ORIGINAL ARTICLE



Defining the usable bandwidth of weak-motion records: application to induced seismicity in the Groningen Gas Field, the Netherlands

Benjamin Edwards • Michail Ntinalexis

Received: 13 February 2021 / Accepted: 27 April 2021 / Published online: 24 May 2021 © The Author(s) 2021

Abstract Seismic hazard and risk analyses are increasingly tapping into the previously underused resource of local weak-motion records. This is facilitating the development of local- or even application-specific models for the characterisation of earthquake ground motion. In turn, this offers the opportunity to derive non- or partially non-ergodic models and significantly reduce bias and uncertainty. However, weak-motion data, while carrying important information about local earthquake source, path and site effects, are susceptible to noise. We show that high-frequency noise has a record-, or regionspecific, impact on pseudo-spectral acceleration (PSA). This impact depends on the shape of the records' Fourier amplitude spectrum (FAS): PSA from moderately to highly damped 'soil' records (e.g. Groningen, the Netherlands) is much less susceptible to high-frequency noise than PSA from weakly damped 'rock' records

Highlights

- We show that PSA from weak-motion records are susceptible to high-frequency noise, which can lead to over-estimation of spectral ordinates.
- A model defining the minimum usable period of 5% damped PSA, T_{\min} , is presented, based on easily measurable characteristics of a waveform.
- A full workflow, defining the upper and lower usable frequencies (of FAS) and periods (of 5 % damped PSA) is presented, with application to an induced seismicity dataset.

B. Edwards (⊠)

School of Environmental Sciences, University of Liverpool, L69 3GP, Liverpool, UK

e-mail: ben.edwards@liverpool.ac.uk

M. Ntinalexis Independent Consultant, London, UK (e.g. Eastern North America). We make use of simulated ground motion records to develop a parametric model for the lower usable period of PSA (T_{\min}) . The model accounts for the impact of high-frequency noise on PSA, conditional on easily measured parameters characterising the shape of a record's FAS. We then present a workflow, describing processing undertaken for records of induced seismicity from the Groningen gas field. The workflow includes the definition of maximum and minimum usable frequencies and periods of FAS and PSA, respectively. As part of the workflow, we present an approach that considers multiple estimates of T_{\min} . These include the parametric model and, additionally, record-specific hybrid simulations that artificially extend or modify time series' FAS beyond the noise floor to assess subsequent impacts on PSA.

Keywords Ground motion records · Seismic hazard · Seismic risk · Induced seismicity

1 Introduction

Ground motion models (GMMs) are a key component of a variety of scientific and engineering products, from seismic hazard and risk analyses, to *shakemaps* and magnitude scales. GMMs are developed using empirical data, either through direct regression, or in the case of simulation-based models, through calibration to recorded data. The quality of the underlying ground motion dataset is therefore of significant importance. The main issues that may reduce the quality of ground motion



records are the instrument and datalogger (i.e. the record's fidelity), and the background noise. The use of modern instrumentation, with broadband flat sensitivity, high-resolution dataloggers and reliable data transmission means that the main constraint on the usability of an earthquake ground motion record now lies with background noise. A great deal of attention has been paid to the processing of earthquake records in strong ground motion datasets to avoid or minimise the impact of noise on response spectra at long periods (Akkar and Bommer 2006) and short periods (Douglas and Boore 2011). However, little attention has been paid to the impact of noise on weak-motion records, which are increasingly utilised in studies developing application- and location-specific GMMs (e.g. Bommer et al. 2016; Novakovic et al. 2018; Edwards et al. 2021).

The recent increase in induced seismicity related to emerging industries, such as hydraulic fracturing, geothermal energy and CO₂ sequestration is of particular global concern, especially within the context of the transition to low-carbon economies. Induced seismicity tends to be of smaller magnitude, however is often at shallow depths and in close proximity to urban areas. This means that smaller magnitude events contribute to the seismic hazard and risk of local population centres. Furthermore, there is increasing recognition that ground motions are regionally variable, particularly for small to moderate earthquakes occurring at shallow depths (Douglas and Edwards 2016). As a result, engineers must either modify existing models, or develop local GMMs using regional or, better still, data local to the target site. This spatial limitation means that motions will inevitably be of smaller amplitude if uncertainty and biases in developed models are to be reduced.

Seismic noise is a continuous, but variable, vibration with various sources. At low frequency, the microseism dominates and is related to natural phenomena such as ocean waves (Webb 1992). At high frequency, sources of seismic noise tend to be anthropogenic, owing to their lower propagation efficiency. Monitoring instruments in close proximity to urban environments are therefore susceptible to both low- and high-frequency disturbances. Cauzzi and Clinton (2013) and Peterson (1993) provide an overview of seismic noise and develop models for 'high' and 'low' noise cases. In terms of the impact of seismic noise on ground motion time-histories or, equivalently, Fourier spectra, an

unambiguous assessment is possible by comparing earthquake records with 'noise records'. Noise records are typically taken from the immediate pre-event time-history (such that transient noise at the time of recording may be captured), but equally an estimate could be reconstructed based on the high- and lownoise models (Cauzzi and Clinton 2013; Peterson 1993). The ratio of signal (plus noise) to noise Fourier spectral amplitudes, typically referred to as the signal-to-noise ratio (SNR), provides a useful measure of signal contamination. Thresholds above three are usually considered suitable, but this varies between applications, with authors typically striking a balance between data quality and quantity.

The impact of noise on response spectral ordinates such as pseudo-spectral acceleration (PSA), which form the basis of seismic hazard and risk analyses, is more difficult to quantify than for Fourier amplitude spectra (FAS). This is because of the non-linear transform between the Fourier domain (representing the signal amplitude at a given signal frequency) and response spectral amplitudes (representing peak motions of an oscillator with characteristic period). Bora et al. (2016) show that this transform results in a roughly linear relationship between low Fourier frequencies and long oscillator periods (i.e. $FAS(1/f) \propto PSA(T)$ for T>0.5 s), but quickly becomes non-linear as periods reduce. At long periods, Akkar and Bommer (2006) showed that a usability limit of $T_{\text{max}} = [0.7 \text{ to } 0.97]/f_1$ was required to avoid the impact of long period noise on PSA, with f_1 (the minimum uncontaminated signal frequency) defined, for example, by a SNR threshold.

On the other hand, Bora et al. (2016) show that the shortest oscillator periods of engineering interest (e.g. 0.01–0.1 s) are driven by motions with longer Fourier periods than those of the corresponding responsespectrum oscillator period. In fact, noise present in the time series at very short periods, such as 0.01-0.03 s does not necessarily affect the corresponding response spectrum at all, even in that specific period range. For instance, PSA at T = 0.01 s (often assumed equivalent to PGA) is typically related to ground motions at 20-30 Hz or lower, and nowhere near the 100 Hz implied by the reciprocal of the oscillator period. This was explored in detail by Douglas and Boore (2011), who concluded, through simulations of moderate to large events, that contaminating records with high-frequency noise had a negligible impact on their response spectra. This is frequently used as



justification to ignore high-frequency noise and T_{\min} when dealing with PSA.

