Accelerations are obtained from the derived velocities by numerical differentiation, and velocities are obtained from the derived accelerations using numerical integration. Spectral accelerations are computed using the algorithm provided in Wang (1996). Ground motion intensities are calculated across a time window from p-wave arrival to 5 seconds after the occurrence of the maximum displacement amplitude. Signal-to-noise ratios for each seismogram are taken as a ratio of the Fourier amplitude spectrum (FAS) evaluated during this time window to the FAS evaluated for a noise window of equivalent duration (Perron et al., 2018). We ignore data with signal-to-noise ratios less than or equal to 3, which removes 3 $SA_{0.05}$ values, 7 $SA_{0.1}$ values and 5 $SA_{0.2}$ values from the PNR dataset, and 1 $SA_{0.05}$ value from the NO dataset. The data considered for both earthquake sequences are summarised in Figure 4. It is important to note that the size of the dataset - 77 earthquakes with a median of 4 and a maximum of 12 data points per earthquake- is sufficient for the proposed evaluation procedure to perform correctly. We can confirm this by repeating Example 3 of Mak et al. (2017), using $N_i = \{x_i, x_{i+1}, ..., x_{n-1}, x_n\}$, where the length ($n$) of $N_i = 77$ and x_i is equal to the number of records available for the ith earthquake. To adequately capture the interaction between sample size and event term, the earthquakes are placed within N_i in ascending order of their inter-event residual with respect to the ASB14$_{\text{hypo}}$ GMPE. We find that the EMD_{total} scores accurately indicate the correct model; EMD_{total} for the correct model is 0.316, which is lower than the value for the model with inflated σ_b (0.332) and the value for the model with deflated σ_b (0.321).

The value of a ground motion intensity measure used for a particular event and distance combination depends on the requirements of the GMPE of interest. For ASB14$_{\text{hypo}}$, ASB14$_{\text{epi}}$, and D13, it is taken as the geometric mean of the values computed for the two horizontal components. For A15, it is taken as the median value for the two horizontal components computed over all nonredundant azimuths, as detailed in Boore (2010). M_L values are converted to M_w values using the empirical relationship derived by Butcher et al. (2019) for coal-mining induced seismicity in the UK:

$$M_w = 0.69M_L + 0.74 \tag{10}$$

All sites sit on alluvial soils so we use a V_{s30} value of 280 m/s, the median value found for these types of soil by Campbell et al. (2016), for site correction factors in ASB14$_{\mathrm{hypo}}$, ASB14$_{\mathrm{epi}}$, and A15. We assume a linear site response for A15. We assume strike-slip style-of-faulting for PNR data and reverse faulting for NO data in ASB14$_{\mathrm{hypo}}$ and ASB14$_{\mathrm{epi}}$, as these are the respective dominant regimes for each type of seismicity (Clarke et al., 2019; Verdon et al., 2017).

3.3 Evaluation Results

Table 2 provides EMD_{total} scores for each GMPE. Also provided for comparison are $\ell\ell(\mathbf{p,V,q})$ scores, calculated according to the evaluation procedure proposed by Mak et al. (2017). Figures 5 and 6 provide the corresponding inter- and intra-event residuals, as well as those expected for a standard normal distribution (i.e. a perfectly fitting GMPE). It can be seen from Table 2 that, according to the proposed evaluation procedure, D13 is the most suitable GMPE for modelling all ground motion intensities examined except $SA_{0.2}$, for which A15 is the most suitable. It is interesting to note that these findings are consistent with those of a similar evaluation study carried out by Cremen et al. (2019) for the same GMPEs, which included only 2018 PNR data from the BGS seismometers and used the GMPE ranking scheme of Scherbaum et al. (2004). Since both ASB14 equations and A15 were calibrated at much higher magnitudes than those examined here (see **GMPEs Examined**), these results provide further support for previous studies, (e.g. Bommer et al., 2007; Douglas and Jousset, 2011; Atkinson and Morrison, 2009) which found that GMPEs derived from larger-magnitude events should not be extrapolated to predict ground motions from earthquakes with smaller magnitudes.

The ranking of GMPEs according to the proposed procedure matches that of the Mak et al. (2017) procedure except in the case of PGV, for which the proposed procedure favours D13 and the procedure of

Mak et al. (2017) favours A15. Figure 5 highlights why the proposed procedure favours D13 for PGV; while the intra-event residuals for A15 compare better overall with the standard normal distribution than those of D13 due to the closer fit of their standard deviation (mean of D13 intra-event residuals = -0.03 and mean of A15 intra-event residuals = 0.31, while standard deviation of D13 intra-event residuals = 0.41 and standard deviation of A15 intra-event residuals = 1.02), the inter-event residuals for D13 perform significantly better relative to the standard normal distribution than those of A15 due to their notably lower bias (mean of D13 inter-event residuals = -0.06 and mean of A15 inter-event residuals = 1.10, while standard deviation of D13 inter-event residuals = 0.30 and standard deviation of A15 inter-event residuals = 0.50). The Mak et al. (2017) procedure's preference for A15 can be explained by A15's significantly smaller variance relative to that of D13 for PGV; A15 intra-event variability for PGV (in natural log units) is 0.645, which is over 60% less than the equivalent value of 1.811 for D13. Even though the error term (i.e. $[\mathbf{q}-\mathbf{p}]'\mathbf{V}^{-1}[\mathbf{q}-\mathbf{p}]$) of the $\ell\ell(\mathbf{p},\mathbf{V},\mathbf{q})$ score is much lower for D13 (73) than A15 (551), the difference in values of the variance term (i.e. $\log|\mathbf{V}|$) is sufficient to yield an overall lower $\ell\ell(\mathbf{p},\mathbf{V},\mathbf{q})$ score for A15 ($\log|\mathbf{V}|$ is 505 for D13 and is -276 for A15).

4 Developing a Modified GMPE

We now analyse the suitability of the most promising GMPE, D13, in greater detail. This equation has the following functional form:

$$\ln Y = a + bM + c \ln \sqrt{r_{hyp}^2 + h^2} + d r_{hyp} + \mathcal{N}(0, \phi) + \mathcal{N}(0, \tau) \tag{11}$$

where Y is the observed ground motion intensity measure of interest for moment magnitude M and hypocentral distance (in km) r_{hyp}, $\mathcal{N}(\mu, \Sigma)$ is a normal distribution with mean μ and standard deviation Σ, ϕ is the intra-event standard deviation, τ is the inter-event standard deviation, and $\sigma = \sqrt{\phi^2 + \tau^2}$ is the total standard deviation.

We examine trends in the residuals with the different predictor variables and update model coefficients to better suit the data as required, similar to the referenced empirical method for fitting GMPEs (Atkinson,

2008) and in line with the procedure detailed in Scasserra et al. (2009). We first investigate the variation of intra-event residuals ($\epsilon_{i,j}$) as a function of hypocentral distance (Figure 7). To highlight trends, we perform a linear regression according to:

$$\epsilon_{i,j} = z_{A,i,j}\sigma_A = a_R + b_R R_{i,j} + (\epsilon_R)_{i,j} \tag{12}$$

where $z_{A,i,j}$ and σ_A are as defined in equation 3, $R_{i,j}$ is hypocentral distance, a_R and b_R are regression parameters, and $(\epsilon_R)_{i,j}$ is the residual for the jth recording from the ith event. The p-values plotted in Figure 7 test the null hypothesis that the slope parameter b_R is equal to zero; since they all have extremely low values (i.e. ≤ 0.01), we can conclude that there is a statistically significant relationship between the residuals and hypocentral distance for each ground motion intensity measure examined. b_R is negative in each case, indicating that there is faster distance attenuation of the observed data relative to the D13 GMPE. To address the distance attenuation discrepancy, we recalculate coefficients related to near-source saturation (i.e. c and h) and the constant term of D13, using non-linear regression of the observed data. (We do not attempt to reevaluate the anelastic attenuation term of D13, given the short distances of interest). Note that the h coefficient is not found to be statistically significant in the initial regression analyses for any ground motion intensity measure examined, so the values for the other two terms are instead computed with h set to 0. We obtain the inter- and intra-event standard deviations of the distance-modified D13 by performing mixed effects regression on the total log residuals $\log(Z_{i,j})$ (e.g. Scasserra et al., 2009), calculated as follows:

$$\log(Z_{i,j}) = \log(im_{obs,i,j}) - \log(im_{GMPE',i,j}) \tag{13}$$

where $\log(im_{GMPE',i,j})$ is the logarithm of the median estimate of the ground motion measure for the model parameters of the distance-modified D13 and $\log(im_{obs,i,j})$ is as defined in equation 1. There is a statistically insignificant relationship between the normalised intra-event residuals of the distance-modified D13 and hypocentral distance (Figure 8), indicating that the updated GMPE is adequately capturing the distance attenuation of the observed data.

We assess the magnitude-scaling of the distance-modified D13 by investigating the variation of the inter-event residuals (η_i) as a function of moment magnitude (Figure 9). To illustrate trends, we conduct a linear regression according to:

$$\eta_i = z_{E,i}\sigma_E = a_M + b_M M_i + (\epsilon_M)_i \tag{14}$$

where $z_{E,i}$ and σ_E are as defined in equation 2, M_i is moment magnitude, a_M and b_M are regression parameters, and $(\epsilon_M)_i$ is the residual for the ith event. There is a statistically significant positive trend in the residuals with moment magnitude for each ground motion intensity measure of interest besides PGA (indicated by the small p-values for b_M plotted in Figure 9). This implies that the magnitude-scaling of the observed data is larger than that predicted by the GMPE in these cases, which makes sense given that D13 was calibrated for slightly higher magnitudes (e.g. Chiou et al., 2010). To rectify this, we use linear regression to recompute the magnitude-related coefficient and the constant term of the distance-modified D13 (except in the case of PGA). Mixed-effects regression is then used to calculate the updated inter- and intra-event standard deviations of the distance- and magnitude-modified D13. It is observed in Figure 10 that the distance- and magnitude-modified D13 correctly accounts for the magnitude-scaling of the observed data. Note that distance-dependent trends in the intra-event residuals of the distance- and magnitude-modified D13 are also found to be negligible.

