Robust Bayesian estimator for $S$-wave spectra, using a combined empirical Green’s function

W. Törnman $^{\spadesuit,1,2}$, J. Martinsson $^{\spadesuit,1,2}$ and S. Dineva $^{\spadesuit,1,2}$

$^1$Luossavaara-Kiirunavaara AB (LKAB), 98186 Kiruna, Sweden. E-mail: wille.tornman@ltu.se
$^2$Luleå University of Technology (LTU), 97187 Luleå, Sweden

Received 2021 April 12; in original form 2020 August 11

**SUMMARY**

We propose a new fully automatic and robust Bayesian method to estimate precise and reliable model parameters describing the observed $S$-wave spectra. All the spectra associated with each event are modelled jointly, using a $t$-distribution as likelihood function together with informative prior distributions for increased robustness against outliers and extreme values. The model includes the observed noise and a combined empirical Green’s function. It captures source-, receiver- and path-dependent terms in the description of the observed spectra by combining a physical source and attenuation model with a spatially and event-size dependent empirical compensation. The proposed method propagates estimation uncertainties along the entire processing chain starting from the hypocentre location and delivers reliable uncertainty description of the estimands. The objective is to automatically provide robust and valid descriptions of the observed $S$-wave spectra generated from an earthquake source in a noisy and heterogeneous environment. The efficiency of the method is tested with synthetic seismograms, and the model is calibrated and cross-validated using 31 640 mining induced seismic events in iron ore mine (in north of Sweden) with an comprehensive seismic network. The model is evaluated using both posterior predictive checks and residual analysis and we found no evidence that indicates any model deficiencies with respect to central tendency, dispersion and residual trends.

**Key words:** Probability distributions; Earthquake dynamics; Induced seismicity; Statistical seismology.

**LIST OF SYMBOLS**

$\mathcal{F}$ Fourier transform.

$\phi_{ik}$ The $i$th frequency coefficient for $j$th sensor in $k$th event.

$Y_j$ Observed $S$-wave spectra.

$\phi$ Parameter vector $[\Omega_0, f_c, n, Q, \beta]$ used in $\mu_\phi(\phi)$ and $\sigma^2_{\mu_\phi}(\phi)$.

$\mu_\phi$ Combined model (for the logged $S$-wave spectra).

$S$ Source model.

$\Omega_0$ Amplitude of source model ($\propto M_0$).

$f_c$ Corner frequency.

$n$ Spectral fit parameter (high-frequency falloff rate).

$G_\theta$ Combined Empirical Green’s Function (CEGF).

$A_\theta$ Modelled attenuation.

$Q$ Quality factor.

$\beta$ Spreading parameter.

$E_\theta$ Empirical compensation.

$N_\theta$ Observed noise spectra (obtained prior to P-wave onset).

$\sigma^2_{\mu_\phi}$ Propagated uncertainty.

$\sigma^2_{E_\theta}$ Uncertainty in $E_\theta$.

$\sigma^2_{\Omega_0}$ Uncertainties in ray length.

$\sigma^2_{f_c}$ Uncertainties in origin time.
\( \sigma_0^2 \) Uncaptured uncertainties.

\( R_{jk} \) Model residual.

\( \hat{p} \) Estimated hypocentre.

\( V \) Expected S-wave velocity between \( j \)th sensor and \( \hat{p} \).

\( r_j \) Expected distance between \( j \)th sensor and \( \hat{p} \).

\( \tau_j \) Expected travel time between \( j \)th sensor and \( \hat{p} \).

\( \Omega_j \) Set of event indices triggered by \( j \)th sensor in the calibration data set.

\( \nu \) Normality parameter, parameter controlling the tail width in a \( t \)-distribution.

\( \theta \) Parameter vector \( \{ \phi, \nu, \sigma_0^2 \} \), containing both the physical parameter vector \( \phi \) and the statistical parameters \( \nu, \sigma_0^2 \).

\( \lambda, \lambda_0 \) Hyperparameters, defines the prior distributions for \( \nu \) and \( \sigma_0^2 \), respectively.

\( \mu_m, \Sigma_m \) Hyperparameters, defines the prior distributions for \( \log ([\Omega_0, f_c]) \) in model \( m \in \{ \text{A, B} \} \).

\( a_B, b_B \) Hyperparameters, defines the prior distributions for \( \eta \) in Model B.

1 INTRODUCTION

Seismic source parameters like seismic moment, seismic energy, source radius and stress drop are important parameters for understanding the physics of earthquakes (Prieto et al. 2007; Van Houtte & Denolle 2018). Additionally, these parameters are commonly used in subsequent analysis, for example in hazard assessments (Cornell 1968; Wesseloo 2018) or aftershock analysis (Kozlowska et al. 2020), but their estimation uncertainties and systematic deviations are often overlooked. For instance, in sequential analysis using cumulative measures (see e.g. Thelen et al. 2010; Nordström et al. 2020) there is a risk of aggravating the biases as a result of the accumulation of systematic errors. For induced seismicity, like mining, the seismic waves are significantly influenced by heterogeneities along the propagation path and ambient noises causing uncertain estimates of the source parameters.

While this is also valid for tectonic seismicity, it is especially challenging for mining induced seismicity as investigated here in LKAB’s underground iron-ore mine located in Kiruna, Sweden (see Fig. 1 for a horizontal overview). In this mine, the rock mass contains large heterogeneities relative to the lengths of the ray paths between the events and the sensors (e.g. extracted rock mass, zones of fractured rock, large geological and stress variations as explained in Sjöberg et al. 2012; Vatcher et al. 2014; Nordqvist & Wimmer 2016; Vatcher 2017; Svartsjaern & Saiang 2012; Troll et al. 2019). Also, the earthquakes are usually small in size (99 per cent with \( M_w < 1 \)) with low signal-to-noise ratio (SNR) and the recorded waveforms are commonly contaminated with disturbances from mining operations (e.g. blasting, scaling, drilling, crushing, mucking, pumps, ventilation and heavy vehicles). Additionally, the sensors are placed within active event clusters as shown in Fig. 1 where multiple sensors may end up within the source radius, causing near-field effects like complex waveforms, flat spectra and resonances patterns (see e.g. Mikumo & Miyatake 1978; Vidale et al. 1995; Legrand & Delouis 1999; Yamada & Mori 2009). Therefore we claim that it is of equal importance to estimate the corresponding uncertainties as the source parameter values themselves. If we do not have any confidence in a value, that value is meaningless, and may impair subsequent analysis and conclusions. For example, estimates of stress drop may vary several orders of magnitude between individual studies (as discussed in Abercrombie 2015; Abercrombie et al. 2016) and if used for estimation of strong ground motions (see e.g. Oth et al. 2017) for hazard assessments it may have severe consequences. Especially, if the estimates lack confidence or the sequential models do not account for it.

Source parameters can be estimated by fitting the source model to attenuation-corrected spectra (see e.g. Oye et al. 2005; Shearer 2006) or by fitting the observed spectra where the model describes the attenuation (see e.g. Ida et al. 2003; Supino et al. 2019). The benefit with attenuation-corrected spectra is that they can be stacked together to estimate a mean spectrum for every sensor and therefore partly compensates for radiation-pattern effects (Oye et al. 2005). It is a computationally fast method that results in a smooth source-spectra, but it relies on removing the frequency coefficients that are under the influence of noise to obtain good model fit. Additionally, stacking based on sample mean is also sensitive to outliers (see e.g. Gelman et al. 2004; Martinsson 2012; Kruschke 2014).

Challenges such as mining disturbances, extreme values and outliers, associated with arrival times obtained from automatic location procedures (see e.g. Martinsson 2012) in heterogeneous environments increase the demand on robust automatic source parameter estimation with respect to uncertainties and outliers in the input data (Ross & Ben-Zion 2016). Estimates of the uncertainty can be obtained in different ways. For example, it can be estimated by: taking the standard deviation of the estimates for each spectrum (Caprio et al. 2011); removing one sensor at a time using the jackknife method (Prieto et al. 2007); taking the variance weighted uncertainty (Supino et al. 2019) where the variance for each spectrum are obtained using Gaussian assumptions and estimated with Markov
chain Monte Carlo (MCMC) methods (Gelman & Hill 2006; Kruschke 2014); evaluating the skewness of the misfit function (De Nolle & Shearer 2016).

Instead of iterative methods, we rely on the benefits provided by a joint estimation procedure (see e.g. Gelman et al. 2004; Kruschke 2014; Martinsson & Törnman 2019) and propose a MCMC estimation method similar as Supino et al. (2019) but here the observed spectra \( Y_j \) are modelled jointly for the \( j \)th frequency coefficient in the \( j \)th site. A valid description of all the observations are an important part of model validation (Pintelton & Schoukens 2001; Gelman et al. 2004; Kruschke 2014; Martinsson & Törnman 2019). If a model is not able to properly describe all the observations, in terms of central tendency and dispersion, then inference obtained using the model can not be trusted. For robust methods, this description must also include disturbances, outliers and extreme values. As opposed to using a Gaussian distribution for the likelihood (Supino et al. 2019), we propose instead a Student’s \( t \)-distribution due to its robustness against outliers and extreme values (Gelman et al. 2004; Martinsson 2012; Kruschke 2014). For example, in Martinsson (2012) it has been shown that the Gaussian distribution in particular, performs poorly and is very sensitive to a single outlier value when estimating the hypocentre. Using it in presence of possible outliers, will result in biased estimates and unrealistic uncertainty regions; favouring heavy-tailed distributions. The \( t \)-distribution is symmetric and bell-shaped like the normal distribution but has heavier tails. The tail mass is controlled with an additional parameter \( \nu \) called the degrees of freedom when used in the Student’s \( t \)-test or the normality parameter when used in the proposed model to increases the robustness against outliers (see e.g. Gelman et al. 2004; Kruschke 2014). When \( \nu = 1 \) the \( t \)-distribution takes the form of the Cauchy distribution and when \( \nu > 30 \) it is similar to the Normal distribution.

To obtain realistic estimates of the source parameters, the model parameters need to be restricted to physically reasonable values (e.g. they need to be larger than zero). To obtain reliable estimates with corresponding uncertainties, the method needs to capture the cumulative uncertainties along the processing chain. We propose a fully automatic Bayesian method as in Pugh et al. (2016) that both restricts the estimates with prior information as in Martinsson (2012); Stähler & Sigloch (2014) and includes the estimated uncertainties prior to source inversion. For instance, we include the estimated uncertainties in the origin time \( t_0 \) and the source receiver distance \( r \), both obtained from the hypocentre estimation algorithm (described in Martinsson 2012).

All these properties described above (e.g. robustness against outliers and extreme values, informative prior distributions to regularize the solution, uncertainty description of the input covariates for proper weighting) are needed to meet our objective: to derive a fully automated and robust method that provides a valid description of the observed \( S \)-wave spectra of mining induced events. A valid description needs to capture: the variations in the propagation medium, the variations at the source and the influence of temporal noises. We will introduce these more in the next section.

The rest of the manuscript is structured in the following way. In Section 2, we introduce the attenuation and source model considered here, together with a description of the background noise. Section 3 is devoted to the mine site, the measurement system and the data. A short introduction to Bayesian inference is given in Section 4 followed by a more detailed presentation of the Bayesian model in Section 5. The stepwise estimation procedure is described in Section 6 and the inferential results are presented in Section 7 followed by discussion and conclusion in Sections 8 and 9, respectively.

## 2 Theory

For earthquakes, the source, the paths and site effects are challenging to distinguish into separate terms from the observations since the effect of each contribution is jointly connected. On the other hand, the additive noise characteristics \( N \) can be estimated directly from the recorded wave forms prior to the \( P \)-wave onset. From a modelling perspective, we need to describe the observed spectra \( Y \) using commonly accepted models describing the source \( S \) and the attenuation \( A \). These models usually are an oversimplification, which leads to uncaptured behaviour embedded in the residuals. If these behaviours are systematic for certain ray paths, they can be estimated from training data by introducing an empirical compensation term \( E \). The origin of \( E \) will remain unknown as Shearer et al. (2019) stated, and we can for simplicity assume that \( E \) describes site effects at the source (as in Shearer et al. 2006) or describes the propagation channel as in this study. However, due to the commutative property of these terms in the model \( Y = S(E)A + N \), the result is the same regardless if the empirical compensation term \( E \) is devoted to the source, the propagation channel or a mix of both. In the following three subsections we will describe the source \( S \), the attenuation \( A \) and the noise \( N \) more thoroughly.

### 2.1 Attenuation

The attenuation can be modeled in many ways and here we use a simple model \( A \) that involves intrinsic attenuation and geometric spreading and is defined as

\[
A = e^{-\frac{\beta f}{Q}} \times \frac{r^{-\beta}}{\beta},
\]

where the quality factor \( Q \) depends on the medium and usually increases with depth (Patanè et al. 1994). At our shallow depths we expect \( Q < 300 \) (O’Neill & Healy 1973; Johnson & McEvilly 1974; Wang 2004; Oye et al. 2005; Ge et al. 2009). The geometric spreading parameter \( \beta \) describes the decrease in spectral level when the wave propagates away from a source. For far-field body-waves in homogeneous medium the wavefront is spherical with \( \beta = 1 \) in comparison to surface waves where the wavefront is cylindrical with \( \beta = 0.5 \). For heterogeneous medium (Drouet et al. 2011) expect \( \beta > 1 \) due to scattering and reflections. Some empirical estimates of \( \beta > 1 \) are observed for example by Zhu et al. (1991), Atkinson (2004), Akinci et al. (2006), Edwards et al. (2011) and Drouet et al. (2011).