Douglas and Boore (2011) investigated records for earthquakes typically found in strong ground motion datasets (M > 4.5) and, furthermore, made use of site conditions representative of such records, typically soil or stiff-soil sites. Both the moderate to high magnitude of events and the relatively high damping result in records with Fourier spectral amplitudes naturally lacking in high-frequency content. The records used by Douglas and Boore (2011) therefore exhibited low source corner frequencies $(f_0 < 1 \text{ Hz})$ and strong exponential decay at high frequency due to damping. For application to smaller events (weak-motion data), with higher source corner frequencies, or to records from rock or hard rock sites with low damping, we must consider signals with very different spectral content to those investigated by Douglas and Boore (2011). As noted previously, in such cases, the degree to which short oscillator period PSA is driven by longer period motions reduces and the conclusions of Douglas and Boore (2011) may, therefore, not be transferable to weak-motion data.

It is the aim of this study to investigate the impact of high-frequency noise on the response spectrum and propose a robust workflow for defining the usable bandwidth of both FAS and PSA from weak-motion records. For clarity, throughout this manuscript, we refer to PSA in terms of oscillator period, T, and FAS in terms of signal frequency, f, as per convention. A parametric model for the lowest usable (uncontaminated) period, T_{\min} , is initially developed using simulations that account for the influence of Fourier spectral shape on the influence of noise in a record's 5% damped response spectrum. The model is then used in direct application to investigate the impact of high-frequency noise on response spectral ordinates from weak-motion records of induced seismicity in the Groningen Gas Field, the Netherlands. The database consists of 803 triaxial recordings from events between 2006 and 2020 with local magnitudes ranging from M_L 2.5 to M_L 3.6 and is being used in the framework of the Groningen gas field hazard and risk analyses (van Elk et al. 2017) to develop a GMM (Bommer et al. 2017). The recordings are from highquality digital accelerographs at 98 sites belonging to the B- and G-networks of the Royal Netherlands Meteorological Institute (KNMI; see Ntinalexis et al. 2019; Dost et al. 2017; KNMI 1993). By virtue of the small magnitude of the events, the recordings contain small-amplitude motions, with as-recorded horizontal PGA values ranging from 0.068 cm/s^2 ($7 \times 10^{-5} \text{g}$) to 108.68 cm/s^2 (0.11 g).

2 Noise and its impact on FAS and PSA

The assessment of noise and its impact in the Fourier domain is relatively straight-forward. Typically, preevent noise samples are taken and compared with the record in the Fourier domain. It is important to account for differences in signal duration when sampling the time-history for noise. Authors sometimes ensure that both the earthquake time-history and the noise timehistory are of equal length, but this is not always possible. In this case, noise FAS should be scaled by the square root of the ratio of duration between the earthquake and noise time-history (after Parseval's theorem) to provide FAS amplitudes that are consistent (i.e. corresponding to equivalent signal lengths). Due to the characteristic 'trapezoidal' shape of the earthquake acceleration spectrum, it stands out over the broadly flat noise floor within the passband that can be considered acceptable (Fig. 1). Lower (f_1) and upper (f_{11}) usable frequency limits are therefore clearly identifiable.

As noted earlier, for the response spectrum (PSA), the correlation with FAS amplitudes at corresponding oscillator periods decreases dramatically as signal frequencies increase above the record's apparent corner frequency (roughly the peak of the FAS) (Bora et al. 2016). While at long periods we can therefore assume a correspondence of the minimum usable frequency of the record's FAS (f_1) and the maximum usable period of the record's response spectrum (i.e. $T_{\rm max} = [0.7 \text{ to } 0.97]/f_1$, after Akkar and Bommer 2006), at high signal frequencies and short oscillator periods, we cannot make this assumption.

In order to assess the impact of high-frequency noise on PSA, we initially work with synthetic data. This allows us an unambiguous definition of the true signal amplitude and corresponding uncontaminated response spectrum, which is not afforded with real data. Time-domain stochastic simulations have been performed using EXSIM (Motazedian and Atkinson 2005) as modified by Boore (2009). Simulations have been performed using the GMMs for:



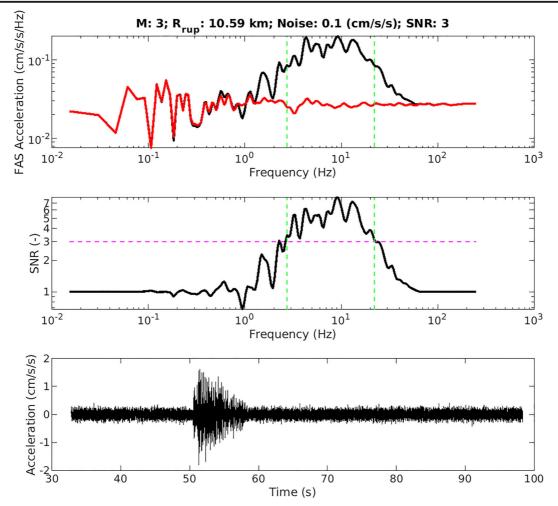


Fig. 1 Example simulation for the Groningen GMM (with κ_0 = 0.01 s). Top: FAS of signal (*black*) and noise (*red*). Middle: SNR (*black*) and threshold (*magenta*). Green-dashed lines show the

lower and upper usable frequencies. Bottom: the corresponding synthetic time-history

- i) the Groningen gas field at a buried rock horizon $(V_{s30} = 1400 \text{ m/s})$, as detailed in Edwards et al. (2019), and
- ii) Eastern North America (ENA, Atkinson and Boore 2006).

For the Groningen simulations, we investigate the impact of damping by varying the simulation parameter κ_0 (Anderson and Hough 1984), using κ_0 = 0.01, 0.03 and 0.05 s (roughly equivalent to damping expected at competent rock outcrops through to low V_s soil site conditions). The ENA GMM specifically allows the investigation of weakly damped motions, with a very hard-rock site condition (κ_0 = 0.005 s) implicit in the GMM. In both cases, the GMMs are calibrated against

local empirical data in the magnitude range of interest for this study (approx. $M_L < 4$), and the simulations can therefore be considered to be representative, yet diverse, in terms of amplitude and frequency content, of real earthquake records.