Coefficients of the distance- and magnitude-modified D13 (henceforth referred to as CWB19) are provided in Table 3, for all ground motion intensity measures examined. Figure 11 provides regional median PGV predictions of the GMPE related to two hypothetical scenario earthquakes at the PNR shale gas site, which are equivalent in size to the two largest events that occurred during operations there in 2019. The applicability of CWB19 is limited to hypocentral distances between approximately 2 and 6 km, and (positive) moment magnitudes less than 3, given the sparsity of available calibration data for other values. CWB19 nevertheless represents a reasonable first attempt at modelling ground motions related to UK shale gas exploration, and will be refined in the future as further data are recorded.

4.1 Comparing CWB19 with existing GMPEs

We now examine the distance-scaling, the magnitude-scaling, and the standard deviations of CWB19, relative to those of the GMPEs previously assessed for suitability to modelling the ground motions of interest. A15 is site corrected to a V_{s30} value of 280 m/s in all distance- and magnitude-scaling comparisons. It should also be noted, as part of interpreting the comparisons, that ground motion amplitudes calculated according to A15 are not strictly equivalent to those calculated using the other GMPEs (see **Data Used** for more details).

Figure 12 compares the distance-scaling of the median predicted amplitudes of CWB19 with those of the previously examined GMPEs, for a fixed focal depth of 2 km and a moment magnitude of 1.5. The ground motion amplitudes predicted by the GMPEs derived from naturally occurring events (i.e. the ASB14 equations) are significantly larger than those predicted by the GMPEs designed for induced earthquakes (i.e. all other equations examined) across most distances and intensity measures of interest. This is not surprising, given that the ASB14 equations have undergone the largest extrapolation from their range of applicability (e.g. Baltay and Hanks, 2014). The very near-source predicted amplitudes of CWB19 are significantly larger than those of A15 and D13 (and even those of both ASB14 equations for PGV). The distance attenuation of CWB19 is faster than that of all other examined GMPEs, such that its predictions are similar to those of either A15 or D13 at the farthest distances considered. We can conclude that, for the ground motion intensity measures studied, close-distance intensities predicted by CWB19 are larger than those expected by the two GMPEs focused on induced events (as well as those expected by the GMPEs derived from naturally-occuring events for PGV), but its predicted intensities at farther distances are in line with expectations for induced earthquakes. This may be explained by the fact that the UK induced earthquakes examined occurred at shallower depths than those used to constrain D13 and A15; all PNR and NO earthquakes occurred at depths less than 3 km, while the mean focal depth of earthquakes used to fit D13 is approximately 5 km, based on visual inspection of Figure 1 in Douglas et al. (2013), and the mean focal depth of earthquakes used to fit A15 is 9 km (Atkinson, 2015).

Figure 13 compares the magnitude-scaling of the median predicted amplitudes of CWB19 with those of the previously assessed GMPEs, at a distance of 3 km (which is hypocentral or epicentral, depending on the functional form of the GMPE). Across all intensity measures examined except PGV, the ground motion amplitudes predicted by the natural GMPEs are notably larger than those predicted by the induced GMPEs for magnitudes less than approximately 2.5, but are similar at greater magnitudes. The magnitude-scaling of CWB19 is comparable to that of D13 for PGA, $SA_{0.05}$, and $SA_{0.1}$, and that of A15 for $SA_{0.2}$; the only notable difference is a marginally steeper scaling for CWB19 in the case of $SA_{0.05}$, $SA_{0.1}$, and $SA_{0.2}$, such that expected ground motion amplitudes are higher for CWB19 than for either D13 or A15 at the largest magnitudes considered. The magnitude-scaling of CWB19 for PGV is significantly different to that of the other GMPEs at very small magnitudes, but very similar to those of ASB14$_{\mathrm{epi}}$, A15, and D13 for magnitudes greater than 1.5. We conclude that the magnitude scaling of CWB19 is generally in line with that of other induced GMPEs, for the ground motion intensity measures examined.

Figure 14 shows intra- and inter-event standard deviation values (in natural log units) for CWB19 across all ground motion intensity measures of interest, compared with equivalent values for the other GMPEs examined. Inter-event values for CWB19 are consistently lower than those of A15 and D13, and are significantly less than those for all other GMPEs assessed in the case of PGV and $SA_{0.1}$. These findings are not surprising, given that CWB19 is derived using (essentially) only two sources, i.e. the shale gas site at PNR and the coal mine at NO. The intra-event variability values of the developed GMPE are generally slightly lower than those of the other GMPEs; this may be explained by the narrow near-source distance range of interest for CWB19. Note that the relatively small standard deviation values underline the fact that CWB19 should not be used outside the seismicity context for which it was created nor the magnitude and distance ranges outlined in **Developing a Modified GMPE**, as underestimating variability in ground motions can have a significant impact on the results of seismic hazard analyses (Bommer and Abrahamson, 2006).

4.1.1 Improvement in GMPE

We can use the EMD_{total} metric developed to quantify the improvement in modelling accuracy offered by CWB19 over D13 for the data of interest, given that the scale of the score is consistent across all GMPEs. The percentage improvement is calculated as follows:

$$\%\ \text{Improvement} = \frac{(EMD_{total})_{D13} - (EMD_{total})_{CWB19}}{(EMD_{total})_{D13}} \times 100 \tag{15}$$

where $(EMD_{total})_z$ is EMD_{total} for the GMPE z. Table 4 contains percentage improvement values for all ground motion intensity measures examined in this study. It can be seen that there is a notable improvement for all intensity measures, with an average improvement of 66%. Thus, adjusting the coefficients of D13 has significantly enhanced its suitability to modelling ground motions induced by UK shale gas exploration.

5 Conclusions

This paper has proposed a new method for evaluating the suitability of GMPEs to modelling the ground motions in a given region of interest. The method leverages a statistical tool from sensitivity analysis to quantitatively compare the distribution of residuals from a GMPE with the distribution expected for an exact fit of the equation to the underlying observations. The proposed method has a number of advantages over similar procedures in the literature. For example, it is based on an intuitive scoring system that yields consistent score values across all GMPEs and observed datasets. It does not rely on statistical hypothesis testing, from which it is difficult to measure the importance of a result. It also correctly accounts for the hierarchical structure of GMPEs. The accuracy of the proposed procedure can be hampered by very small sample sizes (i.e. on the order of 4 earthquakes), however such limited datasets are far from those expected to be used in real-life evaluations of GMPEs.

The proposed evaluation procedure was used to assess the suitability of a number of different GMPEs ($\mathrm{ASB14}_{\mathrm{hypo}}$, ASB_{epi}, A15, and D13) for modelling earthquakes induced by shale gas exploration in the UK. We specifically focused on events related to the PNR shale gas site near Blackpool in Lancashire, and supplemented the dataset with information on a sequence of similar events related to coal-mining that occurred within the same geologic formation at New Ollerton, North Nottinghamshire. We found that D13 was the most applicable GMPE of the four, at least for the considered ground motion intensity measures of PGV, PGA, $SA_{0.05}$, $SA_{0.1}$, and $SA_{0.2}$, and the dataset of observed recordings examined. We further enhanced the suitability of D13 for modelling ground motions associated with UK shale gas exploration, by adjusting its coefficients in line with the observed dataset; details of the modified equation (CWB19) are provided in **Developing a Modified GMPE**.

This paper provides a useful tool for ranking GMPEs that can be used to select suitable candidate models for input to probabilistic seismic hazard analyses (PSHA). Our assessment and development of GMPEs for modelling ground motions related to UK shale gas exploration enhances understanding of the strength of ground shaking associated with this type of seismicity, and the findings have many potential applications in further related work. For example, the developed GMPE could be used as part of future PSHA studies related to UK shale gas seismicity, for accurately modelling ground motion amplitudes at close distances and small magnitudes. These studies could ultimately inform engineering seismic risk calculations, which could be used to aid decision-making related to UK regulations on shale gas operations.

6 Data and Resources

Earthquake catalogs were obtained from the earthquake database of the British Geological Survey (`https://earthquakes.bgs.ac.uk/earthquakes/dataSearch.html`). Seismograms, phase measurements, and data used to correct for instrument response were acquired from the British Geological Survey's seismic database and the UK Oil and Gas Authority's database on 2018 PNR operations (`https://www.ogauthority.co.uk/onshore/onshore-reports-and-data/preston-new-road-pnr-1z-hydraulic-fracturing-operations-data/`). All other data used were retrieved from sources listed in the references.

7 Acknowledgements

We thank Dr. Julian J. Bommer and an anonymous reviewer for very helpful comments that significantly improved the quality of this manuscript. This work has been funded by the Natural Environment Research Council (NERC) Grant Number NE/R017956/1 "Evaluation, Quantification and Identification of Pathways and Targets for the assessment of Shale Gas RISK (EQUIPT4RISK)", the Bristol University Microseismic Projects (BUMPS), and the British Geological Survey.

References

Abrahamson, N. A. and R. Youngs (1992). A stable algorithm for regression analyses using the random effects model, Bull. Seismol. Soc. Am. **82**(1), 505–510.