There exist more complex attenuation models for example where \( Q \) and \( \beta \) are frequency dependent (see e.g. Zhu et al. 1991) or \( Q \) increase with travel time (see e.g. Wang 2004). However, the mining environment is extremely heterogeneous (Martinsson 2012; Martinsson & Törnman 2019), containing excavated volumes, complex geology and it is continuously changing due to continuous mining (see e.g. Vatcher et al. 2015, 2018; Nordqvist & Wimmer 2016; Troll et al. 2019). This changing environment causes frequency-dependent spatiotemporal attenuation effects on the seismic waves. These complex effects are difficult to model but can be empirically estimated using traditional empirical Green’s function EGF (Mueller 1985), multiple empirical Green’s function MEGF (Hough 1997) or combined empirical Green’s function CEGF (Shearer et al. 2006).

Traditional EGF and MEGF use smaller events in the proximity of the source with similar source time function to estimate the impulse response, assuming that the interevent-distance is much smaller than the event-receiver distance. The reason for small events is that
the source time function is shorter, generating a flatter frequency response that hopefully spans the frequencies of larger events. The CEGF proposed by Shearer et al. (2006) uses spatially homogeneous attenuation (i.e. the path effects are a simple function of travel time) together with an empirical correction for the unmodelled behaviours in the EGF events. The CEGF works for events spanning a broad range of magnitudes (Shearer et al. 2006), not limited to events with one or two magnitude units greater than the smallest EGF events (calibration events) as traditional EGF or MEGF, where the smallest events must be discarded.

In Shearer et al. (2006, 2019), the empirical correction is based on stacking of a large number of EGF events, resulting in a smooth correction spectra, where each EGF event has equal impact due to stacking. While in Abercrombie (2015), Abercrombie et al. (2016) and Ruhl et al. (2017) the empirical correction is based only on EGF events in the proximity of the source (i.e. within distance related to the source radius) and that mimics the target event (i.e. with cross-correlation above a threshold). Abercrombie (2015) observed that, including distant or dissimilar EGF events resulted in biased estimates compared to using only similar EGF events in the proximity of the source. Shearer et al. (2019) observed that stacking of a large number of EGF events resulted in more consistent estimates for densely clustered earthquakes compared to traditional EGF approaches.

Here we propose a new CEGF approach that uses all EGF events, resulting in a smooth compensation spectra as in Shearer et al. (2006, 2019), but it also favours similar (i.e. in size) and nearby EGF events as in Abercrombie (2015), Abercrombie et al. (2016) and Ruhl et al. (2017).

However, our method does not truncate calibration data with respect to size, correlation or distance. Instead, we interpolate with respect to both size and distance, where calibration events near the source and with similar size will have a larger impact in the estimate of the Green’s function $G$. We investigate appropriate interpolation of the EGF events based on residual analysis using cross-validation and simply propose one that provides the best description of our observations at hand, in terms of central tendency and dispersion. Additionally, we also estimate the underlying uncertainties $\sigma^2_G$ for the empirical compensation spectra $E$, which is part of $G$. The frequency- and sensor dependent uncertainty $\sigma^2_E$ is used in the Bayesian inference for proper weighting.

### 2.2 Source model

Two commonly used seismic source models are Brune (1970) and Boatwright (1980), both built for circular cracks and based on far-field assumptions. The Brune model has a broader corner compared to the Boatwright model and this causes larger uncertainties in general for the estimates of the frequency that defines the corner (i.e. corner frequency). However, the broader corner may allow for more realistic estimates of the corner frequency near the maximum frequency, as observed by Ruhl et al. (2017), since the broader corner allows for more use of the curvature in spectra. There exist more complex source models that contain an additional source dimension like rectangular or elliptical rupture geometries (see e.g. Savage 1972; Denolle & Shearer 2016), or models that include directional effects (Kaneko & Shearer 2015; Abercrombie et al. 2017). Here we use a simple Brune-type source model

$$S = \frac{2\pi f M_0}{1 + (f/f_c)^n}. \quad (2)$$

expressed in velocity since it is the format of the measured data, with the parameters: corner frequency $f_c$, seismic moment $M_0$, and the spectral fit parameter $n$ (also known as the high-frequency falloff rate for the displacement spectra). Brune (1970) proposed $n = 2$ which is the value most commonly used for source models (see e.g. Oye et al. 2005; Shearer et al. 2006). However, there are many cases where $n$ needs to be adjusted to obtain reasonable fit (see e.g. Brune et al. 1986; Lie 1992; Patanè et al. 1994; Abercrombie 1995; Prieto et al. 2007; Zhang et al. 2011; Ross & Ben-Zion 2016; Van Houtte & Denolle 2018). Despite that $n$ is mainly used as a model fit parameter, there are some attempts to define the physical meaning of $n$. For example, Frankel (1991) suggested that $n$ depends on the strength along the fault zone, Brune (1986) observed a positive correlation to stress drop and Lee et al. (2003) stated that the radiated energy remains finite only if $n > 1.5$. Here we evaluate two possible options of the source models

Model A: Brune-type model with fixed $n$ (i.e. $n = 2.31$)
Model B: Brune-type model with event specific $n$, given some constraints (i.e. $n > 1.5$)

More information about the value and constraints of $n$ in Model A and Model B is defined in Sections 6.1 and 6.2, respectively.

### 2.3 Background noise

The background noise affects the observed spectra for both induced and natural earthquakes (Zamani & Murrell 1978). The recorded signal $y_t = s_t + n_t$ contains both the received signal $s_t$ from the source and the noise $n_t$ for time sample $t$. In a mining environment the characteristics of the noise $n_t$ are spatio-temporal (e.g. caused by vibrations from drilling, pumps, fans, traffic, etc.) and may significantly influence the estimated result if $n_t$ is not properly dealt with. A typical example of spatio-temporal mining noises is shown in Fig. 2 where the second seismogram contains periodic bursts not related to the seismic event.

The common way to handle noise in estimating $S$-wave spectra is to either only consider the frequency coefficients where the signal-to-noise ratio (SNR) is above a specified threshold (usually set to 2 or 3, Shearer et al. 2006; Oth et al. 2011; Drouet et al. 2011; Denolle & Shearer 2016; Ross & Ben-Zion 2016; Van Houtte & Denolle 2018), or to weight the spectral coefficients by the SNR in the estimation process (Boatwright et al. 1991; Maranò et al. 2017). Truncating above a SNR threshold will neglect certain frequencies or distant sensors despite the fact that they still carry useful information. Also, applying SNR weighting increases the impact of sensors near the source, where part of the spectra may violate the far-field assumption and, additionally, these near-source sensors are more sensitive to location errors. Due to the above, near-source sensors should preferably be downweighted in the estimation process, especially those installed close to or inside the estimated source radius.

Estimation of the source model parameters is commonly done in the frequency domain, for example using Brune’s model. A frequency-domain representation is often favourable from an estimation and a modelling point of view, providing not only a simple mapping of models between continuous- and discrete-domain using the discrete Fourier transform (DFT, Pintelon & Schoukens 2001). The DFT also has beneficial asymptotic statistical properties of the measurement noise (Brillinger 1981; Ljung 1985), such as uncorrelated and complex normal distributed DFT coefficients, particularly useful in the estimation step (Kay 1993; Pintelon &
The right-hand panels show the model residuals (Schoukens 2007; Martinsson 2008). The Fourier transform grams for this event are shown in Fig. C1 in Appendix C. The first and third Figure 2. black, respectively, and red parts (i.e. the event. The noise and S-wave signal windows are coloured in grey and in the analysis. The model fit and the model residuals for both seismograms (logarithmically spaced) noise (Schoukens et al. 2020). The rock mass is heterogeneous with large variations in material density (Vatcher et al. 2014), rock strength (Svartsjaern & Saing 2017), wave velocity (Vatcher et al. 2018) and stresses (Sjöberg et al. 2012; Vatcher 2017). Additionally, the rock mass contains extracted volumes, altered zones, clay and crush zones (Vatcher 2017), all together influencing the wave propagation in the medium. Additionally, the sensors are usually placed where the events occur, causing small separation between the waves (i.e. P and S wave) and multiple sensors may end up inside the estimated source radius, especially for larger events.

The seismic measurement system shown in Fig. 1 is dense especially on the footwall side where most infrastructure is placed) and mine wide (i.e. covers the entire length of the orebody) with approximately 250 active geophones (totally 287 installed and some depleted thru time). Approximately half of the geophones are triaxial and the other half are uniaxial. The frequency responses also differ, depending on sensor type, and roughly 47 per cent of the geophones in the system have a natural frequency at 4.5 Hz, another 49 per cent at 14 Hz and finally 4 per cent at 30 Hz. The 30 Hz sensors correspond to the oldest sensors and these are not included in this study. The majority of the sensors are installed inside drilled holes, mainly vertically oriented, and grouted into the rock mass. For the three-axial sensors, we only use the sensor component that is parallel with the hole which is for the moment the only known component direction. The system is provided by a commercial company (Institute of Mine Seismology) and is likely one of the largest underground in-mine seismic systems in the world (Dineva & Boskovic 2017; Kozlowska et al. 2020).

The complete data set contains 31 640 seismic events recorded between the time period from 2018.12.10 to 2019.01.24. The events in the complete data set are chosen to have well defined locations, with 95 per cent confidence volume of the hypocentre distribution (Martinsson 2012) of less than 10$^5$ m$^3$ (equivalent to a sphere with a radius of 28.8 m). The hypocentre locations and corresponding uncertainties are automatically estimated (see Fig 3), including the arrival times, according to the technique described in Martinsson (2012). The events are randomly divided into two data sets where 70 per cent of the events are used for calibration (defined as the calibration data set) and 30 per cent for validation (defined as the validation data set and shown in Fig. 1). For each trace (illustrated in Fig. 2) we use the noise window $n_t$ (in grey) obtained prior to P-wave onset, and the signal window $y_t$ (in black) defined from the onset of the S wave with a length up to 500 ms. The S-wave windows...
Figure 3. Hypocentre and centre origin time (where \( t_0 \) is the posterior mode estimate of the origin time) distribution of the event in Fig. 2, where the cyan star is the posterior mode estimate of the hypocentre. Red areas are extracted rock-mass at the depth of the hypocentre (i.e. \( z \approx 820 \) m), used as hard prior restrictions in the Bayesian hypocentre estimation algorithm (Martinsson 2012). Orange triangles are the triggered sensors by the event. The total number of sensors are visible in Fig. 1). The lines in light and dark grey are the tunnels at two different depths, where the depths are marked as triangles in the z-axis on right-hand side.

considered here are slightly larger than those reported in Oye et al. (2005) for a mine with a similar seismic measurement system. The reason is to capture more energy for larger events. We choose to keep the observed spectra as unampered as possible following the guidelines in Pintelon & Schoukens (2001) using rectangular window functions (for both \( y_t \) and \( n_t \)). This will also preserve the high energetic part in \( y_t \) due to the uneven energy versus time distribution (i.e. the largest part of the energy is generally located in the beginning of the signal window). Furthermore, we are primarily interested in finding a model that describes our frequency-domain distribution (i.e. the largest part of the energy is generally located in the beginning of the signal window). Furthermore, we are primarily interested in finding a model that describes our frequency-domain observations (Kay 1993; Pintelon & Schoukens 2001) rather than in, for example, detecting single frequency components by pushing down side lobes (see e.g. Oye et al. 2005, using cosine taper) or smoothing the observed spectra (see e.g. Thomson 1982; Prieto et al. 2007, 2009, using multitaper) to reduce spectral variation. Adopting a statistical approach as we do here, there is generally no need to reduce spectral variations since this part is included in the model description as dispersion. The noise spectra \( N \) and signal spectra \( Y \) are first estimated using the DFT then resampled to 128 logarithmically spaced frequency coefficients in the interval from 14 to 1000 Hz, to reduce the number of coefficients and enhance the impact at low frequencies (see e.g. Ide et al. 2003; Lior & Ziv 2018). Additionally, to reduce possible electrical disturbances that may contaminate the sensor measurements, that is 50 Hz and overtones, we remove the corresponding DFT coefficients prior to resampling.

4 BAYESIAN INFERENCE

One of the advantages with a Bayesian method is that it can provide a solution even if the data is poor or inconsistent, by leaning on the support of informative prior distributions (see e.g. Martinsson 2012). The aim of adopting Bayesian inference here is to embed prior knowledge when estimating the source spectral parameters of a particular event. Informed prior distributions will regularize the Bayesian inference and provide the necessary robustness for dealing with, for example events triggered by only a few sensors or contaminated with outliers, inconsistent data or disturbances from mining noise.