Noise-free acceleration time series from earthquakes with moment magnitudes from 1.0 to 6.0 (in 0.5 unit increments) are simulated at 20 log-spaced Joyner-Boore distances from 0.1 to 60 km (Fig. 1). Noise is subsequently applied to the simulations in increasing amplitude until the signals are completely lost. We use two noise forms: (i) white noise and (ii) the noise model of Cauzzi and Clinton (2013). The white noise is generated in the time domain and defined by a normal distribution with zero mean and standard deviation:



0.01, 0.1, 1, 10, 100 cm/s². The higher values are not intended to reflect typical noise levels, but to ensure that all records are affected by noise. The model of Cauzzi and Clinton (2013) is used to consider a realistic highnoise scenario, using their high-noise power-spectrum model converted from dB/Hz to absolute units of spectral acceleration. We scale those amplitudes incrementally by factors 0.1, 0.2, 0.5, 0.75, and 1.0, to generate a realistic stochastic-phase noise time-history that is added to the noise-free simulation in the time domain. With this approach, we retain acceleration time-histories for the noise-free simulation, the background noise and the contaminated 'noisy' simulation, with the latter referred in the following as the 'synthetic' time series.

From each of the synthetic time series' FAS, various measures are determined to allow investigation of the impact of the noise level:

 f_u : the upper usable FAS frequency in Hz (defined by a signal to noise ratio of 3);

 f_{peak} : the frequency at the FAS peak;

 A_{peak} : the natural logarithm FAS amplitude at f_{peak} ; A_u : the natural logarithm FAS amplitude at f_u ;

 ΔA : the amplitude difference, $A_{\text{peak}} - A_{\text{u}}$;

 Δf : the frequency difference, $f_{\rm u} - f_{\rm peak}$.

We measure T_{\min} by determining the lowest period at which the 5% damped response spectrum of the acceleration time-history is unaffected by noise. This is defined as where the response spectrum of the synthetic time-history is within a 5% tolerance of the true value (Fig. 2). This is a conservative estimate, as PSA at shorter periods than the subsequently defined T_{\min} may return to within the defined threshold. Our observations show that while this is often the case. PSA then tends to fluctuate within and outside the acceptable tolerance level at periods below T_{\min} (Fig. 2). Low-pass frequency filtering of the time series at, or around, f_u results in severe impact on the PSA for weak-motion data, and, as such, should not be used. It is clear that unfiltered (or high-pass frequency filtered: band-pass $f > f_1$) time series allow calculation of PSA to periods well below $1/f_u$. In fact, at the 5% tolerance level, the PSA from both these cases in Fig. 2 (from the simulation shown in Fig. 1) only just fail, with most PSA amplitudes being within ~ 10% of the true values. On the other hand, the low-pass frequency filtered time series result in up to 50 % underestimation of PSA.

Figure 3 shows the suite of measured $f_{\rm u}$ and $T_{\rm min}$ from simulations using the Groningen GMMs with alternative damping (κ_0) , along with the ENA GMM. Clearly, as expected, for signals with higher $f_{\rm u}$, we obtain PSA with smaller $T_{\rm min}$. However, the spectral shape has a significant impact on the usability of the response spectra: low κ_0 (weakly damped) records require much higher values of $f_{\rm u}$ to maintain usability of PSA down to 0.01 s. In terms of estimating T_{\min} from the time series, we therefore require knowledge of not only the usable FAS bandwidth (i.e. f_u , which can be directly measured from field records), but also the spectral shape. In the following, we therefore develop a model for determining T_{\min} that accounts for spectral shape by using easily measured characteristics of a waveform and its FAS.

First, we define an adjusted upper usable FAS frequency (f_u^*) in Hz. The adjustment normalises f_u to that expected, given the same noise and peak-signal amplitude, for a signal with a reference damping, defined by $\kappa_{\rm ref} = 0.03$ s. It therefore results in a predictor for $T_{\rm min}$ that is unbiased. f_u^* is given by:

$$\begin{split} f_{\rm u}^* &= \\ f_{\rm u.max} \bigg[0.4, \exp \bigg\{ f_{\rm u}. \big[-0.25 \log_{\rm e} \big(\kappa_{\rm ref} + 0.005 \big) - 0.17 \big] \bigg(\frac{\Delta A}{\pi \Delta f} - \big(\kappa_{\rm ref} + 0.005 \big) \bigg) \bigg\} \bigg] \end{split}$$

Figure 4 shows the resulting f_u^* for the three alternative Groningen GMM simulations (κ_0 = 0.01, 0.03 and 0.05 s). Note that f_u^* values for the κ_0 = 0.03 s simulations (the same as our selected reference, $\kappa_{\rm ref}$) are broadly consistent with measured f_u . For the κ_0 = 0.01 s simulations, f_u^* are systematically reduced with respect to f_u , while for the strongly damped κ_0 = 0.05 s simulations, f_u^* are higher than measured f_u . Note that we define a maximum adjustment factor, 0.4, in Eq. 1 based on trial and error after observing over-correction of very weakly damped (low κ_0) signals.

Figure 5 compares the original $f_{\rm u}$ (as Fig. 3) and adjusted $f_{\rm u}^*$ against $T_{\rm min}$. The use of $f_{\rm u}^*$ clearly removes the dependence of spectral shape on the correlation. Based on $f_{\rm u}^*$ and $T_{\rm min}$ for the Groningen GMM simulations (Fig. 5), a best-estimate lower usable period, $\overline{T_{\rm min}}$ (in seconds), within an acceptable tolerance is defined by:



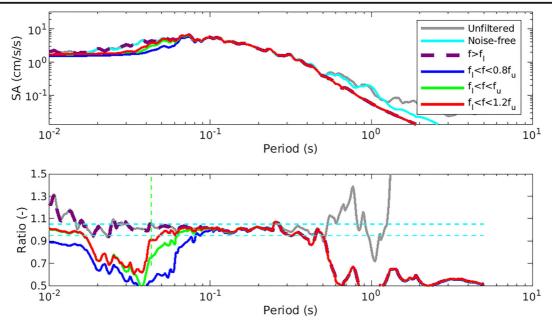


Fig. 2 (*Top*) 5% damped response spectra corresponding to the unfiltered synthetic and filtered data (with legend indicating the frequency band-pass), in addition to PSA from the noise-free time series (the target) in Fig. 1. (*Bottom*) Ratios of PSA (as *top*) with

respect to the noise-free target. $T_{\rm min}$ (dashed vertical line) is defined as the period at which the unfiltered PSA first deviates (with decreasing period) from the $\pm\,5\%$ tolerance level (horizontal dashed lines).

$$\overline{T_{min}} = \max(0.01, exp(a_2 + a_1log_e(f_u^*))) \quad f_u^* < a_3$$

$$\overline{T_{min}} = 0.01 \quad f_u^* \ge a_3$$
(2)

$$T_{min} = exp\left(a_2 + a_1 log_e\left(\frac{f_u^*}{c^n}\right)\right) \qquad \frac{f_u^*}{c^n} < a_3$$

$$T_{min} = 0.01 \qquad \qquad \frac{f_u^*}{c^n} \ge a_3 \qquad (3)$$

where a_3 is the log-mean $f_{\rm u}^*$ for $0.01 < T_{\rm min} < 0.02$ s and a_{1-2} are determined through log-linear regression of $f_{\rm u}^*$ versus $T_{\rm min}$. Bounds on $T_{\rm min}$ are then given by introducing a scaling factor, c, on $f_{\rm u}^*$:

 $\begin{array}{c} 10 \\ & \kappa_0 = 0.005 \text{ s (ENA)} \\ & \kappa_0 = 0.01 \text{ s} \\ & \kappa_0 = 0.03 \text{ s} \\ & \kappa_0 = 0.05 \text{ s} \\ \end{array}$

Fig. 3 Plot of measured T_{\min} versus $f_{\rm u}$ for simulations from the three Groningen and ENA GMMs

f, (Hz)

with the factors c=1.113 (upper bound) and c=1/1.113 (lower bound) designed to encapsulate the data (where $0.01 < T_{\min} < 0.02$ s) at n standard deviations of $\log_e(f_u^*)$. Based on an average over three Groningen GMM simulation scenarios with 1100 simulations in each, and using only data where $T_{\min} > 0.01$ s, we determine $a_1 = -1.753$, $a_2 = 1.946$, and $a_3 = 25.41$ Hz (Fig. 5c). In addition, a maximum threshold of $T_{\min} = 0.1$ s is imposed beyond which it is not possible to reliably estimate T_{\min} from f_u^* . Predicted values of T_{\min} are therefore deemed unresolved if they exceed 0.1 s.

In order to test if there is any sensitivity of the model to the selected magnitude-distance range, the data are split into subsets with magnitude 1–4.5 and 4.5–6 and distances 0–30 and 30–60 km. The model was found to be equally applicable to all of the data subsets. An example is shown for $\mathbf{M}=4.5$ –6 at all distances in Fig. 6. These simulations were subject to unrealistically high levels of noise in order to obtain relatively low f_u for such large events. Interestingly, the model appears



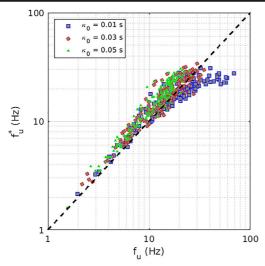


Fig. 4 Comparison of measured ($f_{\rm u}$) and adjusted upper FAS frequencies $f_{\rm u}^*$ according to Eq. 1

equally valid for these very noisy records of larger events, in addition to weak-motion data. While not routinely useful for strong-motion datasets (since noise levels rarely reach such amplitudes), the model would be useful for cases where significant anthropogenic noise levels are present, such as those used for earthquake early warning in industrial settings (Cauzzi et al. 2016). As further validation exercise, we apply the model to the ENA dataset (Fig. 7), which has so far been withheld from the model development. Disregarding the very noisy data with $T_{\rm min} > 0.1$ (which, as noted previously, shows very little correlation with

 $f_{\rm u}$), the consistency with results from the Groningen simulations is very good, and the model for $T_{\rm min}$ is clearly suitable independent of the region.

Our choice of 5% tolerance for selecting observed T_{\min} will clearly have an impact on the results discussed previously: allowing a larger tolerance when measuring T_{\min} means that lower $f_{\rm u}$ are required (for a given T_{\min}). In order to facilitate choice when implementing the T_{\min} model, we have also calibrated coefficients for Eqs. 2 and 3 using alternative tolerances of 10 and 15 % (Table 1, Fig. 8). An alternative to white noise was also explored by implementing the high noise model of Cauzzi and Clinton (2013), which is somewhat more forgiving in the mid-period range than white noise. Here the noise is more realistic, but the larger events, particularly those simulated at near distances, are unaffected by the noise and are therefore not included in the derivation of the alternative model (since $T_{\min} = 0.01 \ s$ for those records). Using the high noise model, a_3 , defining the frequency $f_{\rm u}^*$ above which $T_{\rm min}$ = 0.01 s, is almost unchanged, being instead strongly related to the acceptable tolerance within the true PSA. For tolerance values of 5, 10 and 15%, we observe a_3 values of 24.4–25.4, 19.3–20.3 and 17.0–17.1 Hz, respectively: f_{ij}^* above which results in $T_{\min} = 0.01$ s. The shape of the T_{\min} versus f_n^* slope does change depending on the noise model used, however. This suggests that the shape of the noise spectrum itself, as well as the earthquake time series, has an impact on the usability of PSA.

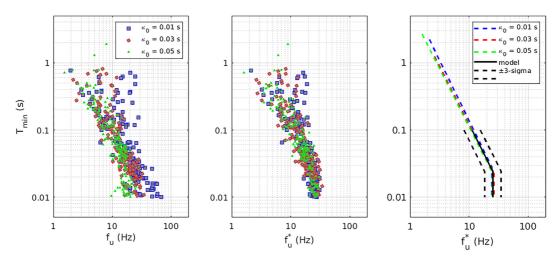


Fig. 5 T_{\min} versus (a) f_{u} for the Groningen GMM simulations, (b) versus f_{u}^{*} (Equation 1), and (c) showing the linear best-fit (using f_{u}^{*}) to the individual simulation datasets, along with the best-

estimate T_{\min} model (Eq. 2) and ± 3 -sigma (Eq. 3, n=3). Predicted T_{\min} (black lines) is assumed unresolved if greater than 0.1 s and not plotted



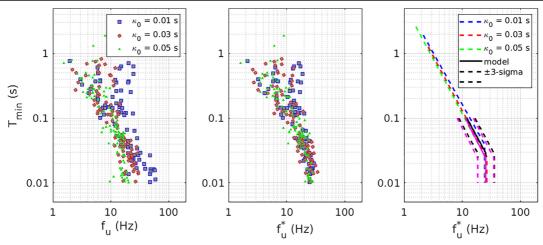


Fig. 6 T_{\min} versus (a) $f_{\rm u}$ for the Groningen GMM simulations for the 3 damping levels with $\mathbf{M}=4.5-6$, (b) versus $f_{\rm u}^*$, and (c) showing the linear best-fit to the individual simulation datasets along with the model (Eqs. 2, 3, n=3). The black lines show the

model for the whole dataset, as Fig. 5, and almost completely overlap the model that would result from using only $\mathbf{M}=4.5$ –6 (*magenta*). Note that predicted $T_{\rm min}$ is assumed unresolved if greater than 0.1 s

3 Workflow: usable frequency range of FAS

In the following sections, we detail the application of a workflow used to define usable frequency (for FAS) and period (for PSA) for an induced seismicity dataset, specifically, a database of 803 recordings from the B- and G-networks of the KNMI in the Groningen region, the Netherlands. The recording networks and instrumentation used to record the acceleration time series are described in Ntinalexis et al. (2019). Prior to 2014, the monitoring network consisted of several GeoSig

digital accelerographs. As a consequence of a M_L3.6 earthquake that occurred in Huizinge on 16 August 2012, more detailed seismic studies were commissioned for the area. A significant upgrade and expansion of the existing network as well as the installation of new networks became part of this effort (Ntinalexis et al. 2019). The KNMI networks now consist of almost 100 modern Kinemetrics accelerometer stations with high rate 24-bit data-logging. We can therefore safely assume that the predominant source of signal contamination in the dataset analysed will be external noise. The