Akkar, S. and J. J. Bommer (2007). Empirical prediction equations for peak ground velocity derived from strong-motion records from Europe and the Middle East, Bull. Seismol. Soc. Am. **97**(2), 511–530.

Akkar, S., M. Sandıkkaya, and J. Bommer (2014a). Empirical ground-motion models for point-and extended-source crustal earthquake scenarios in Europe and the Middle East, Bull. Earthquake Eng. **12**(1), 359–387.

Akkar, S., M. Sandıkkaya, M. Şenyurt, A. A. Sisi, B. Ay, P. Traversa, J. Douglas, F. Cotton, L. Luzi, B. Hernandez, et al. (2014b). Reference database for seismic ground-motion in Europe (RESORCE), Bull. Earthquake Eng. **12**(1), 311–339.

Ambraseys, N., J. Douglas, S. Sarma, and P. Smit (2005). Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: horizontal peak ground acceleration and spectral acceleration, Bull. Earthquake Eng. **3**(1), 1–53.

Ancheta, T. D., R. B. Darragh, J. P. Stewart, E. Seyhan, W. J. Silva, B. S. Chiou, K. E. Wooddell, R. W. Graves, A. R. Kottke, D. M. Boore, et al. (2014). NGA-West2 database, Earthquake Spectra **30**(3), 989–1005.

Arup (2014). Temporary shale gas exploration, Preston New Road, Lancashire: Environmental Statement.

Atkinson, G. M. (2008). Ground-motion prediction equations for eastern North America from a referenced empirical approach: Implications for epistemic uncertainty, Bull. Seismol. Soc. Am. **98**(3), 1304–1318.

Atkinson, G. M. (2015). Ground-motion prediction equation for small-to-moderate events at short hypocentral distances, with application to induced-seismicity hazards, Bull. Seismol. Soc. Am. **105**(2A), 981–992.

Atkinson, G. M. and M. Morrison (2009). Observations on regional variability in ground-motion amplitudes for small-to-moderate earthquakes in North America, Bull. Seismol. Soc. Am. **99**(4), 2393–2409.

Baltay, A. S. and T. C. Hanks (2014). Understanding the magnitude dependence of PGA and PGV in NGA-West 2 data, Bull. Seismol. Soc. Am. **104**(6), 2851–2865.

Beauval, C., H. Tasan, A. Laurendeau, E. Delavaud, F. Cotton, P. Guéguen, and N. Kuehn (2012). On the testing of ground-motion prediction equations against small-magnitude data, Bull. Seismol. Soc. Am. **102**(5), 1994–2007.

Bommer, J. J. and N. A. Abrahamson (2006). Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates?, Bull. Seismol. Soc. Am. **96**(6), 1967–1977.

Bommer, J. J., J. Douglas, F. Scherbaum, F. Cotton, H. Bungum, and D. Fah (2010). On the selection of ground-motion prediction equations for seismic hazard analysis, Seismol. Res. Lett. **81**(5), 783–793.

Bommer, J. J., P. J. Stafford, J. E. Alarcón, and S. Akkar (2007). The influence of magnitude range on empirical ground-motion prediction, Bull. Seismol. Soc. Am. **97**(6), 2152–2170.

Boore, D. M. (2010). Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion, Bull. Seismol. Soc. Am. **100**(4), 1830–1835.

Bourne, S., S. Oates, J. Bommer, B. Dost, J. Van Elk, and D. Doornhof (2015). A Monte Carlo method for probabilistic hazard assessment of induced seismicity due to conventional natural gas production, Bull. Seismol. Soc. Am. **105**(3), 1721–1738.

Bradley, B. A. (2013). A New Zealand-specific pseudospectral acceleration ground-motion prediction equation for active shallow crustal earthquakes based on foreign models, Bull. Seismol. Soc. Am. **103**(3), 1801–1822.

Butcher, A., R. Luckett, J.-M. Kendall, and B. Baptie (2019). Corner frequencies, seismic moments and earthquake magnitudes: The effects of high-frequency attenuation on microseismicity, Bull. Seismol. Soc. Am. (submitted).

Butcher, A., R. Luckett, J. P. Verdon, J.-M. Kendall, B. Baptie, and J. Wookey (2017). Local magnitude discrepancies for near-event receivers: Implications for the UK traffic-light scheme, Bull. Seismol. Soc. Am. **107**(2), 532–541.

Campbell, N., C. Fenton, and S. Tallett-Williams (2016). An Investigation into the Effects of Material Properties on Shear Wave Velocity in Rocks/Soils. In Proceedings of the 5th International Conference on Geotechnical and Geophysical Site Characterization ISC-5, Gold Coast, Australia.

Chiou, B., R. Youngs, N. Abrahamson, and K. Addo (2010). Ground-motion attenuation model for small-to-moderate shallow crustal earthquakes in California and its implications on regionalization of ground-motion prediction models, Earthquake spectra **26**(4), 907–926.

Chiou, B. S.-J. and R. R. Youngs (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra, Earthquake Spectra **30**(3), 1117–1153.

Chun, M.-H., S.-J. Han, and N.-I. Tak (2000). An uncertainty importance measure using a distance metric for the change in a cumulative distribution function, Reliab. Eng. Syst. Saf. **70**(3), 313–321.

Clarke, H., L. Eisner, P. Styles, and P. Turner (2014). Felt seismicity associated with shale gas hydraulic fracturing: The first documented example in Europe, Geophys. Res. Lett. **41**(23), 8308–8314.

Clarke, H., J. P. Verdon, T. Kettlety, A. Baird, and J.-M. Kendall (2019). Real-time imaging, forecasting, and management of human-induced seismicity at Preston New Road, Lancashire, England, Seismol. Res. Lett. **90**(5), 1902–1915.

Cremen, G., M. J. Werner, and B. Baptie (2019). Understanding induced seismicity hazard related to shale gas exploration in the UK. In SECED 2019 Conference: Earthquake Risk and Engineering towards a Resilient World.

Douglas, J., B. Edwards, V. Convertito, N. Sharma, A. Tramelli, D. Kraaijpoel, B. M. Cabrera, N. Maercklin, and C. Troise (2013). Predicting ground motion from induced earthquakes in geothermal areas, Bull. Seismol. Soc. Am. **103**(3), 1875–1897.

Douglas, J. and P. Jousset (2011). Modeling the difference in ground-motion magnitude-scaling in small and large earthquakes, Seismol. Res. Lett. **82**(4), 504–508.

Haney, M. M., J. Power, M. West, and P. Michaels (2012). Causal instrument corrections for short-period and broadband seismometers, Seismol. Res. Lett. **83**(5), 834–845.

Jiang, J. (2007). Linear and generalized linear mixed models and their applications. Springer Science & Business Media.

Kale, Ö. and S. Akkar (2013). A new procedure for selecting and ranking ground-motion prediction equations (GMPEs): The Euclidean distance-based ranking (EDR) method, Bull. Seismol. Soc. Am. **103**(2A), 1069–1084.

Kwak, S. G. and J. H. Kim (2017). Central limit theorem: the cornerstone of modern statistics, Korean J. Anesthesiol. **70**(2), 144.

Laird, N. (2004). Chapter 5: Random effects and the linear mixed model, Volume 8 of Regional Conference Series in Probability and Statistics, pp. 79–95. Beechwood OH and Alexandria VA: Institute of Mathematical Statistics and American Statistical Association.

Lee, S.-Y. and X.-Y. Song (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes, Multivariate Behav. Res. **39**(4), 653–686.

Lindley, D. V. (1991). Making Decisions (2 ed.). John Wiley & Sons.

Mak, S., R. A. Clements, and D. Schorlemmer (2014). Comment on "A new procedure for selecting and ranking ground-motion prediction equations (GMPEs): The Euclidean distance-based ranking (EDR) method" by Özkan Kale and Sinan Akkar, Bull. Seismol. Soc. Am. **104**(6), 3139–3140.

Mak, S., R. A. Clements, and D. Schorlemmer (2017). Empirical evaluation of hierarchical ground-motion models: Score uncertainty and model weighting, Bull. Seismol. Soc. Am. **107**(2), 949–965.

Ornthammarath, T., J. Douglas, R. Sigbjörnsson, and C. G. Lai (2011). Assessment of ground motion variability and its effects on seismic hazard analysis: A case study for Iceland, Bull. Earthquake Eng. **9**(4), 931–953.

Perron, V., A. Laurendeau, F. Hollender, P.-Y. Bard, C. Gélis, P. Traversa, and S. Drouet (2018). Selecting time windows of seismic phases and noise for engineering seismology applications: A versatile methodology and algorithm, Bull. Earthquake Eng. **16**(6), 2211–2225.

Scasserra, G., J. P. Stewart, P. Bazzurro, G. Lanzo, and F. Mollaioli (2009). A comparison of NGA ground-motion prediction equations to Italian data, Bull. Seismol. Soc. Am. **99**(5), 2961–2978.

Scherbaum, F., F. Cotton, and P. Smit (2004). On the use of response spectral-reference data for the selection and ranking of ground-motion models for seismic-hazard analysis in regions of moderate seismicity: The case of rock motion, Bull. Seismol. Soc. Am. **94**(6), 2164–2185.

Scherbaum, F., E. Delavaud, and C. Riggelsen (2009). Model selection in seismic hazard analysis: An information-theoretic perspective, Bull. Seismol. Soc. Am. **99**(6), 3234–3247.

Selley, R. C. (2012). UK shale gas: the story so far, Mar Petrol Geol **31**(1), 100–109.