Bayesian inference takes us from our prior beliefs to our posterior beliefs given some observations (i.e. likelihood) by using the Bayes’ rule which can be expressed as

\[
p(\theta|D) = \frac{p(D|\theta) \cdot p(\theta) }{p(D)},
\]

where \( p \) is the probability density (or mass) function, \( \theta \) represents the parameter vector we like to infer, and \( D \) is the input data that we will specify more in Section 5. The evidence \( p(D) \) scales the posterior distribution to be proper (i.e. to have unit probability mass). In reality we are mainly interested in the location and dispersion of the parameters given the observed data, and not the true probability density for a certain parameter value. We can therefore ignore the evidence \( p(D) \) term which is challenging to compute, and work with the following simplified expression

\[
\frac{p(\theta|D) \propto p(D|\theta) \cdot p(\theta)}{p(D)},
\]

For a more detailed introduction of Bayesian inference (see e.g. Tarantola 2005; Gelman & Hill 2006; Kruschke 2014).

4.1 Inferential methods

We rely on Markov Chain Monte Carlo (MCMC) methods (Gelman et al. 2004; Kruschke 2014) providing realistic estimates of credible intervals of parameter values (Martinsson 2012) compared to, for example approximations based on derivatives. The Bayesian inference here is based on MCMC samples from the posterior distribution and the posterior predictive distribution (Gelman et al. 2004; Kruschke 2014), with converged MCMC chains. The MCMC samples are provided by a slice sampler (Neal 2003) that are here modified to slice in the directions along the eigenvectors of the estimated covariance matrix obtained from the burn in samples (Martinsson & Törnman 2019). This approach is more efficient compared to the traditional slice sampler or the commonly used Metropolis algorithm (Gelman et al. 2004; Kruschke 2014) or variations of it (Martinsson 2012). The efficiency of a sampler is usually defined by the effective sample size (ESS, see e.g. Kruschke 2014), which is a rough measure of number of independent samples in a chain. Events in both the calibration and the validation data set are evaluated with 100 000 samples to obtain smoother histograms. The estimated posterior distributions are summarised by the posterior density for a certain parameter value. We can therefore ignore the evidence \( p(D) \) term which is challenging to compute, and work with the following simplified expression

\[
p(\theta|D) \propto p(D|\theta) \cdot p(\theta),
\]

where \( p \) is the probability density (or mass) function, \( \theta \) represents the parameter vector we like to infer, and \( D \) is the input data that we will specify more in Section 5. The evidence \( p(D) \) scales the posterior distribution to be proper (i.e. to have unit probability mass). In reality we are mainly interested in the location and dispersion of the parameters given the observed data, and not the true probability density for a certain parameter value. We can therefore ignore the evidence \( p(D) \) term which is challenging to compute, and work with the following simplified expression

\[
p(\theta|D) \propto p(D|\theta) \cdot p(\theta),
\]

For a more detailed introduction of Bayesian inference (see e.g. Tarantola 2005; Gelman & Hill 2006; Kruschke 2014).
In Bayesian statistics, the posterior probability is proportional to the likelihood function times the prior probability, and both terms are specified in the next two sections, respectively. After the model is defined we continue in Section 6 to estimate the hyperparameters and the empirical compensation data from the calibration data set, and finally we cross-validate the proposed models in Section 7.

5.1 The likelihood

To obtain robust parameter estimates, we propose a likelihood function [i.e. \( p(D|\theta) \) in (4)] given by the Student’s t-distribution

\[
E_{ij} = p(\mu_{ij}|\phi, \sigma^2, v) \quad \text{for} \quad \nu = \text{normality parameter (Kruschke 2014)}.
\]

The combined model \( \mu_{ij} \) (in 5) and the propagating uncertainty \( \sigma^2_{ij} \) in (6) depends on the additional parameters:

\[
\mu_{ij} = \mu_{ij}(\phi) = \log(S_i(\phi)G_i(\phi) + N_{ij}),
\]

\[
\sigma^2_{ij} = \sigma^2_{ij}(\phi, \sigma^2) = \sigma^2_{ij}(\frac{\partial \mu_{ij}}{\partial r_j})^2 + \sigma^2_{ij}(\frac{\partial \mu_{ij}}{\partial Q})^2 + \sigma^2_{ij}(\frac{\partial \mu_{ij}}{\partial \phi})^2 + \sigma^2_{ij}(\frac{\partial \mu_{ij}}{\partial \sigma})^2,
\]

where the source model and the Green’s function are given by

\[
S_i = S_i(\phi) = \frac{f_i\Omega_0}{1 + (f_i)^2},
\]

\[
G_{ij} = G_{ij}(\phi) = A_{ij}e^{E_{ij}/(\phi)},
\]

respectively, and \( G_0 \) depends on the modelled attenuation

\[
A_{ij} = A_{ij}(\phi) = e^{-a_{ij}/Q} \times r_j^\beta.
\]

The terms above depend on the model parameter vector \( \phi = [\Omega_0, f_c, n, Q, \beta] \). We explain these parameters more in Sections 5.1.1 and 5.1.2 next. To simplify expressions we will, when possible, omit dependencies and use the more compact leftmost expressions in (5)–(9). Note that \( S_i \) in (7) is proportional to \( S \) in (2) as the amplitude is proportional to the seismic moment (\( \Omega_0\times M_0 \)). We have also embedded \( 2\pi \) in this proportionality.

Given the asymptotic statistical properties of the DFT (Brillinger 1981; Ljung 1985; Kay 1993; Pintelon & Schoukens 2007; Martinsson 2008), the joint likelihood is the product over \( i \) and \( j \)

\[
p(D|\theta) = \prod_{\forall i \in I} \prod_{\forall j \in J} L_{ij}(\phi, v, \sigma^2_{ij}|D_{ij})
\]

\[
= \prod_{\forall i \in I} \prod_{\forall j \in J} p(\mu_{ij}|\phi, \sigma^2_{ij}(\phi, \sigma^2), v),
\]

where \( \mathbb{I} \) is a set of frequency coefficients (defined between the cut-off frequencies in observed spectra as shown in Fig. 2), and \( \mathbb{J} \) is a set of spectra obtained from the sensors triggered by the event, as shown in Fig. 3.

5.1.1 The combined model (\( \mu_{ij} \))

The combined model \( \mu_{ij} \) in (5) describes the expected logged observed spectra \( Y_{ij} \) and depends on the source model \( S_i \) in (7), the combined empirical Green’s function \( G_{ij} \) in (8) and the observed noise spectra \( N_{ij} \) obtained prior to the P-wave onset. \( G_{ij} \) contains both modelled physical attenuation \( A_{ij} \) in (9) and the empirical compensation \( E_{ij} \), where \( A_{ij} \) combines both the intrinsic attenuation and geometric spreading where \( Q \) and \( \beta \) are treated as mine wide constants.
5.1.2 The propagating uncertainty ($\sigma_i^2$)

The propagating uncertainty $\sigma_i^2$ in (6) depends on the uncertainties in: the origin time $\sigma_{0i}^2$, the ray-length $\sigma_{ji}^2$, the empirical compensation $\sigma_{Eij}^2$, and the uncaptured uncertainty $\sigma^2_{nc}$. The uncertainties $\sigma_{ji}^2$ and $\sigma_{0i}^2$ are obtained a priori by the location algorithm (Martinsson 2012) and the MCMC samples, where $\sigma_{ji}^2$ is the variance of the distance between the $j$th sensor and the hypocentre position, and $\sigma_{0i}^2$ is the variance of the origin-time estimate. See Fig. 3 for an illustration of the MCMC samples describing the posterior estimates of the hypocentre and the origin time. The last term $\sigma^2_{nc}$ describes the uncaptured uncertainty and is treated as an additional parameter that is estimated jointly together with $v$ and $\phi$ for each event.

5.2 The prior probability

Prior distributions are important to restrict the estimates to realistic values. This is also frequently applied in similar studies using non-Bayesian methods simply by restricting the parameter space to or between certain values (e.g. $2 \leq n \leq 3$ or $0 < f_c \leq 0.7$) as in Huang et al. (2016), which is in a Bayesian context the equivalent of uniform prior distributions spanning these domains. Restrictions are especially important in cases with dependencies between the parameters (Gelman & Hill 2006), as previously observed for the source model parameters (Van Houtte & Denolle 2018; Supino et al. 2019). For example, without restrictions we can obtain a reasonable model-fit with unrealistic parameter estimates (e.g. $n < 1.5$ as observed for the grey dots in Figs 6 and 7) since one can compensate for the other due to the pairwise correlation between the parameters.

Instead of assigning equally likely parameter values within these restrictions, Bayesian methods give us much more flexibility to shape the prior distributions within these restrictions based on relevant data. Relevant data can be, for example historical data as in Martinsson (2012), data obtained from similar studies, hierarchical approaches that estimates the prior distributions jointly with the model parameters as in Martinsson & Törnman (2019), or using a separate calibration data set as we do here.

To avoid unrealistic parameter estimates we define reasonable boundaries according to the literature of credible values for $\phi$ in Table 1, labelled as Region of Reasonable Values (RORV). These boundaries are not to be confused with the parameter space defined by the prior distributions and is simply used to remove unrealistic estimates for proper model calibration of the prior distributions used in the model. We return to it in Section 6 when we describe the estimation process.

Without the filtering step using RORV, occasional unrealistic parameter values and outliers would widen the prior distributions and prevent the regularization effect we are looking for by using informative prior distributions. One of the benefits of Bayesian models is to apply informative prior distributions and the necessary support in cases with noisy, inconclusive, and inconsistent data (see e.g. Martinsson 2012, for the benefits of informative prior distributions in hypocentre estimation).

The prior distribution can be factored into three different parts

$$p(\phi, v, \sigma_{nc}^2) = p(\Omega_0, f_c, n) \cdot p(Q, \beta) \cdot p(v, \sigma_{nc}^2).$$

where $p(\Omega_0, f_c, n)$ is the prior of the source model parameters, $p(Q, \beta)$ is the prior of the attenuation parameters, and $p(v, \sigma_{nc}^2)$ is the prior for the statistical parameters. These three parts are defined in the next three subsections, respectively.

5.2.1 Prior for the source model parameters $p(\Omega_0, f_c, n)$

There is a relationship between $\Omega_0$ and $f_c$ (Aki 1967) that holds for all event sizes (Sellers et al. 2003; Goodfellow & Young 2014), where $f_c$ decrease when $\Omega_0$ increase. To capture the relationship between the logged parameters we use a bivariate normal distribution as prior for $[\log(f_c), \log(\Omega_0)] \sim N(\mu, \Sigma)$, where $m \in \{A, B\}$ for Model A and B, respectively. In comparison to Van Houtte & Denolle (2018), where the logged parameters are assumed independent and assigned a normal distribution for each parameter, we include this dependency with a joint distribution. For the falloff rate $n$ we have two different priors that define the difference between the model structures defined as Model A and Model B (see Fig. 4). In Model A, $n$ is fixed for all events which is equivalent assigning a Dirac delta distribution as prior on $n$ centred at the fixed value, while in Model B the falloff rate $n$ is event specific and therefore we assign a Gamma distributed prior with shift of 1.5 to avoid non-physical estimates according to Lee et al. (2003).

The estimation procedure for the priors based on the calibration data set is explained later in Sections 6.1 and 6.2.

5.2.2 Prior for the attenuation parameters $p(Q, \beta)$

The attenuation parameters $Q$ and $\beta$ are treated as known constants in the modelled attenuation $A_0$ in (9). In Section 6.1, we describe how the values of $Q$ and $\beta$ are obtained. Treating them as known constants is equivalent to assigning Dirac delta distributed priors centred around their known values.

5.2.3 Prior for the statistical parameters $p(v, \sigma_{nc}^2)$

For the statistical parameters $v$ and $\sigma_{nc}^2$ in the likelihood, we use exponential distributed priors following (Kruschke 2014). The prior for the normality parameter $v$ is an exponential distribution shifted with one, using an expected value of 30 as proposed by Kruschke (2014). In Fig. 4 we adopt the simplified notations proposed by Kruschke (2014) for the shifted exponential distribution $v - 1 \sim \text{Exp}(\lambda_v)$, where $\lambda_v^{-1} = 29$. The shift of the prior for $v$ is to restrict the likelihood to not be more extreme in the tail width than a Cauchy distribution (i.e. when $v = 1$ the $t$-distribution is equivalent to a Cauchy distribution). The expected value for the prior of $v$ is assigned as the boundary when the model residual can be assumed as Gaussian distributed (i.e. when the estimate of $v > 30$).

The prior for the uncaptured uncertainty parameter is $\sigma_{nc}^2 \sim \text{Exp}(\lambda_{\sigma_{nc}})$. As estimated values of $\sigma_{nc}^2$ are close to zero for the majority of the events (see e.g. Figs 12 and D1–D7), we follow (Kruschke 2014) and choose to assign a weakly informed prior in relation to these values using an expected value of $\lambda_{\sigma_{nc}}^{-1} = 1$. This is to obtain a more robust Bayesian inference and to obtain converged MCMC chain in all cases.

6 THE ESTIMATION PROCEDURE

We are following the principles described in Gelman & Hill (2006); Gelman & Shalizi (2013); Lemoine (2019) to estimate the parameters: starting with a model with a large degree of freedom to determine the limitations of the inference, and then narrowing it down stepwise to obtain a well restricted model in the final analysis. Here the estimation procedure is summarized in Fig. 5 and performed in 4 steps in a hierarchical manner, where Steps 1–3 are the calibrations performed on calibration data to obtain informed prior distributions
Step 1: Estimate mine wide constants \((n, Q, \beta)\) or \((Q, \beta)\) for Model A or Model B, respectively.