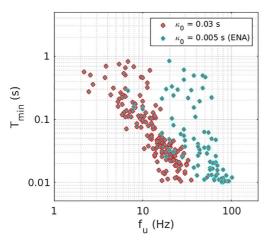
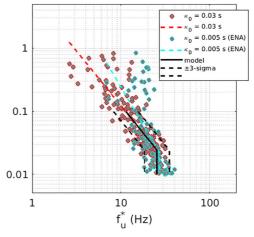


Fig. 7 T_{\min} versus (**a**) f_{u} for the κ_0 = 0.03 s Groningen GMM simulations (red) alongside the ENA data (cyan) and **b** versus f_{u}^* for the same datasets, dashed red and cyan lines show the best fit of



each. The black lines are the proposed model (as Fig. 5) for predicting T_{\min} calibrated using only the Groningen simulations.



Table 1 Coefficients of Eqs. 2 and 3 for various tolerances. *HNM* indicates the use of the high-noise model of Cauzzi and Clinton (2011); otherwise, the white noise model is used

PSA Tolerance	a_1	a_2	<i>a</i> ₃ (Hz)	c
Within 5%	-1.753	1.946	25.41	1.113
Within 5%, HNM	-1.714	1.608	24.40	1.158
Within 10%	-1.670	1.095	20.32	1.200
Within 10%, HNM	-1.733	1.211	19.30	1.182
Within 15%	-1.427	0.081	17.02	1.221
Within 15%, HNM	-1.692	0.807	17.16	1.223

recordings examined were obtained during induced events of local magnitudes ranging from $M_L 2.5$ to $M_L 3.6$ that occurred between 2006 and 2020 in Groningen. The as-recorded horizontal PGA values of the records range from $0.068~\text{cm/s}^2$ to $108.68~\text{cm/s}^2$ and were recorded at epicentral distances ranging from 0.4 to 34~km (Fig. 9).

3.1 Maximum usable frequency

As mentioned previously, the maximum usable FAS frequency, f_u , can be selected via a signal-to-noise ratio analysis. We choose to select f_u as the maximum frequency of the continuous frequency window with SNR above 3. This is the simplest method to select the

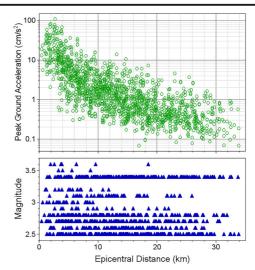


Fig. 9 Peak ground acceleration of the Groningen horizontal components plotted against distance (*upper*) and magnitude-distance distribution of the Groningen database (*lower*)

maximum usable frequency and is also widely employed in engineering and seismology. To conduct the SNR analysis, it is first necessary to obtain a noise model representative of the noise in the record. This is routinely determined as the FAS of the pre-event time series. In most modern recording networks, continuous data streams are available via online services and data portals, which allows the user to select a time window of their choice around the event. In these cases, it suffices for the user to select a time window with a long pre-event memory and select the first several seconds of that

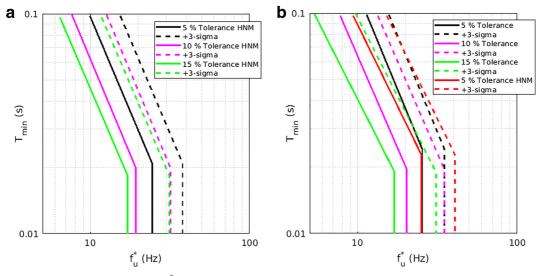
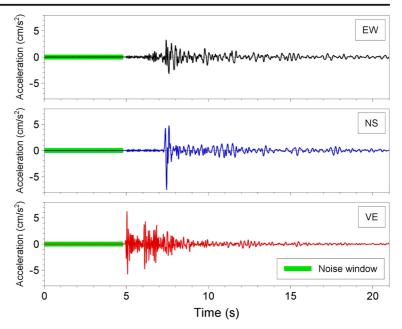


Fig. 8 Comparison of modelled T_{\min} versus f_{u}^{*} for various acceptable tolerances on PSA (see Table 1) using **a** the high-noise model (HNM) of Cauzzi and Clinton (2013) and **b** white noise



Fig. 10 Time-histories and selected noise windows of the MID1 recording of the $M_L3.2$ Westeremden event on 30 October 2008



window to sample noise adequately. However, in networks operating on a triggering-only basis, such as the KNMI B-network in Groningen prior to 2014 (see Ntinalexis et al. 2019), limited time lengths of the preevent memory may be available. In small-amplitude records such as those included in the Groningen database, the SNR at frequencies above 20 Hz can also be very sensitive to the selection of the noise window due to transient signals, and hence it is important to make sure that the noise window is carefully selected.

A technique of dynamic noise window selection is employed in our analyses. We use the vertical component motion to determine the noise window to ensure we avoid P wave energy in the selected analysis window. While small in amplitude on the horizontal components,

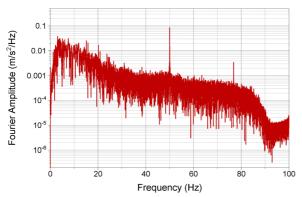


Fig. 11 Example of the presence and amplitude of mains electricity noise in a record's FAS

the P wave has non-negligible high-frequency energy that may bias the noise estimate (and therefore f_u). We begin by locating the time window from the beginning of the record to the point where the Arias Intensity is 0.5% of the total. We then determine short (-1 to +0.5 s) and long-term (-3 to +0.5 s) moving averages (STA and LTA, respectively) and compute the ratio (STA/LTA). A ratio above 1.2 signifies a significant amplitude change that can be associated with the first observable arrivals of the earthquake signal. We choose the end of the noise window to be the earliest of either the 0.5% Arias Intensity or the STA/LTA trigger (assumed to be

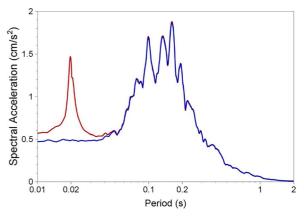


Fig. 12 Response spectra of record G040 from the M_L3.1 Hellum earthquake in Groningen (30 September 2015), removing (*blue*) and without removing (*red*) the 50-Hz noise peak



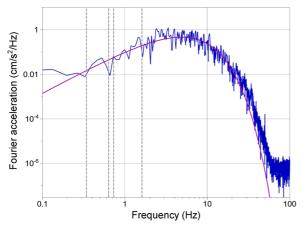


Fig. 13 Fourier amplitude spectrum of acceleration for the North-South component of recording KANT from the $M_L3.2$ Garrelsweer earthquake of 27 June 2011 (*blue*) the fitted Brune model (*purple*) and proposed frequencies for the selection of f_l (as Fig. 14)

the P wave). The noise window, as defined on the vertical component, is then used for the horizontal components. An example is shown in Fig. 10.