Seyhan, E. and J. P. Stewart (2014). Semi-empirical nonlinear site amplification from NGA-West2 data and simulations, Earthquake Spectra **30**(3), 1241–1256.

Stafford, P. J. (2015). Extension of the random-effects regression algorithm to account for the effects of nonlinear site response, Bull. Seismol. Soc. Am. **105**(6), 3196–3202.

Stafford, P. J., F. O. Strasser, and J. J. Bommer (2008). An evaluation of the applicability of the NGA models to ground-motion prediction in the Euro-Mediterranean region, Bull. Earthquake Eng. **6**(2), 149–177.

Stewart, J. P., J. Douglas, M. Javanbarg, Y. Bozorgnia, N. A. Abrahamson, D. M. Boore, K. W. Campbell, E. Delavaud, M. Erdik, and P. J. Stafford (2015). Selection of ground motion prediction equations for the global earthquake model, Earthquake Spectra **31**(1), 19–45.

Verdon, J. P., J.-M. Kendall, A. Butcher, R. Luckett, and B. J. Baptie (2017). Seismicity induced by longwall coal mining at the Thoresby Colliery, Nottinghamshire, UK, Geophys. J. Int. **212**(2), 942–954.

Villani, C. (2008). Optimal transport: old and new, Volume 338. Springer Science & Business Media.

Wang, L.-J. (1996). Processing of near-field earthquake accelerograms, California Institute of Technology.

Wasserstein, R. L. and N. A. Lazar (2016). The ASA's statement on p-values: context, process, and purpose, The American Statistician **70**(2), 129–133.

Tables

Table 1: Scores calculated using the proposed procedure, for the cases in Examples 1-3 of Mak et al. (2017).

Example	Case	EMD_{total}	$\ell\ell(\mathbf{p},\mathbf{V},\mathbf{q})$
1	Case 1	**0.25**	38.8
	Case 2	0.39	**38.5**
2	Correct	**0.25**	**61.2**
	Biased	0.48	61.5
3	Correct	0.25	**38.8**
	Inflated σ_b	0.35	39.6
	Deflated σ_b	**0.17**	39.1

$\ell\ell(\mathbf{p},\mathbf{V},\mathbf{q})$ scores of the Mak et al. (2017) procedure are also shown for comparison. Note that the smallest score for a given procedure (marked in bold) indicates the best model.

Table 2: Ranking of GMPEs for suitability to modelling ground motions produced by UK shale gas-related seismicity, using both the proposed procedure and the procedure of Mak et al. (2017).

Intensity Measure	Metric	$\mathrm{ASB14_{hypo}}$	$\mathrm{ASB14_{epi}}$	A15	D13
PGA	EMD_{total}	4.56	3.47	1.48	**0.74**
	$\ell\ell(\mathbf{p},\mathbf{V},\mathbf{q})$	1830	1297	569	**505**
PGV	EMD_{total}	1.88	0.95	1.25	**0.92**
	$\ell\ell(\mathbf{p},\mathbf{V},\mathbf{q})$	763	605	**493**	645
$SA_{0.05}$	EMD_{total}	4.64	3.63	1.39	**0.62**
	$\ell\ell(\mathbf{p},\mathbf{V},\mathbf{q})$	1912	1428	610	**550**
$SA_{0.1}$	EMD_{total}	4.82	3.78	1.54	**1.17**
	$\ell\ell(\mathbf{p},\mathbf{V},\mathbf{q})$	1912	1404	571	**535**
$SA_{0.2}$	EMD_{total}	5.31	4.26	**1.06**	1.81
	$\ell\ell(\mathbf{p},\mathbf{V},\mathbf{q})$	2178	1579	**474**	605

Note that the smallest score for a given procedure (marked in bold) indicates the best model.

Table 3: Coefficients of CWB19 for all ground motion intensity measures (IMs) examined. Note that the functional form of the GMPE is presented in equation 11.

IM	a	b	c	h	d	ϕ	τ	σ
PGA	-5.096	2.146	-2.611	constrained to zero	-0.023	0.563	0.437	0.712
PGV	-10.213	2.913	-2.719	constrained to zero	-0.046	0.553	0.158	0.575
$SA_{0.05}$	-5.027	2.717	-2.890	constrained to zero	-0.008	0.696	0.378	0.792
$SA_{0.1}$	-4.988	2.814	-2.723	constrained to zero	-0.039	0.632	0.227	0.672
$SA_{0.2}$	-7.704	3.639	-2.276	constrained to zero	-0.057	0.549	0.430	0.698

Table 4: Percentage improvement in modelling accuracy offered by CWB19 over D13 for the data of interest in this study.

IM	$(EMD_{total})_{CWB19}$	$(EMD_{total})_{D13}$	% Improvement
PGA	0.21	0.74	72
PGV	0.48	0.92	48
$SA_{0.05}$	0.26	0.62	58
$SA_{0.1}$	0.40	1.17	66
$SA_{0.2}$	0.25	1.81	86

Note that IM stands for ground motion intensity measure. Values for $(EMD_{total})_{D13}$ are taken from Table 2.

Figures

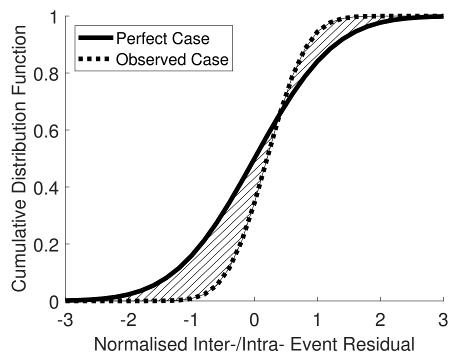
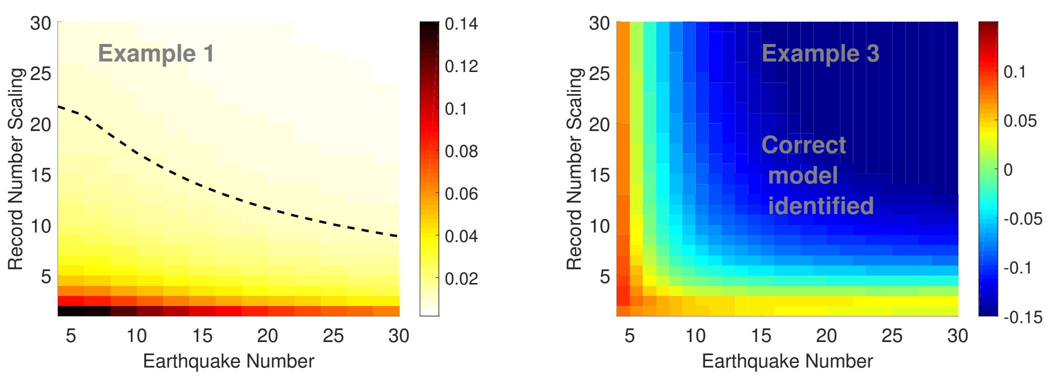


Figure 1: A graphical representation of the scoring system for our proposed GMPE evaluation procedure, which quantifies the distance between the CDF of the standard normal distribution (perfect case) and that of the maximum likelihood normal distribution (observed case) for each type of normalised residual. $\mu_x = 0.5$ and $\sigma_x = 0.5$ for the observed case, therefore $EMD_x = \sqrt{0.5^2 + (0.5-1)^2} = 0.7$ in this case.



(a)

(b)

Figure 2: Understanding the sample sizes necessary for the proposed evaluation procedure to perform correctly in Examples 1 and 3 of Mak et al. (2017). (a) Absolute difference between the EMD_{total} values for Case 1 and Case 2 in Example 1, (black dashed line indicates a value of 0.01) and (b) difference between the EMD_{total} values for the correct model and the model with deflated σ_b, as a function of earthquake number and the scaling of record number per earthquake. Note that lighter colours in (a) and darker blue colours in (b) indicate a more correct performance of the score.

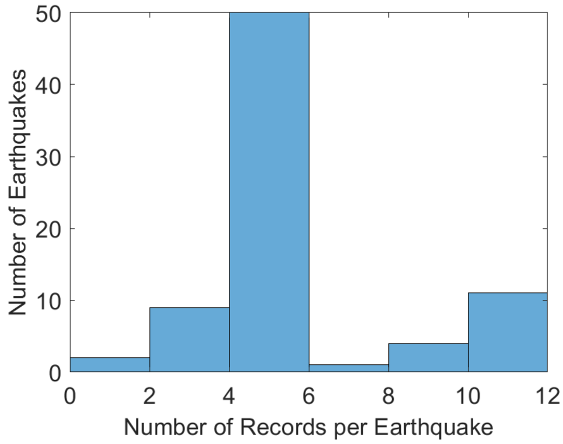


Figure 3: Histogram of the complete observed ground motion record database used in this study.