Step 2: Estimate informative priors \(p(\Omega_0, f_c)\) or \(p(\Omega_0, f_c, n)\) for Model A or Model B, respectively.

Step 3: Estimation of empirical compensation data \((R_{ij}, \phi_k, \tilde{p}_k)\) necessary for \(E(\phi, p, j)\) and \(\sigma^2_{E}(\phi, p, j)\).

Step 4: Model validation of Model A and Model B with \((E = E)\) and without \((E = 0)\) empirical compensation for different spatial \((c_1)\) and size \((c_2)\) weighting coefficients.

> Model calibration, based on the calibration data set (70%)
> Model validation, based on the validation data set (30%)

Figure 5. A summary of the estimation procedure. Steps 1–3 is model calibration based on the calibration data set and Step 4 is model validation based on the validation data set (i.e. cross-validation). In Step 1 we estimate the fixed values assumed as known constants in the models using weakly informative priors on \(\phi\) defined in Appendix B. In Step 2 we estimate informative priors for the event specific parameters in the models using weakly informative priors on remaining parameters in \(\phi\), given the known constants in Step 1. In Step 3 we estimate the empirical compensation data necessary for the empirical model \((E_0\) and \(\sigma^2_{E}\)), given the estimates in Step 1, Step 2 and Model A. In Step 4 we validate the models with different weighting coefficients to determine proper spatial and size impact in the estimate, given Step 1, Step 2 and the empirical model relying on data from Step 3.

Table 1. Region of Reasonable Values (RORV). To be on the safe side we use boundaries of \(\phi\) that are slightly larger than those reported in the literature. For estimates of \(Q\): O’Neill & Healy (1973), Johnson & McEvilly (1974), Wang (2004), Oye et al. (2005), Ge et al. (2009), and for estimates of \(n\): Brune et al. (1986), Iio (1992), Patané et al. (1994), Abercrombie (1995), Zhang et al. (2011), Van Houtte & Denolle (2018) and for estimates of \(\beta\): Zhu et al. (1991), Kijko (1994), Legrand & Delouis (1999), Akinci (2004), Akincı et al. (2006), Edwards et al. (2011) and Drovet et al. (2011).

| \(\phi\) | \(\phi_{\text{min}}\) | \(\phi_{\text{max}}\) |
|-------|----------------|------------------|
| \(n\) | 1.5 | 4.5 |
| \(Q\) | 1.0 | 1000.0 |
| \(\beta\) | 0.5 | 3.0 |

and underlying data for the empirical compensation, and the final Step 4 is the validation performed on validation data.

Note that in Steps 1–3 we use good quality events from the calibration data set (i.e. with the certain requirements defined in Table 1) to avoid extreme values and outliers that may affect calibration. In the final Step 4 we process and estimate source model parameters for all events in the validation data set, relying on the constraints provided by the informative prior distributions and the help from the empirical compensation given by the calibrations.

In the first step, we estimate the fixed values assumed as known constants in the models. These fixed parameters correspond to \(n, Q, \beta\) in Model A and \(Q, \beta\) in Model B. The second step is to estimate the event specific priors for both models, given the fixed values in Step 1. The third step is to estimate the data used for the empirical compensation \(E_0\) in the Green’s function \(G_{ij}\) given the fixed values in Step 1 and the informative priors in Step 2. The final step is the model validation performed on all events in the validation data set. Here we evaluate both source models (Model A and Model B) and we determine the proper impact between size and distance for the training events used in the empirical compensation, given the fixed parameters (Step 1), estimated priors (Step 2) and the empirical compensations (Step 3). We describe each of the four steps in more detail next.

6.1 Step 1: Estimation of fixed parameters \((n, Q, \beta)\)

The attenuation parameters \((Q, \beta)\) and the source parameter \(n\) are often set to constants for the combined model \(\mu_0\) (see e.g. Oye et al. 2005; Shearer et al. 2006). The parameters are dependent with respect to model fit, as observed in Fig. 8, and changing one parameter’s value will affect the estimate of the other parameter’s value. Therefore we estimate all parameters in \(\phi\) simultaneously in this step (Gelman et al. 2004; Kruschke 2014), using weakly informative priors on \(\phi\) defined in Appendix B. The priors \(p(v, \sigma^2_{E})\) are defined in Section 5.2.3.

Since the parameters are correlated and not well restricted in Step 1, there will be estimates where one parameter value will compensate for another causing non-realistic parameter estimates. Therefore, we only accept estimates of \(\phi\) in the calibration set where 95 per cent HDI of \(\phi\) are in the RORV defined in Table 1. The accepted events within the RORV are represented as the magenta coloured dots and magenta coloured histograms in Fig. 6. The fixed values \((n, Q, \beta)\) are estimated from the accepted events as the weighted average, where the weights are the variance of each prior. The resulting estimates in Step 1 of the mine wide constants to be used in Steps 2–4 are: \(Q = 88.33, \beta = 1.19\) (and \(n = 2.31\) in Model A).

6.2 Step 2: Estimation of event specific priors

In this step we narrow down credible parameter values further for the sequential steps, by estimating informative prior distributions for the event specific parameters \((\Omega_0, f_c)\) for Model A and \((\Omega_0, f_c, n)\) for Model B. These informative priors will increase both the robustness against outliers and extreme values and also the precision of the estimates by adding the necessary support needed in cases with noisy, inconclusive, and inconsistent data (Martinsson 2012) and also limit the effect of intercorrelation. The prior distributions are illustrated in the top panel in Fig. 4 together with the hyperparameters describing them. Here the hyperparameters are estimated from the calibration data set for both models, given the fixed parameters from Step 1 and weakly informative priors (see Appendix B) for

![Figure 6. Posterior mode estimate of \(\phi = [\Omega_0, f_c, n, Q, \beta]\) for events in the calibration set, estimated with flat and weakly informative priors. The magenta coloured histograms of \(n, Q, \beta\) are shown together with the weighted average estimate for each parameter in cyan. The magenta dots in the scatter plot, represent the accepted events with the criteria that the entire 95 per cent HDI of \(\phi\) are inside the RORV (see Table 1). The corresponding grey dots are part of the rejected events where the estimated posterior mode of \(n\) is less than 1.5.](https://academic.oup.com/gji/article/227/1/403/6273640)
6.3 Step 3: Estimation of empirical compensation ($E_d$ and $\sigma^2_{E_d}$)

Here we process the events in the calibration data set using Model A, given the fixed parameters in Step 1 and the informative priors in Step 2, to obtain data for the estimation of the empirical compensation $E_d$ and corresponding uncertainties $\sigma^2_{E_d}$.

The estimates obtained from Step 3 are: the model residuals $R_{ij} = \log(Y_{ij}) - \mu_{ij}$, the size $\log(\Omega_{ij})$ and the location $\beta_k$ for the $k$th event. Here the size and the distance (between the target event and the events in the calibration set) are used to weight the $k$th impact of $R_{ij}$ on the $j$th sensors. This approach will not truncate the calibration data with respect to size or distance as Shearer et al. (2006) proposed. Instead, all calibration events are used and those located near the target event with similar size will have a larger impact on the estimation of the empirical compensation $E_d$ and the corresponding uncertainties $\sigma^2_{E_d}$. However, using a truncation free approach it is important to find the appropriate impacts (or weights) of the calibration events with respect to the size and distance when estimating the target event. Here we use the weighted average of $R_{ij}$ to estimate $E_d$ and the weighted variance to estimate $\sigma^2_{E_d}$ as

$$
E_{ij} = \frac{\sum_{k \in \mathcal{K}} w_k R_{ijk}}{\sum_{k \in \mathcal{K}} w_k},
$$

$$
\sigma^2_{E_{ij}} = \frac{\sum_{k \in \mathcal{K}} w_k (R_{ijk} - E_{ij})^2}{\sum_{k \in \mathcal{K}} w_k}.
$$

The weights

$$
w_k = \left( \frac{\prod_{l=1}^2 \frac{w_2(c_l)}{\sigma^2(c_l)}}{\sum_{k \in \mathcal{K}} \prod_{l=1}^2 \frac{w_2(c_l)}{\sigma^2(c_l)}} \right)^{-1/2},
$$

are jointly estimated from the distance and size related impact functions

$$w_{1k}(c_1) = \frac{1}{\| \tilde{p} - \tilde{p}_k \|^2 + \sigma^2},
$$

$$w_{2k}(c_2) = \mathcal{L}(\mu_{ij}, \sigma_{c_2} \log(\Omega_{ij}) \forall k \in \mathcal{K}_j),
$$

parametrized by the coefficients $c_1$ and $c_2$. The impact function $w_{1k}$ in (15) is defined as the inverse distance (between target event and the events in the calibration set) interpolation (see e.g. Martinsson 2012; Wettainen & Martinsson 2014) with the power of $c_1$ in (15).

For the size dependent impact function $w_{2k}$ in (16) we apply likelihood weighting based on a fitted normal distribution, centred around the current target size ($\mu_{ij} = \log(\Omega_{ij})$) and with a standard deviation ($\sigma_{c_2}$) estimated from all events triggered by the $j$th sensor in the calibration set (i.e. the standard deviation of $\log(\Omega_{ij}) \forall k \in \mathcal{K}_j$, where $\mathcal{K}_j$ is the set of event indices triggered by the $j$th sensor in the calibration data set). In Section 7.2, we investigate the impact of different values of the coefficients $c_1$ and $c_2$.

6.4 Step 4: Estimation of source model parameters

Here we process the source models based on the validation data set, using different combination of the compensation coefficients ($c_1$, $c_2$). The input data for Step 4 are: the fixed values obtained in Step 1, the informative priors given in Step 2, and the underlying data for the empirical compensation gathered in Step 3.
The data obtained from Steps 1–3 needed to estimate the source model parameters for one event consists of: the observed spectra \(Y_j\), noise spectra \(N_j\), calibration data \([R_{ij}, \phi_j, \tilde{p}_j]\), ray-lengths \(r_j\), traveltimes \(t_j\), hypocentre \(\hat{\rho}\), uncertainty in origin-time \(\sigma_{t0}^2\) and uncertainties in ray-lengths \(\sigma_{rj}^2\). Here we use the modelled travel times \(t_j = r_j/V(\hat{\rho}, j)\) to reduce the sensitivity to inconsistent arrival-time estimates, where the expected velocity along the ray path \(V(\hat{\rho}, j)\) is both sensor and spatial dependent and explained in Martinsson (2012).

7 Inferential results

Inferential results using different variations of the proposed models are presented here and the main focus is to address our aim, declared in Section 1, to obtain: robust, reliable and precise source model parameter estimands. We will start by using synthetic data to investigate the: accuracy, precision, pairwise correlation, credible regions, and the impact of underlying covariates. We then continue with real data and cross-validation to: assess the two alternative models; find appropriate values of the compensation coefficients; investigate model deficiencies with respect to size, frequency and distance using residual analysis and PPC; evaluate necessity of a robustness against extreme values and outliers indicated by the normality parameter.

7.1 Synthetic model evaluation

With synthetic data we evaluate the Bayesian inference in terms of accuracy, precision, pairwise correlation between parameters and the ability to capture the synthetic data within the estimated credible regions. Synthetic data are also used to evaluate the impact of the covariates (i.e. \(\sigma_{t0}^2\), \(\sigma_{rj}^2\) and \(N_j\)) using PPC (Gelman et al. 2004; Kruschke 2014).

Fig. 8 shows the Bayesian inference of 10 synthetic spectra generated at different distances \(r_j\) from the source, where the upper triangular figures are scatter plots of the MCMC samples and depict the pairwise correlations \(\rho\) of credible values of the parameter vector \(\phi\). The 95 per cent Confidence Interval \(\text{CI}\) of the correlation coefficient \(\rho\) is based on the Fisher’s \(Z\) transform. The diagonal figures are the histograms of posterior samples of \(\phi\), with corresponding 95 per cent HDI. The centre plot in the last row shows the true (i.e. \(\phi = \phi_{\text{true}}\)) and the estimated (i.e. \(\hat{\phi}\)) source spectra \(S_0(\phi)\) in cyan and black, respectively, together with the estimated 95 per cent HDI in grey. The HDI regions are calculated for each frequency component given the source spectra \(S_0(\phi)\) in (7) calculated from the posterior samples of \(\phi\) and summarizes the dispersion of the posterior distribution of the source given the observed data \(p(S_0(\phi)|D)\). Histograms showing the posterior samples of the statistical parameters \(\nu\) and \(\sigma_{\nu j}^2\) are also provided, influencing the shape and location of the likelihood.

In Fig. 8, we can see that the true parameters are within estimated confidence regions and that there are statistically significant correlations \(\rho\) (i.e. CI of \(\rho\) excludes zero) of different amounts between the credible parameters in \(\phi\). The largest correlation is observed between the credible values of \(\phi_0\) and \(\beta\), suggesting either to treat \(\beta\) as a fixed parameter value or to assign informative prior distribution to increase inferential robustness (Gelman & Hill 2006). The model residuals of the estimates can be assumed Gaussian due to the large estimated values of the normality parameter \(\nu\). This makes sense as the synthetic spectra are generated using Gaussian assumptions.