For the Groningen dataset, we found that often the G-station sensors were located close to the electricity mains network. In this case, it was very likely that the record was contaminated with 50-Hz noise. For small-amplitude records, this may result in a significant peak in the FAS (Fig. 11) and affect the calculation of f_u , as well as the response spectra of the record (Fig. 12). Douglas and Boore (2011) recommend the removal of this peak at 50 Hz with a narrow notch filter and in our case, we found it absolutely necessary to remove the 50-Hz noise in order to obtain correct estimates of short-period PSA (Fig. 12).

3.2 Minimum usable frequency

Determining the lower usable frequency (f_l) by employing the same SNR>3 criterion as used for f_u is a choice that is often employed. However, because the SNRs of small-amplitude records are smaller and the resulting bandwidth can be very limited, it is desirable in our case to use a method that results to more forgiving estimates of f_l . The first step is to obtain an initial estimate of f_l . This is defined as the first point (with decreasing frequency) that the linear trend of the recording's FAS is observed to systematically decay more slowly than a theoretical Brune (1970) spectrum. The next step is to low-cut filter the record using f_l as the

filter corner-frequency and then compute the displacement trace through double integration of the acceleration time series. The filter used is an 8th order acausal Butterworth filter. Any low-frequency noise can then easily be observed in the time domain. If the total displacement is zero and long-period noise cannot be readily observed in the displacement trace, then the initial estimate is selected as the final f_i value. If the user judges the displacement trace to still be unacceptably contaminated with noise, a higher frequency is selected, and the process is iterated until f_i is found that results in a noise-free displacement time series.

An example of the application of the iterative selection of f_l is shown in Figs. 13 and 14. Figure 13 shows the FAS of the North-South component of recording KANT from the M_L3.2 Garrelsweer earthquake of 27 June 2011. The identification of different possible low-cut filter frequencies from the FAS of the record is illustrated. The displacement traces obtained after the application of the different filters are compared in Fig. 14. It is obvious that applying a filter of 0.342 Hz (the initial estimate based on spectral shape) is insufficient, as long-period waves are still clearly observable in the displacement trace (Fig. 14). After iterating through increased values of f_l , we observe that a frequency of 1.611 Hz is excessive as it results in a reduction in the amplitude of the record. Low-cut frequencies of 0.635 Hz and 0.732 Hz both produce acceptable results; hence, the lowest, 0.635 Hz, is selected.

3.3 Removal criteria

Figure 15 shows ratios of the PGV and PSA of noisecontaminated synthetic recordings to the noise-free versions using the Groningen GMM with $\kappa_0 = 0.03$ s. The ratios are plotted as a function of the maximum usable frequency (f_u) . It is immediately apparent that, when f_u is low, PGV and the short-period spectral ordinates have significantly increased amplitudes. We therefore recommend that records with f_u below 15 Hz should not be used at all and should be discarded from ground-motion databases. As shown in Fig. 14, low-cut filtering with an excessively high cutoff frequency can result in a reduction in amplitude and should be avoided. Therefore, when f_t is identified above 2 Hz, we also consider a record to also be unusable. When either horizontal component fulfils at least one of these removal criteria, we discard the entire triaxial recording, as



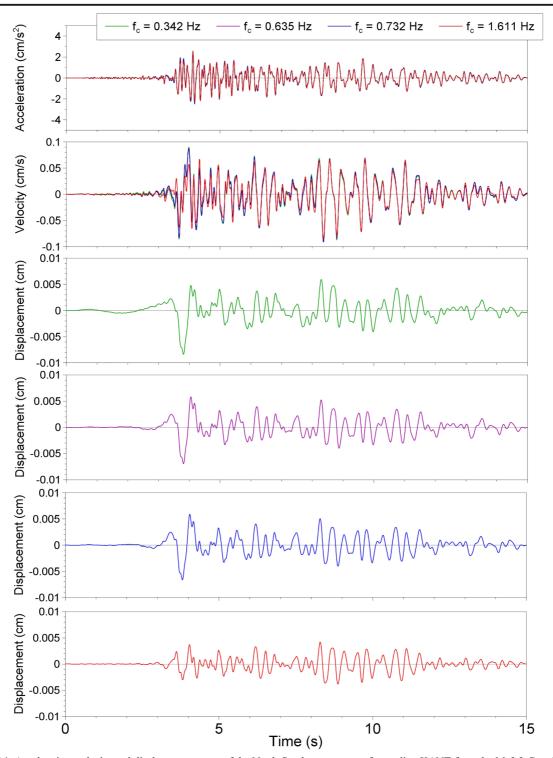


Fig. 14 Acceleration, velocity and displacement traces of the North-South component of recording KANT from the $M_L3.2$ Garrelsweer earthquake of 27 June 2011 after the application of different low-cut filters. The displacement traces are shown in individual panels



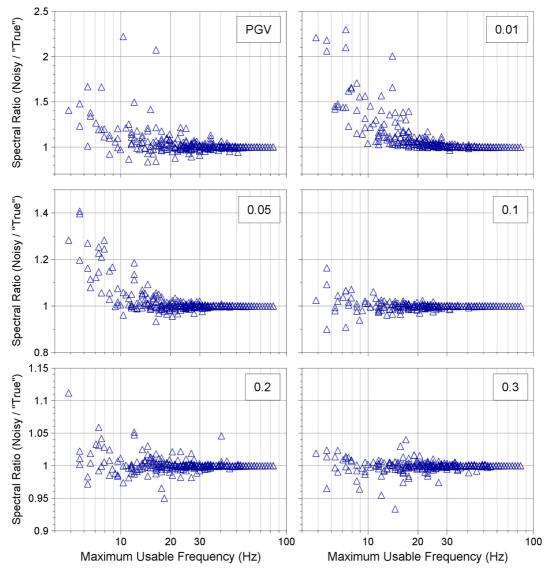


Fig. 15 Ratio of the PGV and PSA of noise-contaminated synthetic recordings (using the Groningen GMM with $\kappa_0 = 0.03$ s) to the noise-free versions

both horizontal components are required to compute the intensity measures commonly predicted by GMPEs and GMMs.

Figure 16 illustrates which records of the database were removed entirely by applying the constraints on f_l and f_u . A total of 96 out of the 800 records (12%) from the Groningen database were removed. As expected, these recordings correspond to the relatively weaker motions within the database, which come from the lower end of the magnitude range and stations at longer epicentral distances (Fig. 16).