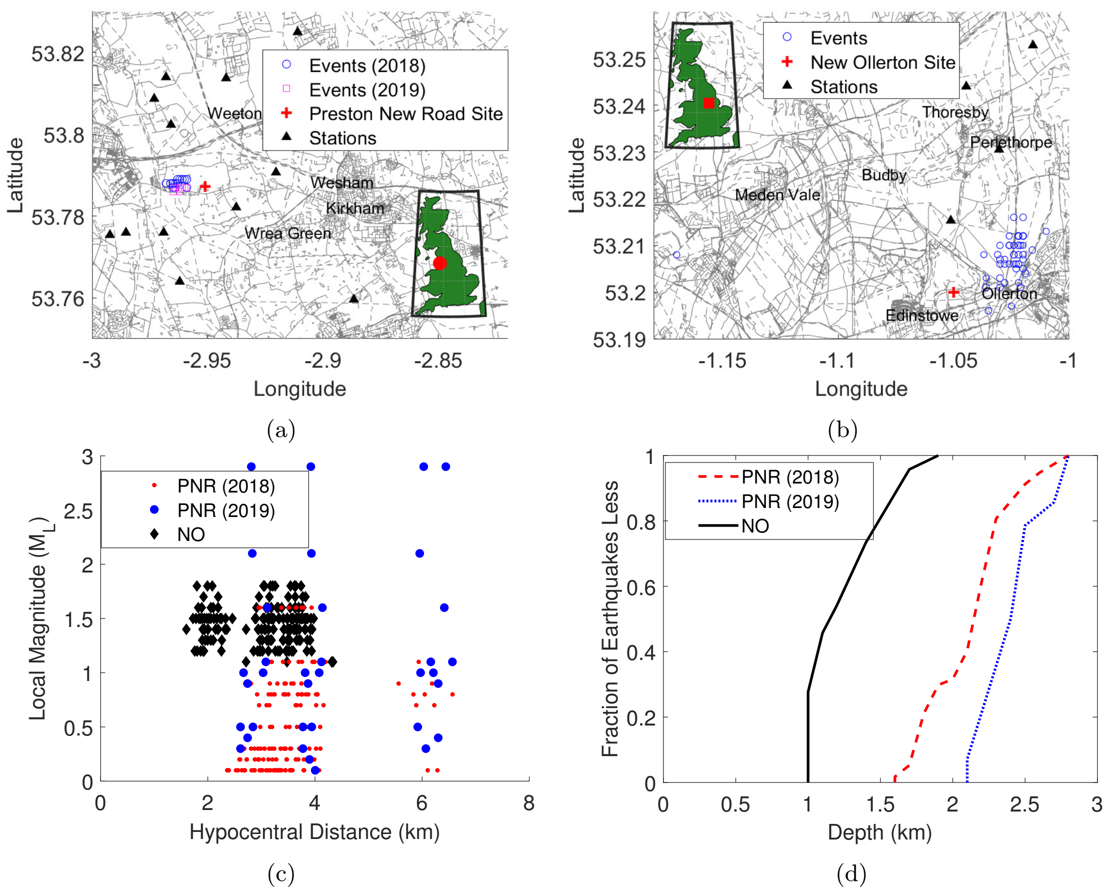


Figure 4: A summary of the data examined in this study. (a) Locations of the considered seismicity and seismic monitoring stations for the Preston New Road (PNR) shale gas site in Lancashire and (b) the Thoresby Colliery at New Ollerton (NO), North Nottinghamshire (insets highlight locations relative to all of Great Britain). (c) Magnitude, hypocentral distance, and (d) depth data examined.

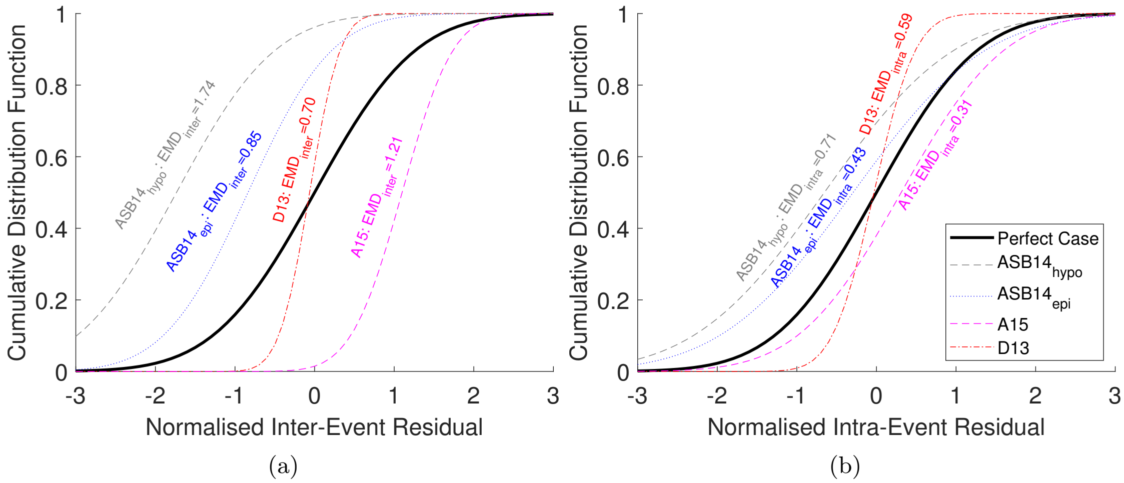
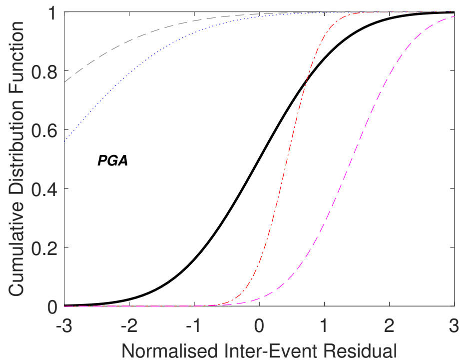
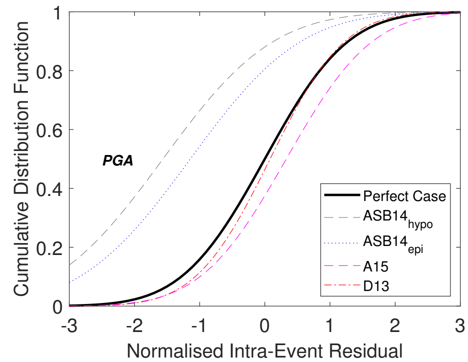


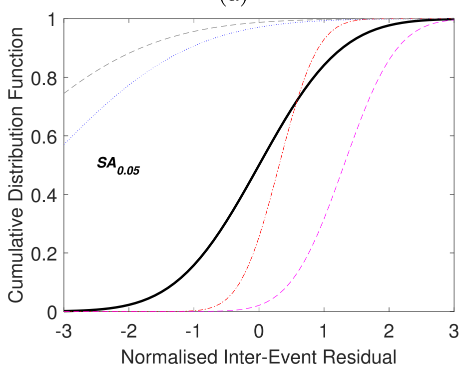
Figure 5: Normalised (a) inter- and (b) intra-event PGV residuals for the four GMPEs evaluated, compared with those expected from a standard normal distribution (the 'Perfect Case'). Also plotted are EMD scores for each type of residual.



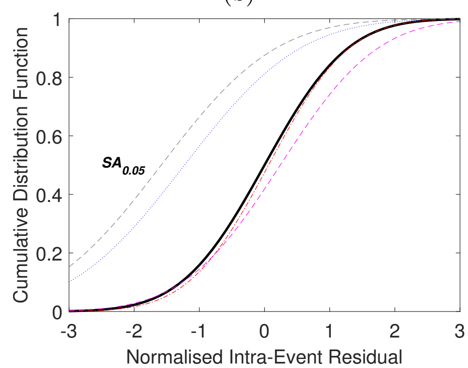
(a)



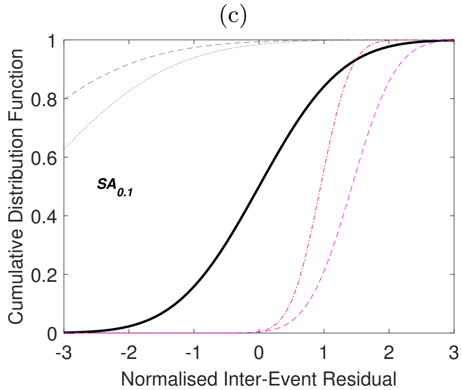
(b)



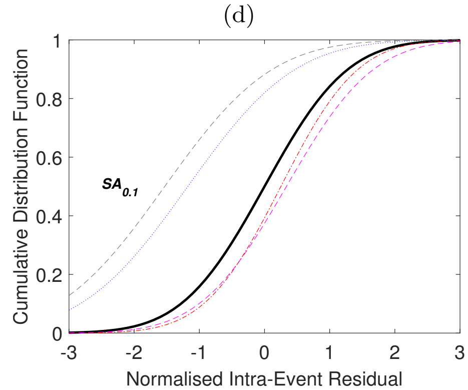
(c)



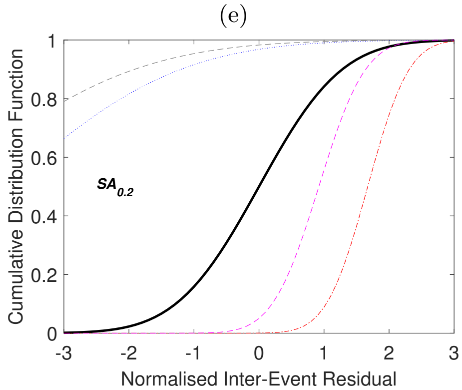
(d)



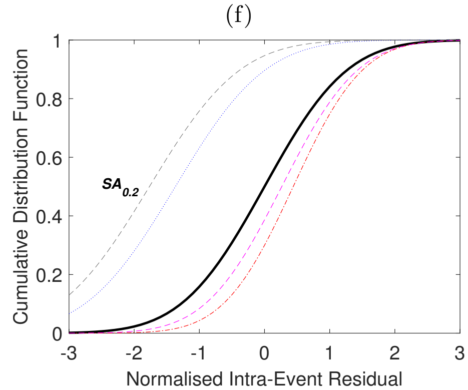
(e)



(f)



(g)



(h)

Figure 6: Normalised (a, c, e, g) inter- and (b, d, f, h) intra-event PGA, $SA_{0.05}$, $SA_{0.1}$, and $SA_{0.2}$ residuals for the four GMPEs evaluated, compared with those expected from a standard normal distribution (the 'Perfect Case').