Also, for synthetic data generated from the model, we expect credible values of \(\sigma_{\nu j}^2\) around zero as there is no additional uncertainty term in the description of the total propagating uncertainty \(\sigma_{\nu j}^2\), since we assign \(\sigma_{\nu j}^2\) as the variance of the model residual. We additionally observe that the true source spectra \(S_0\) (in cyan) are close to the estimated \(\hat{S}_j = S_0(\hat{\phi})\) (in black) and within the 95 per cent credible regions (see e.g. Kruschke 2014).

Fig. 9 shows the credible regions of the posterior predictive distribution (see e.g. Gelman et al. 2004; Gelman & Hill 2006; Kruschke 2014) for the same synthetic data as in Fig. 8, illustrating the impact of the underlying covariates like \(\sigma_{t0}^2\), \(\sigma_{rj}^2\) and \(N_j\). Synthetic spectra from top to bottom represent: the closest sensor, the middle sensor, and the most distant sensor, respectively. The spectra are coloured by the distance \(r_j\) to the hypocentre. Three different 95 per cent HDIs (shown in red dashed, light grey area and dark grey area) obtained from the replicated data drawn from the posterior predictive distribution (see e.g. Gelman et al. 2004; Kruschke 2014), illustrate the separate impact of the propagating uncertainty \(\sigma_{\nu j}\) given different values of \(\sigma_{t0}\) and \(\sigma_{rj}\).

In Fig. 9 we also observe that an increase in \(\sigma_{t0}\) results in larger uncertainties at higher frequencies and an increase in \(\sigma_{rj}\) results in larger uncertainties for sensors near the hypocentre. The impact of the added noise description \(N_j\) in the combined model \(\mu_j\) is clearly observed when the received signal hits the noise floor, as the credible region describes the observed spectra, in comparison to a model neglecting the noise (e.g. as the model represented by the red curves in describing the real event in Fig. 2). As mentioned in Section 2.3, by including a description of the noise in the model, we can properly model the spectral coefficients under great influence from the noise (Ljung 1987; Söderström & Stoica 1989; Pintelon & Schoukens 2001).

7.2 Compensation coefficients \(c_1\) and \(c_2\)

The compensation coefficients \(c_1\) and \(c_2\) adjust the weights in the empirical compensation \(\hat{E}_{ij}\) and the corresponding uncertainty \(\sigma_{\nu i j}\), where appropriate values will improve the general model-fit. The impact of the compensation coefficients are explained in Section 6.3 and illustrated in Fig. 10, and their effect on the model-fit is summarized in Table 2.

Fig. 10 illustrates the impact of different values of \(c_1\) and \(c_2\), event sizes and locations on the unit-less empirical compensation \(\hat{E}_{ij}\) and corresponding uncertainties \(\sigma_{\nu i j}\). Here \(\hat{E}_{ij}\) is represented as solid lines and \(E_{ij} \pm \sigma_{\nu i j}\) as dotted lines coloured by event size \((\Omega_0 \in \{10^{-6}, 10^{-3}, 10^{-2}, 10^{-1}\})\) for the \(j\)th sensor, and evaluated at the average hypocentre location \(\hat{\rho}\) for the calibration events that is triggered by the \(j\)th sensor. The top six figures are plotted using the same sensor \(j = 122\) but with different values of \(c_1\) (related to distance) and \(c_2\) (related to the size). The bottom three figures use the same \(c_1\) and \(c_2\) but evaluated at different sensors \(j \in \{140, 241, 246\}\) to illustrate the differences at the sensor specific response.

In Fig. 10, the upper left-hand figure (where \([c_1, c_2] = [0, \infty]\)) corresponds to the stacked average of the model residuals spectra \(R_{ik}\) for the \(j\)th sensor, and for these values there are no dependencies on either distance or size. The weighting in (14) will therefore solely act as a sensor-dependent compensation. However, neglecting dependencies on distance and size will increase the residuals in Table 2 and reduce the model’s description of the validation data set. The upper middle and upper right-hand figures show the impact when dependencies on the size only are introduced. In general, we
observe flatter empirical compensations for smaller $\Omega_0$ and more frequency dependent compensation for larger $\Omega_0$. The left-hand and middle figures in the second row show the empirical compensation using only the dependency on distance, neglecting the dependency on event size. We observe a slight difference between the inverse distance (i.e. $c_1 = 1$) and the inverse squared distance weighting (i.e. $c_1 = 2$). The right-hand figure in the middle row shows the compensation for the same sensor using the values that provides the best model-fit given in Table 2. Finally, the bottom three figures use the same and proposed values $c_1 = 2$ and $c_2 = 1$, but evaluated at different sensors $j \in \{140, 241, 246\}$ to illustrate the large sensor specific differences observed between sensors, like: resonances at different frequencies, high-pass or low-pass compensation.

Table 2 summarizes the validation of the model fit in terms of dispersion of the residual, given the two different source models and different values of the empirical compensation coefficients. The inference is based on dispersion measures of the model residuals $R_{ijk}$ for all frequency coefficients in the validation set. Model A use fixed $n = 2.31$ while in Model B the parameter $n$ is estimated for each event. The columns are: Model type, size of the 25 per cent HDI of $R_{ijk}$, size of the 75 per cent HDI of $R_{ijk}$, and the standard deviation $\sigma_{R_{ijk}}$ of $R_{ijk}$.

In Table 2, we observe that the compensation coefficients $c_1 = 2$ and $c_2 = 1$ result in the lowest dispersion of the model residuals, therefore indicating favourable values in terms of model-fit relative to the other evaluated alternatives. These values correspond to the inverse square distance weighting, combined with default likelihood weighting with respect to the size (i.e. using simply the standard deviation $\sigma_w$ in (16)).

7.3 Residual analysis

Residual analysis is conducted by investigating the histograms of all residuals in the validation data set. The main purpose is to detect general remaining model deficiencies and asymmetries with respect to: frequency, distance and size (similar to Drouet et al. 2011). Residual analysis can be done on individual seismograms as shown in Fig. 2, but is too comprehensive for the large data set and would not detect general model deficiencies and trends. Also, quantitative measures like those in Table 2 are important in determining the model fit, but they don’t indicate remaining dependencies in the residuals.

Fig. 11 shows the histograms for all frequency coefficients of the model residuals $R_{ijk}$ in the validation set using the less flexible Model A. The histogram in a) shows the residual with $(E = E)$ and without $(E = 0)$ empirical compensation in magenta and cyan colours, respectively. The figures in b), c) and d) represent 2D histograms estimated with empirical compensation, to highlight remaining dependencies of $R_{ijk}$ on either: sensor’s hypocentre distance, frequency or event size, respectively. The colours of the bin counts in b)–d) are in log-scale to enhance tail differences.

In Fig. 11(a), with empirical compensation $(E = E)$ we obtain significantly less dispersion in the model residuals compared to
Posterior predict checks (PPC) replicated from the posterior estimated individual seismograms is also observed in Fig. 12 and in Figs D1–D7 in noise components are reduced. The more precise model fit for individual cases where the estimate of $\nu < 30$ the data contains either extreme values (e.g. sensor resonances) or outliers (e.g. poorly associated seismograms) with respect to the model and underlying covariates. In Fig. 12 we observe both extreme values like sensor specific resonances (top right-hand panel) and outliers caused by wrongful arrival times (see e.g. bottom left-hand panel and low amplitude spectra that is located near the hypocentre). Using a normal distribution as a likelihood function for the events where the estimate of $\nu < 30$ would result in biased parameter estimates and a poor model fit. Extreme values are here described by a more complex model (e.g. with the empirical compensation) and the outliers are handled by adjusting the tail width in the likelihood function, improving the robustness and model fit of the proposed model (see e.g. Gelman et al. 2004; Martinsson 2012; Kruschke 2014).

Table 3 shows the estimates of the normality parameter $\nu$ for the validation set, with and without empirical compensation, and for Model A and Model B. The columns are: the model, the compensation coefficients, the percentage of events where the estimates of $\nu < 30$ with 95 per cent confidence level, and the percentage of the events where the mode estimate $\nu < 30$.

In Table 3, we can see a vague difference of the estimates of $\nu$ between Model A and Model B in row two and three, respectively, in comparison to the estimates with and without empirical compensation in row one and two, respectively. Without empirical compensation 50 per cent of the events have a mode estimate of $\nu < 30$ in comparison with empirical compensation, where the corresponding result was less than 15 per cent of the validation events. An explanation is that many of the observed extreme values are caused by sensor specific resonances (see e.g. Fig. 10) that are captured by the distribution of the empirical compensation described by $E_{ij}$ and $\sigma_{E_{ij}}$.

7.5 Model A or Model B

Here we evaluate the less flexible Model A compared to Model B where $n$ is not fixed but instead event specific. In Section 7.4 the focus was on outliers and extreme values and here the evaluation is based on quantifying the distribution of the model residuals in terms of dispersion that is summarized in Table 2. We also observe the impact of model choice on the estimates of $f_c$ and $\omega_0$ in Fig. 13.

Fig. 13 compares the estimates of $f_c$ and $\omega_0$ obtained using Model A with and without empirical compensation, and Model A and Model B both with empirical compensation. Different panels show posterior mode estimates of $f_c$ (with confidence regions) and $\omega_0$ for the events in the validation set.

In Figs 13(a) and (c), we observe the relationship (Aki 1967) between $f_c$ and $\omega_0$. In panel (c), with Model B, we observe a increased number of low frequency events ($f_c < 10$) and less high frequency events ($f_c > 1000$) compared to Model A. In panels (b) and (d), the impact on $f_c$ for both models with and whiteout empirical compensation are evaluated. We observe that for events where the estimates between the models differ more, the confidence regions are usually larger as indicated by the thin whiskers.
Figure 10. The impact of different compensation coefficients, event sizes, and sites on the distribution of the estimated empirical compensation, parametrized by $E_0$ and $\sigma_{E_0}$, for an arbitrary point in space. Here $E_0$ is represented as solid lines and $E_{ij} \pm \sigma_{E_{ij}}$ in dotted lines, coloured by size, respectively. Each figure shows the empirical compensation for different sizes $\Omega_0 = [10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]$ for the $j$th sensor, evaluated at the average hypocentre location for all triggered events in $\mathcal{K}_j$ (i.e. set of event indices triggered by the $j$th sensor in the calibration data set). The top six figures are plotted using same sensor $j = 122$ but different compensation coefficients $[c_1, c_2]$, where $c_1$ is related to distance and $c_2$ to log ($\Omega_0$). The bottom three figures use the same $c_1$ and $c_2$ but evaluated at different sensors $j \in \{140, 241, 246\}$ to show the differences at the sensor specific response. (The empirical compensation is unit-less).

Table 2. Validation of the model-fit, evaluated with and without empirical compensation using different compensation coefficients ($c_1$, $c_2$) and source models (A,B). The inference is based on different dispersion measures of the model residuals $R_{ijk}$ for the entire validation data set. The columns are: Model, size of the 25 per cent HDI of $R_{ijk}$, size of the 75 per cent HDI of $R_{ijk}$, and the standard deviation $\sigma_{R_{ijk}}$ of $R_{ijk}$.

| Model | $(c_1, c_2)$ | [25 % HDI] | [75 % HDI] | $\sigma_{R_{ijk}}$ |
|-------|--------------|------------|------------|------------------|
| A     | None         | 0.4398     | 1.6729     | 0.7963           |
| A     | (0, $\infty$)| 0.3869     | 1.4562     | 0.6900           |
| A     | (1, $\infty$)| 0.3662     | 1.3878     | 0.6644           |
| A     | (2, $\infty$)| 0.3404     | 1.2991     | 0.6207           |
| A     | (0,1)        | 0.3733     | 1.4062     | 0.6622           |
| A     | (1,1)        | 0.3582     | 1.3530     | 0.6404           |
| A     | (2,1)        | 0.3344     | 1.2761     | 0.6075           |
| B     | (2, $\infty$)| 0.3294     | 1.2592     | 0.5997           |
| B     | (2,1)        | 0.3244     | 1.239      | 0.5883           |

shows a difference in the estimated corner frequencies between the two models. In panel (e), where we compare the estimates of $\Omega_0$ between Model A and Model B, we only observe minor differences in comparison to the estimates of $f_c$. A possible reason is that there is a stronger correlation between $f_c$ and $n$ compared to $\Omega_0$ and $n$ as indicated in Fig. 12 and Figs D1–D7. The strong correlation between $f_c$ and $n$ has also been observed in other studies (see e.g. Van Houtte & Denolle 2018). In panels (a) and (b) we observe that the empirical compensation has only a minor effect on the source model parameters in terms of the central tendency of the estimates, due to the robustness and flexibility of the proposed likelihood function in presence of extreme values as mentioned in Section 7.4.
Figure 12. Bayesian inference of the challenging validation event (same as in Figs 2 and 3) using Model B, where \( c_1 = 2 \) and \( c_2 = 1 \) for the weighting of \( E_{ij} \). The lower left-hand corner depicts the observed spectra \( Y \) coloured by distance \( r_j \) to the estimated hypocentre \( \hat{p} \). The scatter plots to the right of the diagonal show the MCMC samples, and also depict the pairwise correlations \( \rho \) (with 95 per cent CI) of credible values of \( \Omega_0, f_c \) and \( n \) in \( \phi \). Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra \( \hat{S} \) (in black) and the 95 per cent HDI (in grey). The plots in the right-hand column are the observed spectra for closest sensor, middle sensor and furthest sensor from top to bottom coloured by distance, together with the estimate of the observed spectra \( \hat{Y} = \exp(\mu_{ij}(\hat{\phi})) \) (in black) using the posterior mode \( \phi \) and 95 per cent HDIs (in grey). Histograms for the statistical parameters \( \nu \) and \( \sigma^2_0 \) are also provided to the left of the diagonal.