4 Workflow: usable period range of PSA

4.1 Maximum usable period

Once the usable bandwidth of the FAS is defined, the next step is to low-cut filter the records. We recommend the use of an 8th order acausal (zero phase) Butterworth filter, which has been found to be more suitable for use on digital records (Boore and Akkar 2003). For the correct use of this type of filter, it is necessary to zero-pad both ends of the record (Boore and Bommer 2005).



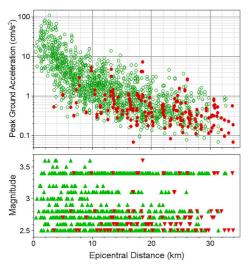


Fig. 16 Peak ground acceleration of the Groningen as-recorded horizontal components plotted against distance (*upper*) and magnitude-distance distribution of the Groningen database (*low-er*). Unusable recordings are shown in *red*

The pad length is a calculated using the function of Converse and Brady (1992) which is dependent on both the chosen filter corner frequency and the order of the filter. We apply the same filter to both horizontal components, using the lowest cutoff, f_l , of the two components, as they are typically used in conjunction when calculating intensity parameters for use in GMPE/GMM development.

The amplitudes of long-period spectral ordinates are highly sensitive to the application of low-cut filters. As the filter removes both signal and noise, an unknown combination of both is left behind by the filter at frequencies lower and close to the cutoff frequency. Therefore, the response spectra are reliable for use only up to a certain period, lower than the long-period cutoff (T_c , the inverse of the cutoff frequency, f_u). Different studies have employed schemes to define this usable period limit. Some examples are described in Boore and Bommer (2005) and Akkar and Bommer (2006). The most widely employed technique—and the one adopted in this study—to define the usable period limit is to identify the ratio $T_c/T_{\rm max}$.

According to Akkar and Bommer (2006), for digital records from soft soil sites such as those in Groningen, this ratio is between 0.7 and 0.97. The method we adopted to select from this range consists of comparing the PSA before and after filtering and only using the spectral ordinates where the change in amplitude is

within a certain threshold. For the Groningen data, we selected this threshold to be 5%. Figure 17 shows ratios of PSA post- to pre-filtering, plotted as a function of the ratio of each period to the cutoff period. In this case, it can be observed that more than 95% of the response spectra have changed by less than 5% up to a period of 70% of the cutoff period. Hence, we selected the ratio of 0.7 and define the maximum usable period for each record as $T_{\rm max} = 0.7T_{\rm c} = 0.7/f_{\rm l}$. It must be noted that, for databases with a small number of available records, it may be preferable to define a larger ratio to maximise the available data, using a more generous threshold.

4.2 Minimum usable period

Filtering high frequencies prior to computing PSA is not recommended as it may have a knock-off effect on a wide range of periods (see Fig. 2). However, as shown earlier, it is still necessary to define a minimum usable period in order to exclude noise-contaminated PSA from use. The first estimate of T_{\min} is the result of the upper-bound T_{\min} model presented earlier at n=3 (Eq. 3), which we apply for a threshold 5% using the white noise model (Table 1).

In addition to the parametric T_{\min} model, we devise additional measures to constrain T_{\min} . We create two hybrid-synthetic records using the FAS of each record under analysis. To create the first synthetic, we fit an idealised Brune (1970) spectrum to the FAS of the record (Fig. 18), and use the FAS of the record within its usable frequency range (f_1 to f_u) and the Brune spec-

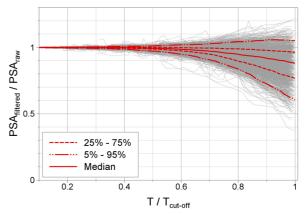
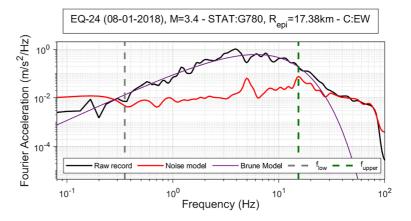


Fig. 17 Ratios of PSA post- to pre-filtering, plotted against a ratio of period to cutoff period



Fig. 18 FAS of record G780 from the M_L3.4 Zeerijp event of 8 January 2018 in Groningen



trum in the unusable frequencies. Thus, we create an idealised 'noise-free' version of the record when performing an inverse Fourier transform. To create the second synthetic, we use the full FAS of the record but double it for frequencies higher than $f_{\rm u}$. In this way, we obtain a noisier version of the same record.

By comparing the response spectrum of the original record to the idealised 'noise-free' version, we obtain an estimate of the periods that are affected by noise. At the same time, by comparing the original response spectrum with the 'noisier' version, we can observe which periods are sensitive to additional noise. From these comparisons, we can define two additional estimates of $T_{\rm min}$, based on the divergence (with 5 % tolerance) of the hybrid-synthetic and the original response spectra. Finally, we select $T_{\rm min}$ using the following logic (Fig. 19):

- If the parametric T_{\min} model is 0.01 s (the shortest period defined), we retain that value.
- If two of the three T_{min} estimates are within 10% of one another, we retain the average value of those T_{min}.
- Otherwise, we select the result of the parametric T_{\min} model, but restrict T_{\min} between the values calculated using the two hybrid-synthetics.

The number of usable PSA, as defined by $T_{\rm min}$ and $T_{\rm max}$, is shown in Fig. 20 over 13 approximately linearly spaced periods from 0.01 to 1.5 s. The largest quantity of usable spectral accelerations correspond to the intermediate periods $(0.1-0.7~{\rm s})$, a smaller number (498) is available at 0.01 s and a rapid decay

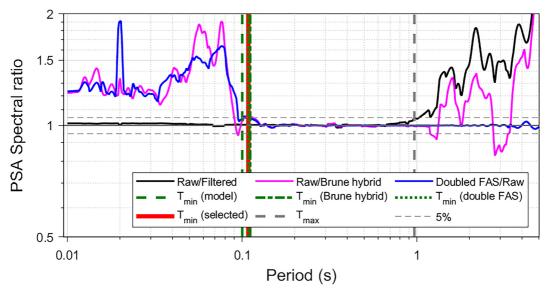


Fig. 19 Spectral ratios and selection of T_{\min} for the H2 component of record G310 of the 2 May 2020 M_L2.5 Zijldijk earthquake in Groningen



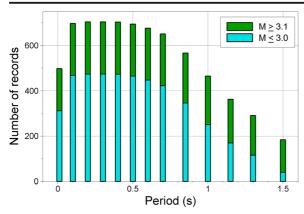
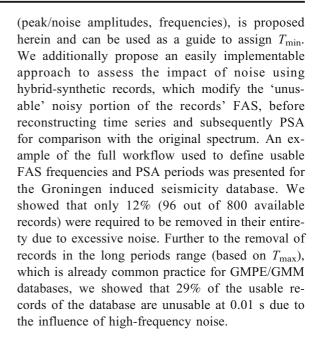


Fig. 20 Number of usable records as a function of oscillator period, showing the total number and those corresponding to different earthquake magnitude ranges

can be observed with increasing period from 0.85 s onward. At 1.5 s, the number of usable spectral accelerations is 184, which can still be considered sufficient for the limited distance ($R_{\rm epi}$ < 35 km) and magnitude range covered by the database. In total, 206 records (29.2% of the 704 usable records) are unusable at 0.01 s due to noise.