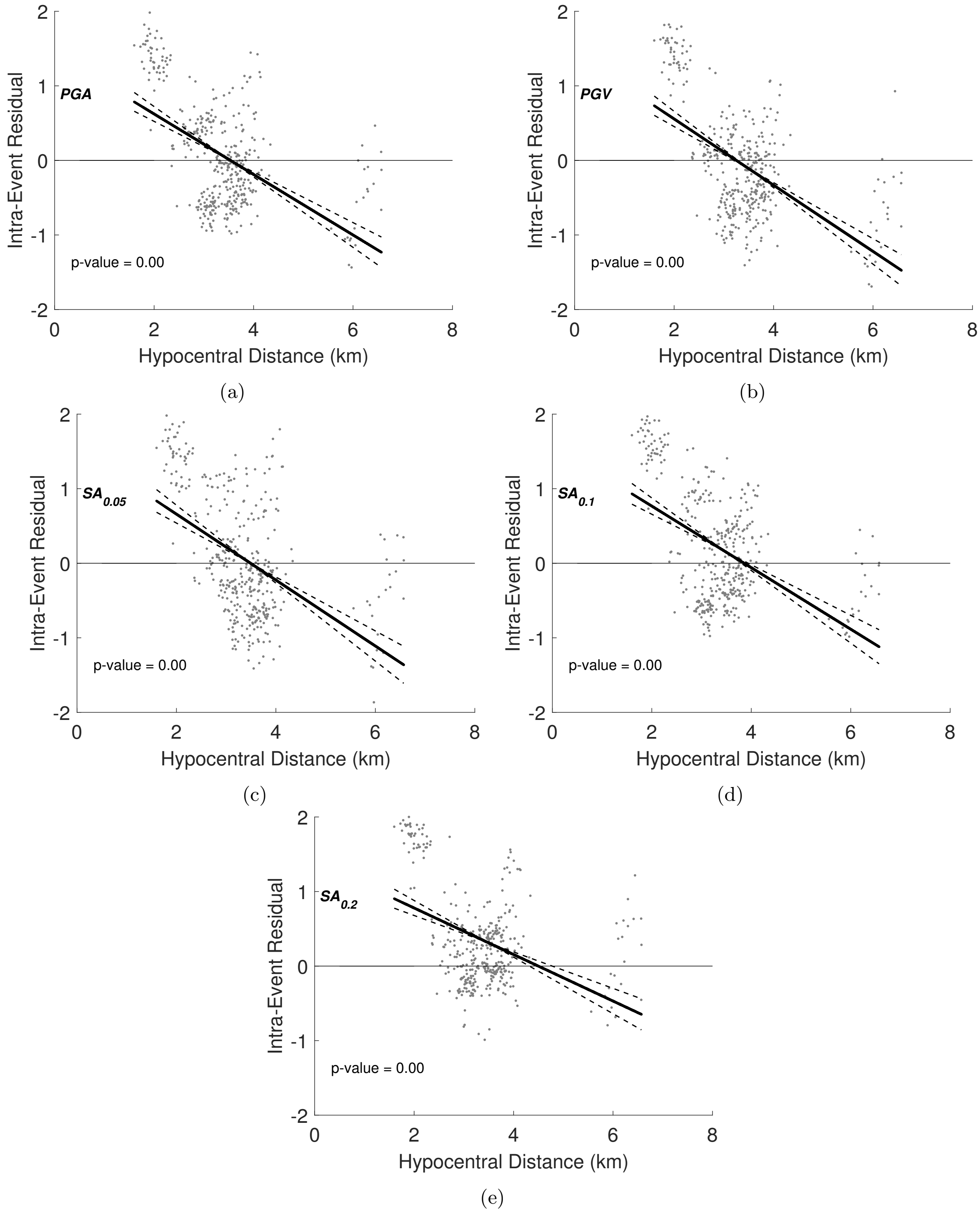


Figure 7: Variation of the D13 normalised intra-event residuals with hypocentral distance for (a) PGA, (b) PGV, (c) $SA_{0.5}$, (d) $SA_{0.1}$, and (e) $SA_{0.2}$. Also shown are the lines fit using linear regression (solid black lines) and their 95% confidence intervals (dashed lines). The p-value for a given plot tests the null hypothesis that the slope of the fitted line equals zero.

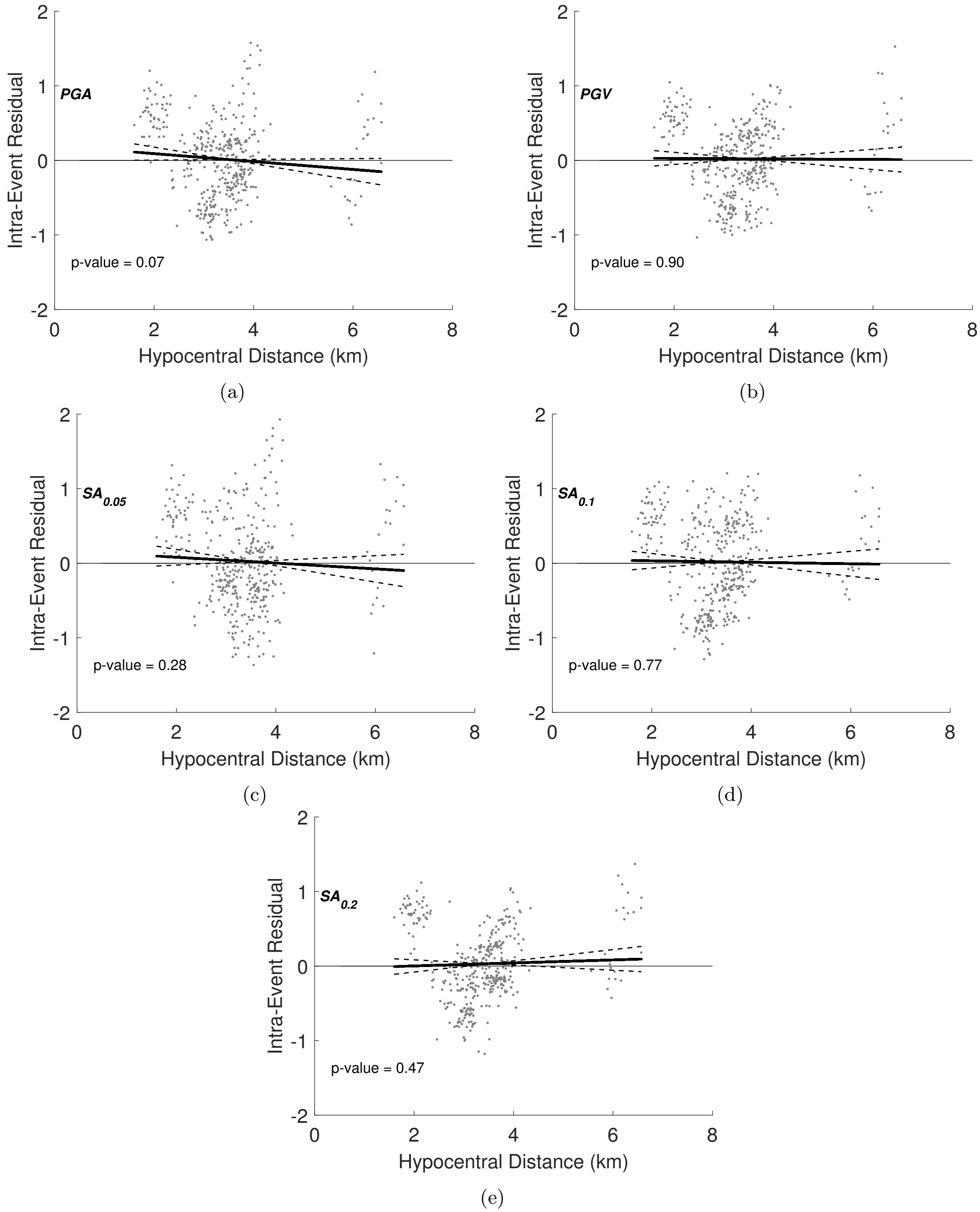


Figure 8: Variation of the distance-modified D13 normalised intra-event residuals with hypocentral distance for (a) PGA, (b) PGV, (c) $SA_{0.5}$, (d) $SA_{0.1}$, and (e) $SA_{0.2}$. Also shown are the lines fit using linear regression (solid black lines) and their 95% confidence intervals (dashed lines). The p-value for a given plot tests the null hypothesis that the slope of the fitted line equals zero.