Table 3. The normality parameter \( \nu \) describes how Gaussian distributed the model residuals are. If \( \nu > 30 \) the model residuals can be assumed Gaussian. Columns in table are the model, the compensation coefficients, the percents of events where the estimates of \( \nu < 30 \) with 95 per cent confidence level and the percents of the events where the mode estimate \( \nu < 30 \).

| Model | \( (c_1, c_2) \) | max(CI(\( \nu \)) 95%) | \( \nu < 30 \) |
|-------|------------------|--------------------------|--------------|
| A     | None             | 22.1% 50.0%              |              |
| A     | (2,1)            | 3.3% 14.9%               |              |
| B     | (2, 1)           | 2.7% 11.5%               |              |

(see e.g. Gelman et al. 2004; Martinsson 2012; Kruschke 2014). However, in panel (a) we observe that with empirical compensation the most extreme dots are contracted towards the centre, indicating more robust estimates due to the increased description provided by the empirical compensation as previously shown in Section 7.3. In panel (f) we observe truncation effects in the histogram of \( n \) using Model B for approximately 10 per cent of the events, these events seem to favour a flatter model than is allowed by the prior restrictions (i.e. \( n < 1.5 \)). In panels (g) and (h) we observe that the corresponding uncertainties \( \sigma_{Omega} \) and \( \sigma_{f_c} \) (i.e. the standard deviation of the posterior samples of \( \Omega_0 \) and \( f_c \)) increase with the estimated size for both \( \Omega_0 \) and \( f_c \), respectively, for both models. The thin whiskers (i.e. HDI 95 per cent) in panel (b) and (d) overlap for 75.89 per cent (for Model A with and without empirical compensation) and 49.64 per cent (between Model A and Model B with empirical compensation) of the events, respectively.

The use of empirical compensation for Model A reduced the average uncertainties in \( f_c \) (i.e. mean of \( \sigma_{f_c} \)) by 27.62 per cent (i.e. from 39.06 to 28.27 Hz) and median uncertainties (i.e. median of \( \sigma_{f_c} \)) by 13.55 per cent (i.e. 5.69 to 4.91 Hz). For Model B with empirical compensation the mean and median uncertainty of \( \sigma_{f_c} \) are 24.24 and 8.21 Hz, respectively. The median value of \( \sigma_{f_c} \) increases for Model B compared to Model A, indicating more uncertain estimates of \( f_c \) in general. This is caused by the increased degrees of freedom in Model B. A reduced mean value for Model B indicates that there are less extreme values.

Table 2 shows different dispersion measures of the model residuals as an indication of model fit. We see that given same \( c_1 \) and \( c_2 \) for both models, Model B is always favourable in terms of model fit. Also, a similar model fit is given using Model B without size compensation (\( c_2 \rightarrow \infty \) in row 8), compared to using Model A with size compensation (\( c_1 = 1 \) in row 7). The best model fit are obtained using both size and distance compensation (\( c_1 = 2, c_2 = 1 \)) together with the more flexible Model B.

8 DISCUSSION

The focus in this study is to derive a robust Bayesian framework that provides a precise and valid description of the observed S-wave
spectra recorded in challenging environments. The model uses distributions to handle uncertainties and unknowns together with physical models for the central tendency in combination with a spatial and earthquake size dependent empirical compensation to capture systematic deviations. The proposed model is trained and evaluated on two separate data sets, consisting of 31 640 mining induced events in total. The processing is fully automatic and is applied without manual interaction. Processing of mining induced events is especially challenging due to the extreme environment containing excavated volumes, complex geology, stresses that is continuously changing (see e.g. Sjöberg 2014; Vatcher et al. 2014; Vatcher 2017; Svartsjaern & Saiang 2017, for further description of the complexities in the mine) along with recorded waves influenced by mining noises (e.g. shown for multiple seisograms for the event in Fig. C1). Also, due to the extreme environment, it is difficult to obtain and set reliable physical properties such as material densities, velocities and stiffness in the rock mass in the studied region. Therefore, we mainly focus here on a description of the source model parameters (i.e. $\omega_0$, $f_c$ and $n$) that can be estimated directly from the observed spectra and is not dependent on material properties as opposed to: seismic moment, radiated seismic energy, and stress drop (Kijko 1994).

Stress drop is often assumed to be independent of source size (Aki 1967) and log-normally distributed (Baltay et al. 2011) and is commonly used as a quality measure of source parameters. Despite this, it seems to depend on the source model (Abercrombie et al. 2016), frequency bandwidth (Ruhl et al. 2017) or what EGF events are included (Abercrombie 2015) and may therefore be a poor measure to quantify variations in earthquake source spectra (Shearer et al. 2019). In Fig. A3, we explore the values of $\omega_0 \times f_c^3$ (which is proportional to stress drop) with respect to: location, $f_c$, and $n$. We find no statistical significant correlation (i.e. CI 95 per cent of $\rho$ includes zero) with respect to $\omega_0$ and we have no evidence that would suggest size dependency and reject standard scaling (i.e. $\omega_0 \propto f_c^{1/3}$ (Aki 1967). We observe some correlation with respect to $f_c$ and $n$, together with some spatial patterns (i.e. cluster of events with larger stress drop) as in Shearer et al. (2006). The positive correlation with respect to $f_c$ may indicate that stress drop is overestimated for larger frequencies (which is the opposite of the observations made by Ruhl et al. 2017) and the positive correlation with respect to $n$ coincides with the finding made by Brune et al. (1986). The histogram in the bottom right rejects the log-normal assumptions, since the data is slightly skewed towards higher estimates in log-scale. This may be related to limited bandwidth since the high frequency events, especially those above $f_{\text{max}}$, have significantly larger stress drop.

8.1 The Bayesian approach and possible model extensions

To avoid additional complexity, the Bayesian framework is presented here using a well known and simple description of the source (Brune 1970) and the attenuation (Shearer et al. 2006). It is straightforward to replace the components in $\mu_\eta(\phi)$ to more complex ones.
Modelling errors (in the path, at the site, and of the source) are always inevitable and the emphasis here is on capturing the errors adequately using a combination of empirical compensation terms and statistical parameters.

Our model empirically captures the unmodelled errors by favoring similar (in terms of earthquake size and spatially closeness) EGF events with respect to the target event. The empirical compensation terms (described in Step 3) capture both path and site and sometimes source related terms (e.g. if a cluster of similar sized events has similar rupture behaviour as observed in Folesky et al. (2016)). In general, systematic errors that are not captured by our source and attenuation model in \( \mu_L(\phi) \) will end up in the empirical compensation term \( E_c \). Non-systematic errors will end up in the uncertainty terms (\( \sigma^2_{ij} \) or \( \sigma^2_{N_i} \)) in the statistical description and therefore expand the HDIs for the inferred parameters and spectra in different ways depending on their origin.

This does not mean that the modelling errors are lost and if we would like to extend the model, the information is still available in the residuals (see e.g. Fig. C2) and the empirical compensation spectra (see e.g. Fig. 10). With a more complex model we would reduce the systematic errors related to the source and consequently also reduce these unmodelled source characteristics embedded in the EGF (e.g. as we do with the more complex Model B).

An interesting extension is to include directional effects, e.g. by fitting some directivity function (Boatwright 2007) to the variations in spectral amplitude (Kane et al. 2013; Calderoni et al. 2015). However, we do not observe any evidence of directional effects such as spectral splitting (i.e. between the modelled and observed spectra) in the representative events in Figs D1–D7. This may be since 99 per cent of the events in the validation data set are less than magnitude one which was the detectable limit for rupture effects in Folesky et al. (2016).

Another aspect we need to consider when extending the model, is that an increased number of degrees of freedom generally leads to increased estimation uncertainties for the individual model parameters. With additional parameters and possible multicollinearity (i.e. intercorrelation), it is essential to restrict the parameters with informed prior distributions to provide the necessary robustness as mentioned in Section 4. This was observed when investigating the estimation uncertainty \( \sigma_{ij} \) of \( f_2 \) between Model A and Model B in Section 7.5. In general, model parameters in physical models (e.g. eqs (1) and (2)) are often prone to intercorrelation, as we are constrained to a specific parametrization. One of the strengths with Bayesian modelling in this regard is that we have an inbuilt mechanism to deal with overparametrization. This opens up the possibility to apply complex models with many parameters in combination with insufficient or inconsistent data, as long as these parameters are constrained by informative prior distributions (see e.g. Gelman & Hill 2006).

An important additional value associated with the full probability description proposed here, is that the uncertainties in the input data are propagated all the way to the resulting source model parameters, providing realistic uncertainty descriptions of the model parameters and observed spectra. This means that they can in turn be propagated to the next step when deriving source parameters. In the next step, the uncertainties in density and velocity can easily be incorporated and combined with those from the source model parameters to provide realistic uncertainties of, for example seismic moment, seismic energy and stress drop.

The extreme environment means that the model needs to be robust with respect to the input data and corresponding uncertainties. This is particularly important in this work when we rely on auto-processed input data in the form of hypocentre location, origin time and arrival times (described in more detail in Martinsson 2012).

Here we depend on the unique strengths of Bayesian statistics using informed prior distributions of the source model parameters in combination with robust likelihood functions that incorporates the uncertainties delivered together with the input data (Martinsson 2012) as shown in Fig. 3.

The estimation procedure could be addressed differently, for example by correcting the observed spectra for attenuation and fit the stacked average spectrum to the source model (Oye et al. 2005; Shearer et al. 2006). The stacked average alternative is computationally fast and results in a simple source spectra. However, using our proposed joint estimation method includes several benefits in comparison. For example:

(i) The possibility to include additive model terms such as the measured noise \( N_0 \) in the combined model \( \mu_{ij} \) in (5). By including a description of the noise in the model, we can properly model the spectral coefficients under great influence from the noise (Pintelton & Schoukens 2001; Ljung 1987; Söderström & Stoica 1989). Without this description a stacked method requires large SNRs, resulting in valuable information being discarded in the estimation process.

(ii) Obtaining a valid description of the underlying individual observations that can be tested using standard model validation procedures (Ljung 1987; Söderström & Stoica 1989; Pintelton & Schoukens 2001; Gelman et al. 2004; Kruschke 2014) such as residual analysis and PPCs described in Section 4.1 and illustrated in Figs 2, 9, 12 and Figs D1–D7.

(iii) Proper uncertainty weighting of \( f_j \)th frequency coefficient in the \( f_j \)th spectra, depending on the underlying covariates (see e.g. Figs 9, 10 or 12). Standard stacking based on sample mean results in equal weights for all spectra and over all frequencies.

(iv) Increased robustness against outliers and extreme values using heavy-tailed likelihood functions (see e.g. Gelman et al. 2004; Martinsson 2012; Kruschke 2014). Stacking based on sample mean follows Gaussian assumptions and is therefore sensitive to outliers (Ross & Ben-Zion 2016).

8.2 Prior knowledge

There are different alternatives to obtain informed prior distributions. One option is to adopt a full hierarchical Bayesian model, where the source model parameters for all the events in a study are jointly estimated along with the parameters in the prior distribution (see e.g Gelman & Hill 2006; Kruschke 2014; Martinsson & Törnman 2019). This approach is theoretically appealing but unfortunately not practical in this study with 31 640 events. Instead, we take an empirical Bayesian approach (see e.g. Gelman et al. 2004) where the prior knowledge that will be described by our prior distributions is based on parameter estimates obtained from a separate calibration data set. Here parameter estimates from the calibration data set will shape our prior distributions used for estimating the source model parameters of an event. Additionally, we also incorporate hard restrictions of the parameters that are physically motivated (see e.g. Table 1). These restrictions should be considered using any of the above options to obtain informative prior distributions.

Another direction is to adopt non-informative or vague prior distributions (Gelman et al. 2004; Kruschke 2014) to reduce the influence of the prior distribution and consequently neglect available prior information. The estimates of each source model parameters
would be driven by the data observed for that particular event only. This would, however, suffer from the same limitations as with classical inference methods, with estimates available only for events providing adequate data. To increase robustness, reduce intercorrelation, and limit the model’s flexibility, model parameters are often set to constant fixed values (e.g. \( n = 2 \) as in Shearer et al. 2006) or restricted by measurement limitations (e.g. \( f_r \) restricted to frequency bandwidth as in Huang et al. 2016) based on optimization strategies prior to inference. These approaches can also be seen as empirical Bayesian but with very informative prior distribution (i.e. a Dirac delta distribution centred at the particular parameter value).

The approach taken here is not set parameters to fix values (e.g. \( n = 2 \)), but instead let the calibration data set shape our prior distributions for the model parameters as we describe in Step 1 and Step 2 (see e.g. Fig. 5).