5 Conclusions

Short-period noise in acceleration time series has the potential to influence response spectral accelerations at short oscillator periods. This has previously been investigated by Douglas and Boore (2011) in the context of data found in typical in strong-motion datasets. Analysis of 'strong-motion' data, however, generally avoids the influence of high-frequency noise. This is both due the relative amplitude of signal and noise, and also due to the fact that the dominant frequency of motion of strong-motion data is much lower than any high-frequency noise. Our simulations show that PSA from noisy weak-motion records, as present in many ground motion databases such as those for induced seismicity, is susceptible to high-frequency noise. This is particularly so for weakly damped records, such as those on 'hardrock' sites. The impact of high-frequency noise on PSA should be considered by assigning record specific T_{\min} and without any form of low-pass frequency filtering. A parametric T_{\min} model, based on easily measurable properties of waveform FAS



Acknowledgements We thank Nederlandse Aardolie Maatschappij (NAM) for funding this study. The authors are also grateful for the funding from EU project New challenges for Urban Engineering Seismology (URBASIS) (Grant Number 813137), which has supported this work. We also thank KNMI for providing access to the data used during this study (KNMI 1993).

Availability of data and material Recorded waveform data used in this study are available from the Royal Netherlands Meteorological Institute (KNMI) on request, and online at http://rdsa.knmi.nl/dataportal/ (last accessed January 2021).

Code availability The software EXSIM used for this study is available on request from the author, and online at http://www.daveboore.com/software online.html (last accessed January 2021).

Funding This work has been funded by the Nederlandse Aardolie Maatschappij (NAM).

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative



Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Akkar S, Bommer JJ (2006) Influence of long-period filter cut-off on elastic spectral displacements. Earthq Eng Struct Dyn 35(9):1145–1165
- Anderson JG, Hough SE (1984) A model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies. Bull Seismol Soc Am 74(5):1969–1993
- Atkinson GM, Boore DM (2006) Earthquake ground-motion prediction equations for eastern North America. Bull Seismol Soc Am 96(6):2181–2205
- Bommer JJ, Dost B, Edwards B, Stafford PJ, van Elk J, Doornhof D, Ntinalexis M (2016) Developing an application-specific ground-motion model for induced seismicity. Bull Seismol Soc Am 106(1):158–173
- Bommer JJ, Dost B, Edwards B, Kruiver PP, Ntinalexis M, Rodriguez-Marek A, Stafford PJ, van Elk J (2017) Developing a model for the prediction of ground motions due to earthquakes in the Groningen gas field. Neth J Geosci 96(5):203–213
- Boore DM (2009) Comparing stochastic point-source and finitesource ground-motion simulations: SMSIM and EXSIM. Bull Seismol Soc Am 99(6):3202–3216
- Boore DM, Akkar S (2003) Effect of causal and acausal filters on elastic and inelastic response spectra. Earthq Eng Struct Dyn 32(11):1729–1748
- Boore DM, Bommer JJ (2005) Processing strong-motion accelerograms: needs, options and consequences. Soil Dyn Earthq Eng 25(2):93–115
- Bora SS, Scherbaum F, Kuehn N, Stafford P (2016) On the relationship between Fourier and response spectra: implications for the adjustment of empirical ground-motion prediction equations (GMPEs). Bull Seismol Soc Am 106(3): 1235–1253
- Brune JN (1970) Tectonic stress and the spectra of seismic shear waves from earthquakes. J Geophys Res 75(26):4997–5009
- Cauzzi C, Clinton J (2013) A high-and low-noise model for highquality strong-motion accelerometer stations. Earthquake Spectra 29(1):85–102
- Cauzzi C, Behr Y, Le Guenan T, Douglas J, Auclair S, Woessner J, Clinton J, Wiemer S (2016) Earthquake early warning and

- operational earthquake forecasting as real-time hazard information to mitigate seismic risk at nuclear facilities. Bull Earthq Eng 14(9):2495–2512
- Converse A, Brady AG, (1992). BAP: basic strong-motion accelerogram processing software, version 1.0. US Department of the Interior, US Geological Survey.
- Dost B, Ruigrok E, Spetzler J (2017) Development of seismicity and probabilistic hazard assessment for the Groningen gas field. Neth J Geosci 96(5):235–245
- Douglas J, Boore DM (2011) High-frequency filtering of strongmotion records. Bull Earthq Eng 9(2):395–409
- Douglas J, Edwards B (2016) Recent and future developments in earthquake ground motion estimation. Earth Sci Rev 160: 203–219
- Edwards B, Zurek B, Van Dedem E, Stafford PJ, Oates S, Van Elk J, DeMartin B, Bommer JJ (2019) Simulations for the development of a ground motion model for induced seismicity in the Groningen gas field, The Netherlands. Bull Earthq Eng 17(8):4441–4456
- Edwards B, Crowley H, Pinho R, Bommer JJ (2021) Seismic hazard and risk due to induced earthquakes at a shale gas site. Bulletin of the Seismological Society of America. https://doi.org/10.1785/0120200234.
- KNMI (1993) Netherlands Seismic and Acoustic Network. Royal Netherlands Meteorological Institute (KNMI). https://doi. org/10.21944/e970fd34-23b9-3411-b366-e4f72877d2c5.
- Motazedian D, Atkinson GM (2005) Stochastic finite-fault modeling based on a dynamic corner frequency. Bull Seismol Soc Am 95(3):995–1010
- Novakovic M, Atkinson GM, Assatourians K (2018) Empirically calibrated ground-motion prediction equation for Oklahoma. Bull Seismol Soc Am 108(5A):2444–2461
- Ntinalexis M, Bommer JJ, Ruigrok E, Edwards B, Pinho R, Dost B, Correia AA, Uilenreef J, Stafford PJ, van Elk J (2019) Ground-motion networks in the Groningen field: usability and consistency of surface recordings. J Seismol 23(6):1233–1253
- Peterson J (1993) Observations and modeling of background seismic noise, open-file report 93-322, U. S. Geological Survey, Albuquerque, NM, pp 42.
- van Elk J, Doornhof D, Bommer JJ, Bourne SJ, Oates SJ, Pinho R, Crowley H (2017) Hazard and risk assessments for induced seismicity in Groningen. Neth J Geosci 96(5):259–269
- Webb SC (1992) The equilibrium oceanic microseism spectrum. The Journal of the Acoustical Society of America 92(4): 2141–2158

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