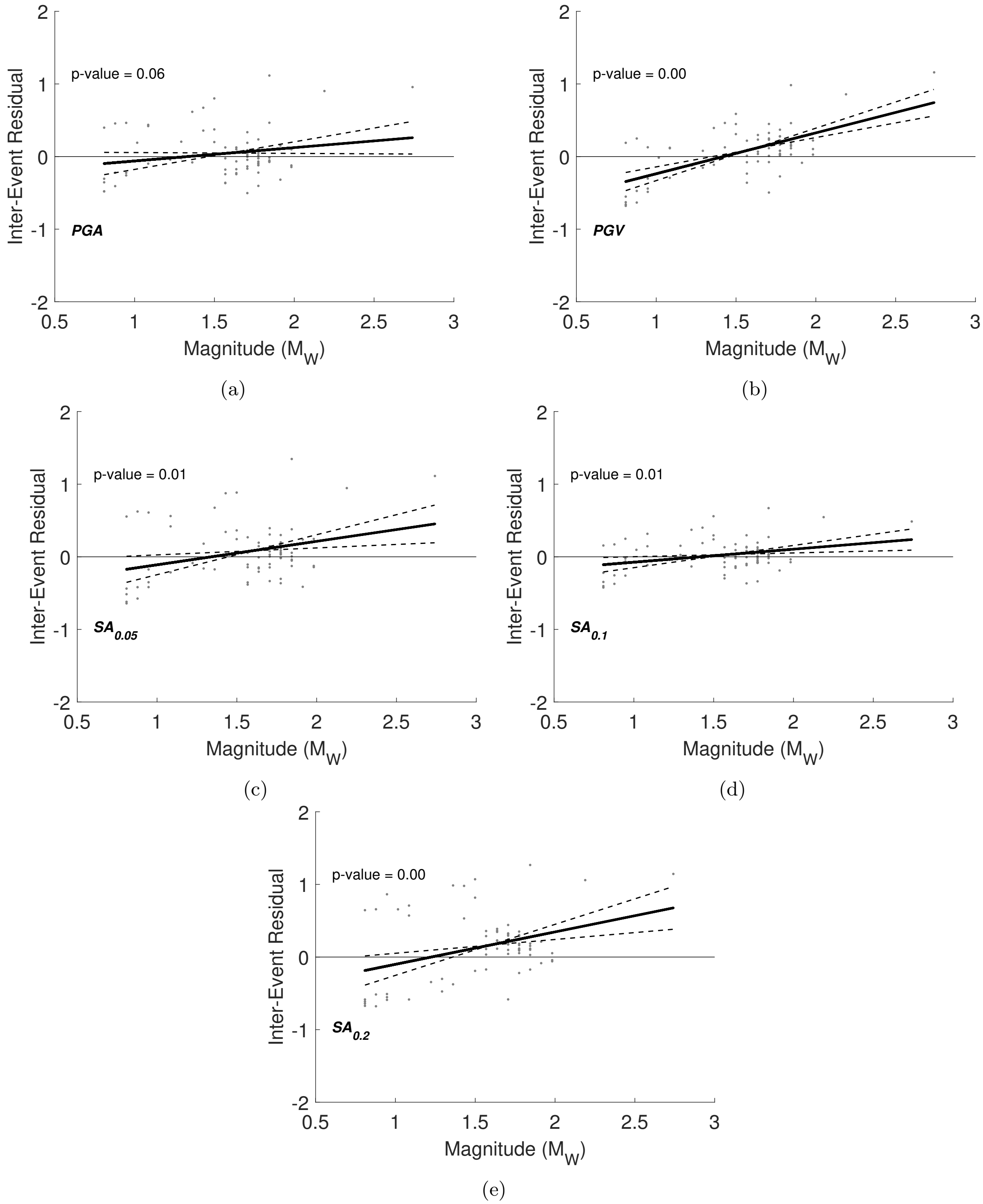


Figure 9: Variation of the distance-modified D13 normalised inter-event residuals with magnitude for (a) PGA, (b) PGV, (c) $SA_{0.5}$, (d) $SA_{0.1}$, and (e) $SA_{0.2}$. Also shown are the lines fit using linear regression (solid black lines) and their 95% confidence intervals (dashed lines).The p-value for a given plot tests the null hypothesis that the slope of the fitted line equals zero.

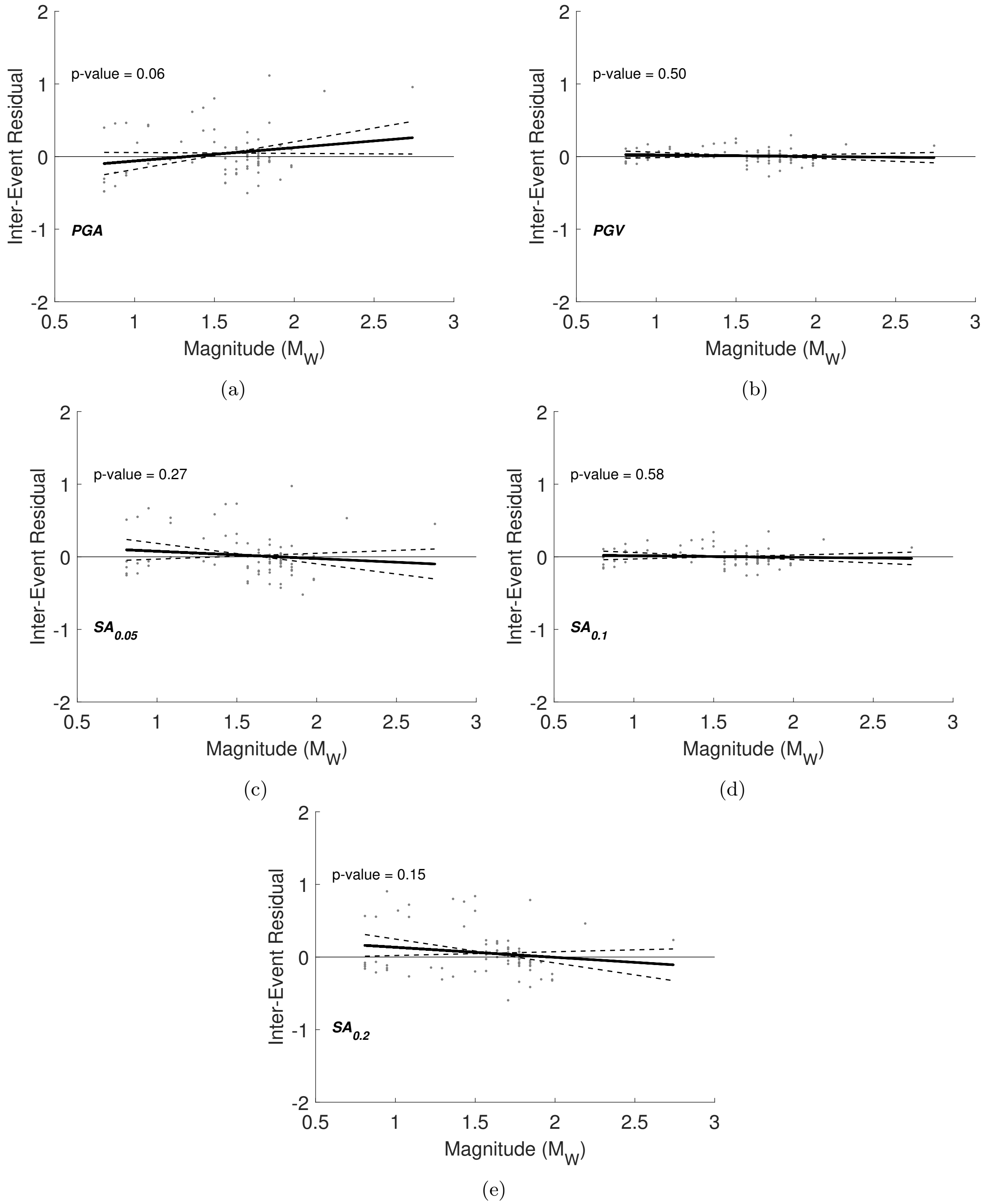
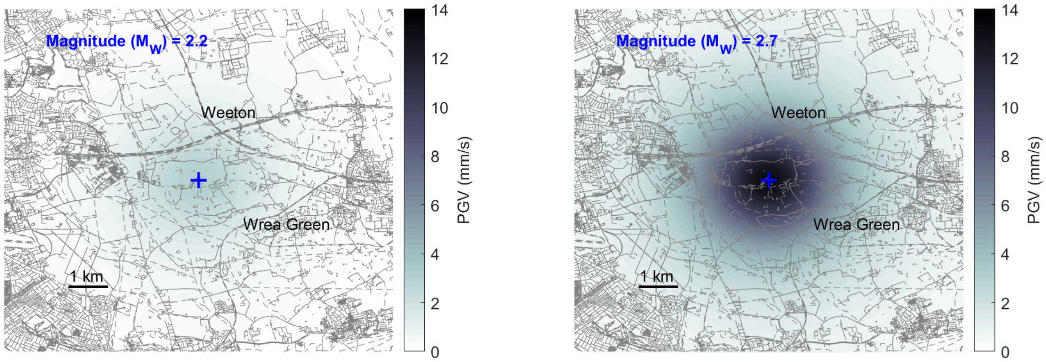


Figure 10: Variation of the CWB19 normalised inter-event residuals with magnitude for (a) PGA, (b) PGV, (c) $SA_{0.5}$, (d) $SA_{0.1}$, and (e) $SA_{0.2}$. Also shown are the lines fit using linear regression (solid black lines) and their 95% confidence intervals (dashed lines). The p-value for a given plot tests the null hypothesis that the slope of the fitted line equals zero.



(a) (b)

Figure 11: CWB19 median predictions of PGV within the PNR greater region, for two hypothetical scenarios: (a) an earthquake with $M_w = 2.2$ and (b) an earthquake with $M_w = 2.7$, that are co-located with the PNR shale gas site (blue cross) at a depth of 2 km.