### 8.3 Empirical compensation

The empirical compensation term is sensor, frequency, spatial and event-size dependent, to account for the heterogeneous rock-mass and unique sensor effects (shown e.g. in Figs 2, 3, 12 and C1). This term reduces the uncertainties of credible intervals of the parameters as well as prediction intervals for the observed spectra, and consequently decreasing the estimated value of \( \sigma^2_0 \) that describes uncaptured variations. Without the empirical term, the proposed model is still applicable but the uncaptured variations (parametrized by \( \sigma^2_0 \)) would increase in order to describe the unmodelled variations in observed spectra caused by the heterogeneous rock-mass and sensor effects. This is also the case when the combined model (5) is not flexible enough (e.g. when the spectra is flatter than the model allows) to capture the observed spectra, as shown in Fig. E1 where \( \sigma^2_0 \) is significantly larger then zero. This can be compared to the posterior distributions of \( \sigma^2_0 \) in, for example the representative events of the data set shown in Figs D1–D7 where, the 95 per cent HDIs include zero. This is an indication that the model describes the data well and also an indication that data originates from genuine mining-induced seismic events. Here \( \sigma^2_0 \) would likely act as an important additional classification variable to reject non-genuine events like, for example ore-pass noises, mucking, drilling or crushing that are commonly observed in the mine.

Typical challenges in a mining environment such as: induced noises, resonances and wrongfull arrival times are all shown for the event in Figs 2, 3, 12 and C1. The results show the strengths of our model in coping with these challenges by using the spatial- and size-dependent CEGF and including the observed noise spectra \( N_j \) in the combined model, together with proper weighting based on underlying covariates. In Appendix D, we provide additional modelling results in Figs D1–D7 that are more representative of the entire validation data set and for different event sizes (\( \Omega_0 \)). The events are not picked to show challenges and strengths, but instead correspond to the specific event indices matching exactly the: 1st, 5th, 25th, 50th, 75th, 95th and 99th percentiles with respect to size (\( \Omega_0 \)) for all events in the validation data set. These results will provide a more representative picture of the validity and modelling performance in general. For model validation, we assess replicated data from the distribution to detect aspects of the data not captured by the model (see e.g. Gelman et al. 2004) and we found no evidence that would indicate model deficiencies or flaws as seen from a statistical modelling perspective.

In a mining environment there is likely to be considerable \( Q \) variations in the rock mass, with lower values in more disturbed rock. These variations are likely to be systematic for certain ray paths and should manifest as slope variations in the residual spectra \( R_{95} \) if not accounted for. Therefore, these variations are likely to be captured by our spatial and size dependent empirical model \( \mathcal{E}_P \), in a similar manner as seen e.g. in the lower left-hand subfigure in Fig. 10.

By observing the residual spectra in Fig. C2, or the model fit for the representative events shown in Figs D1–D7, we cannot see any signs of uncaptured \( Q \) (i.e. slope) variations in the proposed method. Anyhow, these variations in the slope for the residual spectra \( R_{95} \) obtained in the model calibration step (see Step 3 in Fig. 5) could be studied further to see, for example, if the attenuation being less per unit travel time for more distant sensors (e.g. shown in Wang 2004; Eberhart-Phillips 2016) or if the slope of the residuals (i.e. \( R_{95} \) obtained in Step 3) for the ray paths are included in tomographic models the result may prevail interesting anomalies in the rock mass.

### 8.4 Near-field effect and P-wave residues

Source parameter estimation is usually done in the frequency domain. Here we use the S-wave spectra due to their relatively large energy content compared to P-wave spectra in general (see e.g. Haskell 1964; Hofstetter & Shapira 2000; Oye et al. 2005). However, the method could easily be extended to apply for P-wave spectra as well, at least for more distant sensors from the source where the measurement window for the P wave is larger and the P-wave near-field effects are less prominent (see e.g. Vidale et al. 1995; Yamada & Mori 2009). Other common near-field effects like: complex interference patterns and resonances (see e.g. Fig E2) have been reported in Bouchon & Aid (1977), flatter frequency responses (see e.g. Fig E1) in Mikumo & Miyatake (1978) observing \( n = 1 \) and larger geometrical decays (Legrand & Delouis 1999) proposing \( \beta = 2 \); may be present in our data set as well. These effects would cause poor model behaviour for observed spectra in the proximity of the source if not compensated for.

The remaining part of the P wave in the observed S-wave spectra is usually neglected due to its low energy content (see e.g. Haskell 1964; Hofstetter & Shapira 2000). However, it will have a notable effect for sensors near the source (Kijko 1994) as the P wave usually contains higher frequencies than the S wave and may cause poor model fit at higher frequencies if not compensated for.

The model we propose here reduces the impact of these effects on estimated parameters with the following measures:

1. Weights the estimates from \( t_{13} \)th frequency coefficient in \( \mathcal{E}_P \) in (6) with the underlying uncertainties (\( \sigma_{t1} \) and \( \sigma_a \)) obtained from the posterior samples (see Fig. 3) of the hypocentre and origin-time distribution (Martinsson 2012), where:

   \[
   \sigma_{t1} \text{ represents the uncertainty in origin time (see Fig. 3), will down-weight high frequencies in spectra (6) as shown in Fig. 9, } \sigma_{t1} \text{ is the uncertainty in ray length between sensor and hypocentre.}
   
   \text{This uncertainty is given as the standard deviation of all credible distances between the } \mathcal{E}_P \text{ and the credible hypocentre locations given by the posterior MCMC samples shown in Fig. 3. For sensors near the source, this uncertainty is large relative to its distance, and will therefore downweight the estimates in (6) from spectra in the proximity of the source as shown in Fig. 9.}
   
2. Spatial- and size-dependent empirical compensation will partly compensate for these challenges (i.e. P-wave residues and
near-field effects). If similar sized training events near the target reveal similar behaviour then \( E_{ij} \) will describe them (see e.g. Figs E1 or E2). Otherwise, if there is large discrepancy between these training events then \( \sigma_{E_{ij}} \) will increase and therefore downweighted in the estimate.

(iii) Truncate the spectra above the upper cut-off frequency (explained more in Appendix A). This is primarily to increase the processing speed, but it also reduces possible impact of high-frequency artefacts from P-wave residue and uncaptured noises.

Additionally, for a properly configured measurement system, the sensors closer to the source are outnumbered by more distant sensors, relative to the source radius (Haenggi 2005; Martinsson & Jonsson 2018). Instead of empirically compensating the observed spectra, with respect to distance and size, an alternative approach for dealing with some of these challenges would be to remove sensors in the proximity of the source in the estimation process as in Oth et al. (2017), at the cost of losing valuable information. A more appealing approach would be to include the source radius as an additional term in (6) that depends on the distance to the source, similar to \( \sigma_{r_i}^2 \), parametrized by \( f_c \) and \( V(p) \) (Kijko 1994). This would downweight the nearest sensors even more and therefore reduce the impact of possible near-field effects. Possible near-field effects are today compensated by \( \sigma_{r_i} \), together with the size- and distance-dependent empirical compensation where we rely on the outnumbered assumption above. However, we have no evidence that supports increasing the uncertainties for sensors near the source further, since the confidence regions are already valid for these spectra (see e.g. Figs D1–D7). On the contrary, a valid description of the close-to-source spectra suggests that the measures taken in the list above are sufficient.

8.5 High frequency falloff rate \( n \)

In the calibration Step 1 and Step 2, when \( n \) is estimated using a flat positive prior, we obtain a fair amount of events with estimates of \( n < 1.5 \) shown as grey dots in Figs 6 and 7. These dots deviate significantly from the trend in the scatter between \( f_c \) and \( \Omega_0 \) with respect to the majority of the dots. They are also unrealistic according to Lee et al. (2003), arguing that the radiated energy remains finite only when \( n > 1.5 \) under certain assumptions. In the histogram of \( n \) in Fig. 13(f), we also see a fair amount of events (\( \approx 10 \) per cent) favouring a flatter spectra with values of \( n \) at its lower bound (i.e. \( n = 1.5 \)). If these low estimates of \( n \) are due to: near-field effects (Mikumo & Miyatake 1978); correlation between \( n \) and \( f_c \) (Ross & Ben-Zion 2016; Van Houtte & Denolle 2018); the rupture propagation (e.g. direction or velocity of the cracking) in relation to sensor placement (Kaneko & Shearer 2015); deficiencies in the measurement system (e.g. high-frequency artefacts for these sensors reported in Cochrane et al. (2014) and observed in Figs E3 and D7), we leave to further debate.

The model could be extended further in a hierarchical structure as in Van Houtte & Denolle (2018), to incorporate, for example near-field effects on \( n \) over distance (Mikumo & Miyatake 1978). Instead of treating \( n \) as a fixed effect it would be more informative to consider it as a random effect and applying, for example group-level regression (Kruschke 2014; Martinsson & Törnman 2019) where the location of the prior distribution of \( n \) could depend on distance.

Although, in terms of fit we only expect minor improvements by increasing the complexity as proposed, since part of these behaviours will be captured by the empirical compensation.

8.6 Spatial and size interpolation coefficients \( (c_1, c_2) \)

In Table 2, we see that using both size and distance interpolation in the calculation of the weights \( w_{ij} \) in (14) improves the overall model fit compared to using the average log residual spectra over all calibration events for each sensor (i.e. corresponding to \( c_1 = 0 \) and \( c_2 \to \infty \)). However, using the average is still a good choice compared to no empirical compensation (as shown in Table 2 labelled ‘None’), since it will results in memory efficient code and fast processing without the need to store all log residual spectra and avoid computationally heavy interpolation in (14). Using a more complex source model (e.g. Model B) will also improve the general model fit (see Table 2) but it will also impact the behaviour of the source model estimates (see Fig. 13), with increased number of low \( f_c \) estimates.

A few common alternatives of \( c_1 \) and \( c_2 \) are evaluated due to study limitations. We use inverse distance weighting because of the simplicity to implement and understand, remarking that further investigation on type of interpolation method would be beneficial to improve the model-fit. Here we simply evaluate \( c_1 \in \{0, 1, 2\} \) corresponding to: no distance weighting (i.e. average over all), and the commonly used inverse distance and inverse squared distance interpolation, respectively (see e.g. Martinsson (2012; Wettain & Martinsson 2014, applying these in similar context). The compensation coefficients seem to favour large values for \( c_1 \) and small values for \( c_2 \) in terms of model-fit (see Table 2), providing more weight on similar-sized events (\( c_2 \)) and spatially close events (\( c_1 \)) in the calibration set compared to the target event. Since the weighting is jointly combined in (14), and going to the extreme in the proposed direction for either \( c_1 \) or \( c_2 \) may cause poor fit for individual events, since the weighting may only rely on the closest or the most similar sized event in the calibration set. More complex alternatives like Gaussian process regression (or Kriging, Krige 1951) could also be considered including covariation dependencies on distance and size. Less complex alternatives like the nearest neighbour interpolation could also be considered to reduce computational costs.

8.7 Robustness and extreme values

The data used in both calibration and validation are of higher quality compared to the majority of the triggered events observed in the mine (see Section 3). Despite this, we still observe spikes, resonances and poor frequency behaviour in many individual spectra, indicating the importance of a robust likelihood function. Part of these poor behaviours is captured by including the noise \( N_{ij} \) and the empirical compensation \( E_{ij} \) in the combined model \( \mu_{ij} \) in (5) and its uncertainty (6). However, we still observe estimates of the normality parameter \( \nu < 30 \) (see Table 3) for some events, implying the necessity of allowing the data to adjust the tail width in our likelihood function. Robust likelihoods are needed here to reduce biased parameter estimates that would otherwise be the result of extreme values and outliers in the data set. Another alternative to the Student’s \( t \)-distribution used here, is to consider a mixture of two likelihoods with different dispersion as in Kruschke (2014), Martinsson & Gustafsson (2018) and where the mixing coefficient has the role of the normality parameter and is often called the contamination parameter. This approach may be beneficial from a computational perspective, avoiding the computationally expensive gamma functions in the \( t \)-distribution.
9 CONCLUSION

We present a fully automatic approach for estimating the model parameters describing the S-wave spectra. In comparison to traditional empirical Green’s function approaches (Mueller 1985; Hough 1997), our method works for earthquakes of any size where all spectra associated with an event are modelled jointly within a Bayesian framework.

For precise estimates we use a combined empirical Green’s function (CEGF), similar as proposed by Shearer et al. (2006), consisting of both the modelled attenuation together with the spatial and event-size dependent empirical compensation, where the model fit significantly improved when favouring data from training events of similar size in the proximity of the source. In addition to Shearer et al. (2006) our new CEGF: (1) do not truncate the calibration data but instead interpolate it with respect to distance and size between target event and calibration events and (2) describes the variation in the estimates of the empirical compensation for proper weighting of the spectral coefficients.

To obtain reliable estimates, we additionally include the underlying uncertainties in ray-length and origin-time obtained from the hypocentre and origin-time distribution (Martinsson 2012) in the Bayesian model. The proposed method is first tested on synthetic data and then calibrated and cross-validated using real mining induced events. The model gives a valid description of observed spectra and we found no evidence of model deficiencies with respect to frequency, distance or size of the earthquakes, using both PPCs and residual analysis. The model and underlying uncertainties are incorporated in a heavy-tailed likelihood function together with informative prior distributions (estimated from calibration data), that together provide a Bayesian solution for robust, precise and reliable estimates of S-wave spectra and corresponding model parameters.