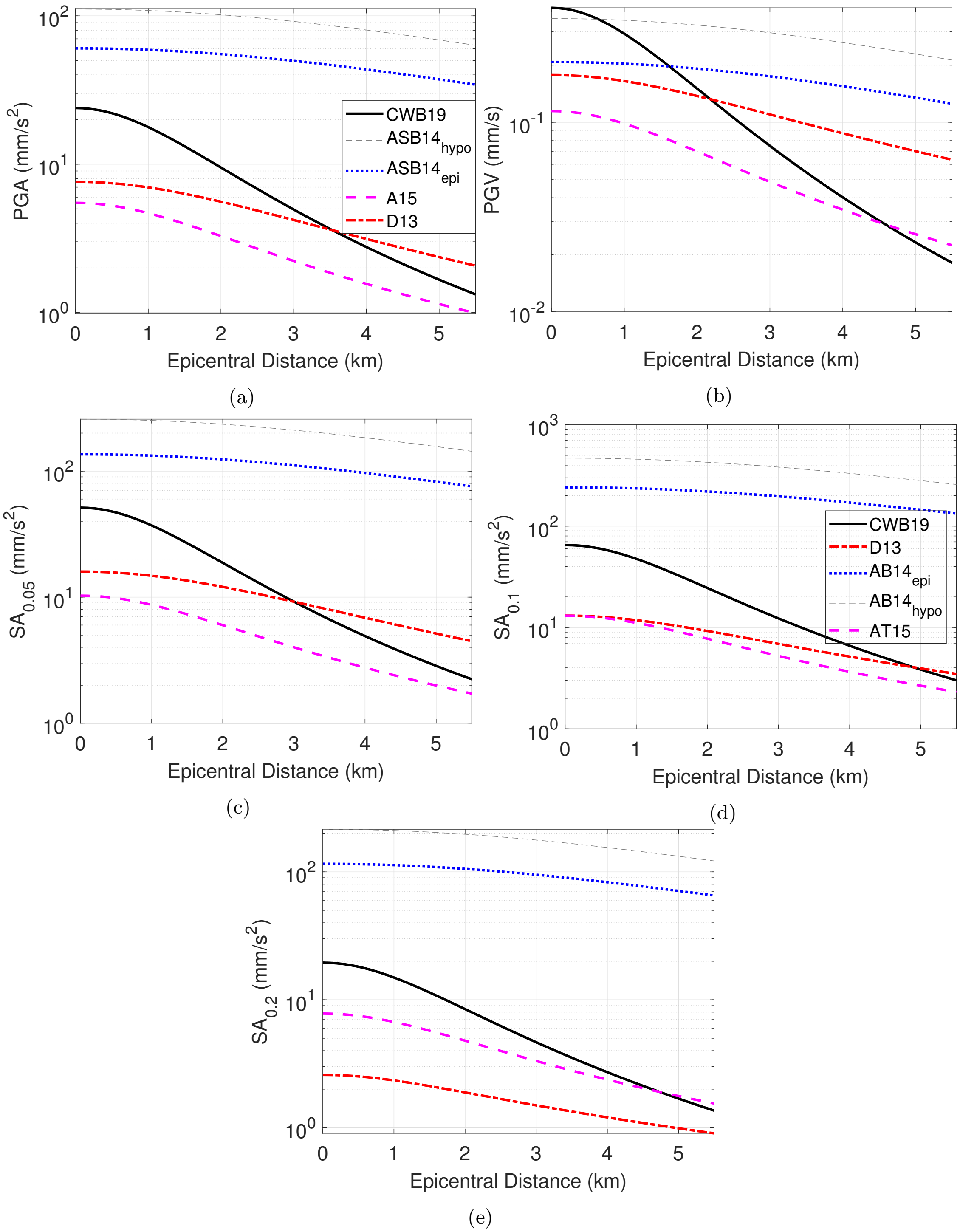


Figure 12: Distance-scaling of CWB19 for a fixed focal depth of 2 km and a moment magnitude of 1.5, compared with the equivalent distance-scaling of other GMPEs examined in this study, for (a) PGA, (b) PGV, (c) $SA_{0.05}$, (d) $SA_{0.1}$, and (e) $SA_{0.2}$.

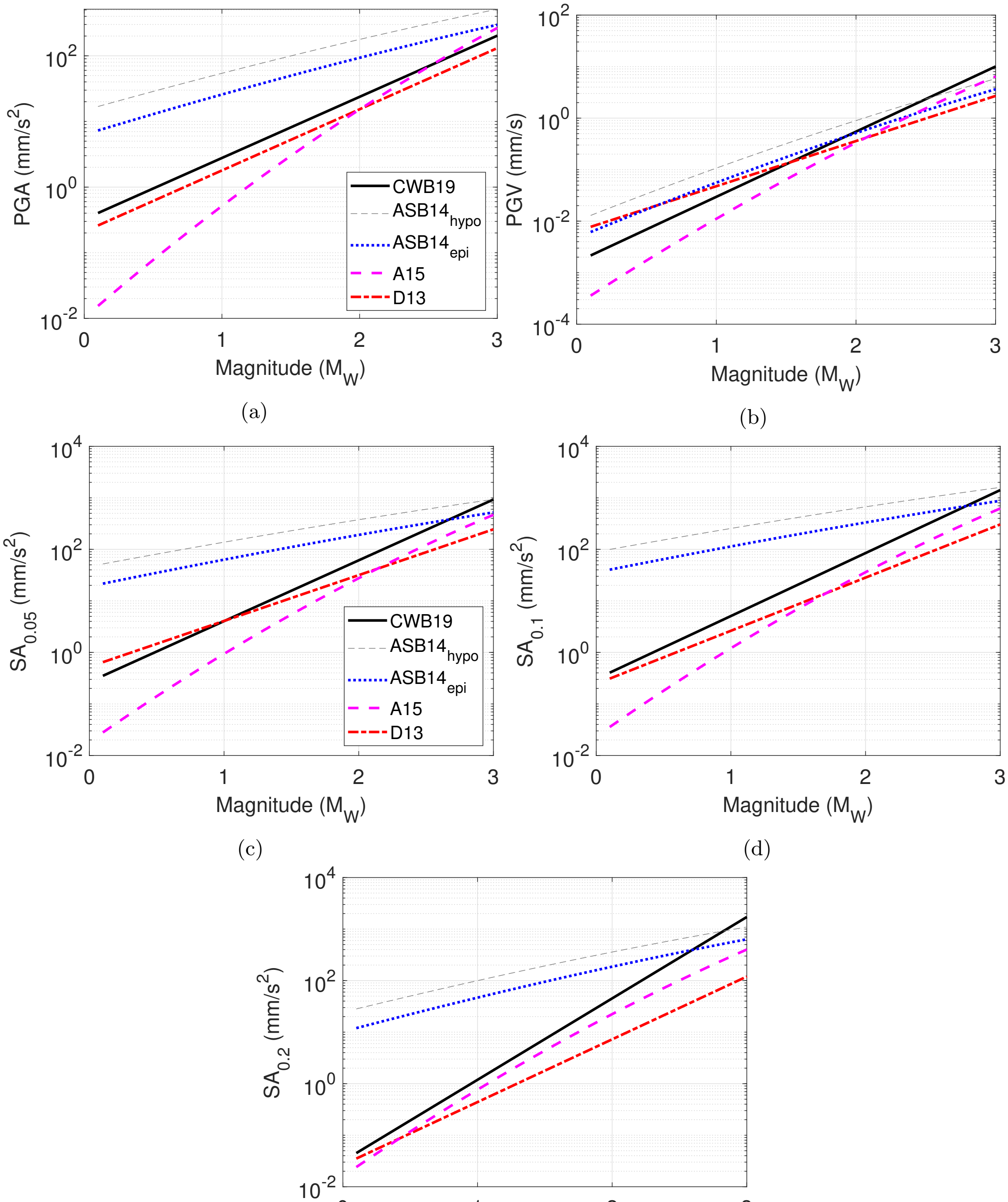
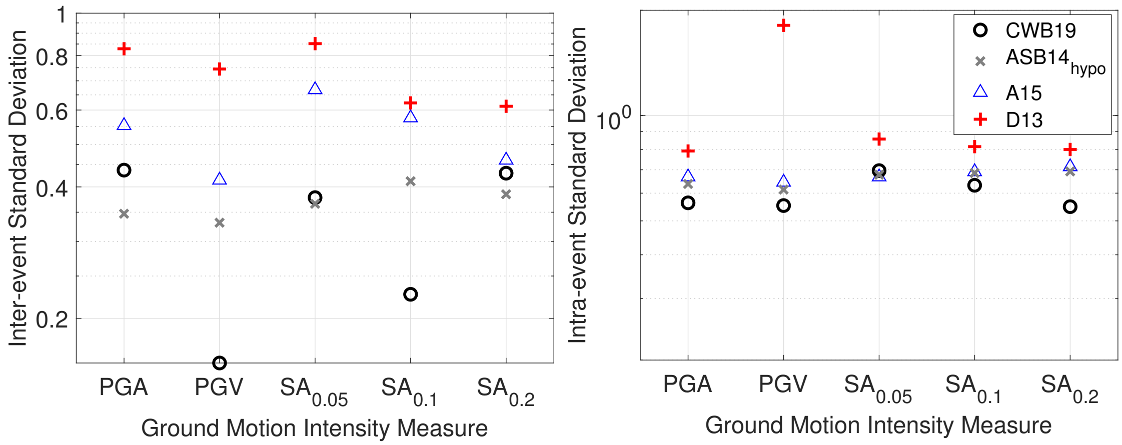


Figure 13: Magnitude-scaling of CWB19 at 3 km, compared with the equivalent magnitude-scaling of other GMPEs examined in this study, for (a) PGA, (b) PGV, (c) $SA_{0.05}$, (d) $SA_{0.1}$, and (e) $SA_{0.2}$.



(a) (b)

Figure 14: (a) Inter- and (b) intra-event standard deviations (in natural log units) for CWB19, compared with equivalent values for other GMPES examined in this study. Note that $\mathrm{ASB14_{epi}}$ data are not included for clarity, since they are almost identical to those of $\mathrm{ASB14_{hypo}}$ (Akkar et al., 2014a).