ACKNOWLEDGEMENTS

The authors would like to thank Rachel Abercrombie, an anonymous reviewer, and Andrew Valentine for their constructive comments and suggestions, which significantly improved the manuscript. The authors acknowledge the financial support from Luossavuara-Kirunavaara AB (LKAB, Sweden) and strategic innovation program STRIM, a joint venture of VINNOVA, Formas and the Swedish Energy Agency.

DATA AVAILABILITY

The data underlying this paper is obtained from the seismic monitoring system in LKAB’s underground mine in Kiruna. The data are owned by the mining company and can be accessible by signing an NDA. This can be initiated by contacting either Mikael Myrzell or Matthias Wimmer, both employed at LKAB.

REFERENCES

Abercrombie, R.E., 1995. Earthquake source scaling relationships from -1 to 5 ml using seismograms recorded at 2.5-km depth, J. geophys. Res., 100(B12), 24015–24036, doi.org/10.1029/95JB02397.

Abercrombie, R.E., 2015. Investigating uncertainties in empirical green’s function analysis of earthquake source parameters, J. geophys. Res., 120(6), 4263–4277, doi.org/10.1002/2015JB011984.

Abercrombie, R.E., Bannister, S., Ristau, J. & Doser, D., 2016. Variability of earthquake stress drop in a subduction setting, the Hikurangi Margin, New Zealand, Geophys. J. Int., 208(1) 306–320, doi.org/10.1093/gji/ggw393.
APPENDIX A: ESTIMATION OF CUT-OFF-FREQUENCIES

The cut-off frequencies shown as magenta dots in Fig. 2 defines the frequency interval $f_{ij}$ in (A1) for jth spectrum

$$ f_{ij} \in \{ f_{ij} \mid S(A_{ij}(\hat{\phi}_j)) > e^{\xi_j(\hat{\phi}_j) + \delta N_j} \} \tag{A1} $$

where $\hat{\phi}_j$, $\hat{\xi}_j$ and $\delta N_j$ are estimates given by

$$ \hat{\phi}_j = \arg \min_{\phi} \sum_i \left| \log(Y_{ij}) - \mu_j(\phi_j) \right| \tag{A2} $$

$$ \hat{\xi}_j = \arg \min_{\xi} \sum_i \left| \log(N_{ij}) - \hat{N}_j(\hat{\xi}_j) \right| \tag{A3} $$

$$ \delta N_j = \frac{1}{f_{j-1} - f_{j+1}} \sum_i \left( \log(N_{ij}) - \hat{N}_j(\hat{\xi}_j) \right)^2 \tag{A4} $$

using $\mu_j$ in (F) and

$$ \hat{N}_j(\hat{\xi}_j) = \xi_j + \xi_2 \log(f_{ij}) \tag{A5} $$

where $\xi_j = [\xi_{1j}, \xi_2]$ is a parameter vector.
APPENDIX B: WEAKLY INFORMATIVE PRIORS USED IN ESTIMATION STEP 1 AND 2

In the model calibration steps (i.e. Step 1 and Step 2) we use weakly informative priors to find the informative priors used in the sequential steps (i.e. Step 3 and Step 4). The priors for the parameters in $\phi = \{\Omega_0, f_c, n, Q, \beta\}$ are here weakly informative with respect to: the event size (see Fig. 6), the literature (see Table 1), and the frequency bandwidth (with $f_{\text{max}} = 1000$ Hz).

$$\Omega_0 \sim U(0, 100),$$
$$f_c \sim U(0, 3000 = 3f_{\text{max}}),$$
$$n \sim U(0, 5),$$
$$Q \sim \text{Exp}(1/500),$$
$$\beta \sim U(0, 4),$$

where $U(a, b)$ denotes a uniformly distributed parameter between $a$ and $b$ and $\text{Exp}(\lambda)$ denotes exponentially distributed parameter with the expected value at $\lambda^{-1}$.

APPENDIX C: SUPPORTIVE FIGURES

To provide additional evidence to manuscript we include four supportive figures (i.e. Figs C1, C2, C3 and C4). Figs C1 and C2 containing all the waveform and residuals, respectively for the events shown in Fig. 2, Fig. 3 and in Fig. 12. Fig. C3 addresses scaling and stress drop with respect to $\Omega_0$, $f_c$ and $n$. Fig. C4 provide an additional synthetic example using heterogeneous assumptions for Model B.

APPENDIX D: REPRESENTATIVE EVENTS

Here we provide additional modelling results in Figs D1–D7 that are more representative of the entire validation data set and at different event sizes. The events are not picked to show challenges and strengths, but instead correspond to the specific event indices matching exactly the: 1st, 5th, 25th, 50th, 75th, 95th and 99th percentiles with respect to $\Omega_0$ in the entire validation data set. These results will provide a more representative picture of the modelling performance in general.

APPENDIX E: ADDITIONAL EVENTS

Here we provide additional modelling results in Figs E1–E3 that are picked to show additional challenges and strengths with the data and proposed method.

Figure C1. Amplitude normalized seismograms for the event shown in Fig. 2, Fig. 3 and in Fig. 12. The seismograms are plotted on the vertical axis at distance $r$ between sensors and hypocentre and coloured by peak amplitude. Black vertical line marks the origin time and the red seismograms highlight both traces shown in Fig. 2.
Figure C2. Residual plot using Model B for the event in Figs 2, 3, 12 and C1 where the radial and tangential axis in the polar plot shows the distance $r_j$ and azimuth (where $0^\circ$ and $90^\circ$ correspond to the $x$- and $y$-axis, respectively, in the mine’s coordinate system shown in Fig. 1), respectively with respect to the hypocentre of the triggers sensors (see Fig 3). Upper right-hand figure show the elevation between hypocentre and sensor (negative values correspond to sensors below the hypocentre), with respect to distance. Lower right-hand figure shows the residual spectra $R_{ij} = \log Y_{ij} - \mu_{ij}(\hat{\phi})$ for the event, where $\hat{\phi}$ is the posterior mode of $\phi$. All figure are coloured by the mean of the residual spectra (i.e. $\overline{R}_j = \text{mean}(R_{ij}) = \|i\|^{-1} \sum_{i\in\mathcal{I}} R_{ij}$) for corresponding $j$th sensor.
Figure C3. Estimates of $f_c^3 \times \Omega_0$ (∝ stress drop) for Model B with respect to location, size ($\Omega_0$), corner frequency ($f_c$) and falloff rate ($n$). Top figure shows the location of validation events. The scatter plots from left- to right-hand side shows posterior mode estimates for $\Omega_0$, $f_c$ and $n$ with respect to $f_c^3 \times \Omega_0$ together with pairwise correlations $\rho$ (with 95 per cent CI). All figures are colored by the estimate of $f_c^3 \times \Omega_0$ including the histogram of $f_c^3 \times \Omega_0$ to the right.
Figure C4. Bayesian inference of 10 synthetic spectra generated from an arbitrary point $p = [2223, 6247, -1048]$ to arbitrary sensors in the mine, with our heterogeneous velocity model $V(p, j)$ and empirical rock mass model described by $E(\phi, p, j)$ and $\sigma_{Eij}^2(\phi, p, j)$. The inference is provided with Model B. The lower left-hand subfigure depicts 10 noisy synthetic spectra $Y_{ij}$ coloured by distance $r_j$ to source, given predefined model parameters and noise characteristics ($\phi, N_{ij}$) defined in the legend. The scatter plots to the right of the diagonal shows the MCMC samples, and also depict the pairwise correlations $\rho$ (with 95 per cent CI) of credible values $\Omega_0, f_c$ and $n$ in $\phi$. Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The centre plot in the last row show the true source spectra $S_i$ (in cyan) shown together with the estimated source spectra $\hat{S}_i = S_i(\hat{\phi})$ (in black) and 95 per cent HDI (in grey). The plots in the right-hand most column are the synthetic spectra for closest sensor (this sensor is not triggered by the training data set, since $E = 0$), middle sensor and furthest sensor from top to bottom coloured by distance, together with the estimate of the observed spectra $\hat{Y} = \exp(\mu_{ij}(\hat{\phi}))$ (in black) using the posterior mode $\hat{\phi}$ and 95 per cent HDIs (in grey). Histograms for the statistical parameters $\nu$ and $\sigma_0^2$ are also provided left of the diagonal.
Figure D1. Bayesian inference of the event index corresponding to the 1st percentile with respect to \( \Omega_0 \) in the entire validation data set. Estimated using Model B, where \( c_1 = 2 \) and \( c_2 = 1 \) for the weighting of \( E_{ij} \). The lower left-hand corner depicts the observed spectra \( Y_{ij} \) coloured by distance \( r_j \) to the estimated hypocentre \( \hat{\rho} \). The scatter plots to the right of the diagonal shows the MCMC samples, and also depict the pairwise correlations \( \rho \) (with 95 per cent CI) of credible values of \( \Omega_0, f_c \) and \( n \) in \( \phi \). Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra \( \hat{S}_i = S_i(\hat{\phi}) \) (in black) and 95 per cent HDI (in grey), where \( \hat{\phi} \) is the posterior mode. The plots in the right-hand most column are the observed spectra for closest sensor, middle sensor and furthest sensor coloured by distance, together with the estimate of the observed spectra \( \hat{Y} \) (in black) and 95 per cent HDIs (in grey). Histograms for the statistical parameters \( \nu \) and \( \sigma_0^2 \) are also provided left of the diagonal.
Figure D2. Bayesian inference of the event index corresponding to the 5th percentile with respect to $\Omega_0$ in the entire validation data set. Estimated using Model B, where $c_1 = 2$ and $c_2 = 1$ for the weighting of $E_0$. The lower left-hand corner depicts the observed spectra $Y_{ij}$ coloured by distance $r_j$ to the estimated hypocentre $\hat{p}$. The scatter plots to the right of the diagonal shows the MCMC samples, and also depict the pairwise correlations $\rho$ (with 95 per cent CI) of credible values of $\Omega_0$, $f_c$ and $n$ in $\phi$. Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra $\hat{S}_i = S_i(\hat{\phi})$ (in black) and 95 per cent HDI (in grey), where $\hat{\phi}$ is the posterior mode. The plots in the right-hand most column are the observed spectra for closest sensor, middle sensor and furthest sensor coloured by distance, together with the estimate of the observed spectra $\hat{Y}_i$ (in black) and 95 per cent HDIs (in grey). Histograms for the statistical parameters $\nu$ and $\sigma_0^2$ are also provided left of the diagonal.
Figure D3. Bayesian inference of the event index corresponding to the 25th percentile with respect to $\Omega_0$ in the entire validation data set. Estimated using Model B, where $c_1 = 2$ and $c_2 = 1$ for the weighting of $E_{ij}$. The lower left-hand corner depicts the observed spectra $Y_{ij}$ coloured by distance $r_j$ to the estimated hypocentre $\hat{p}$. The scatter plots to the right of the diagonal show the MCMC samples, and also depict the pairwise correlations $\rho$ (with 95 per cent CI) of credible values of $\Omega_0$, $f_c$ and $n$ in $\phi$. Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra $\hat{S}_i = S_i(\hat{\phi})$ (in black) and 95 per cent HDI (in grey), where $\hat{\phi}$ is the posterior mode. The plots in the right-hand most column are the observed spectra for closest sensor, middle sensor and furthest sensor coloured by distance, together with the estimate of the observed spectra $\hat{Y}$ (in black) and 95 per cent HDIs (in grey). Histograms for the statistical parameters $\nu$ and $\sigma_0^2$ are also provided left of the diagonal.
Figure D4. Bayesian inference of the event index corresponding to the 50th percentile with respect to $\Omega_0$ in the entire validation data set. Estimated using Model B, where $c_1 = 2$ and $c_2 = 1$ for the weighting of $E_{ij}$. The lower left-hand corner depicts the observed spectra $Y_{ij}$ coloured by distance $r_j$ to the estimated hypocentre $\hat{p}$. The scatter plots to the right of the diagonal shows the MCMC samples, and also depict the pairwise correlations $\rho$ (with 95 per cent CI) of credible values of $\Omega_0$, $f_c$ and $n$ in $\phi$. Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra $\hat{S}_i = S_i(\hat{\phi})$ (in black) and 95 per cent HDI (in grey), where $\hat{\phi}$ is the posterior mode. The plots in the right-hand most column are the observed spectra for closest sensor, middle sensor and furthest sensor coloured by distance, together with the estimate of the observed spectra $\hat{Y}$ (in black) and 95 per cent HDIs (in grey). Histograms for the statistical parameters $\nu$ and $\sigma^2_0$ are also provided left of the diagonal.
Figure D5. Bayesian inference of the event index corresponding to the 75th percentile with respect to $\Omega_0$ in the entire validation data set. Estimated using Model B, where $c_1 = 2$ and $c_2 = 1$ for the weighting of $E_{ij}$. The lower left-hand corner depicts the observed spectra $Y_{ij}$ coloured by distance $r_j$ to the estimated hypocentre $\hat{p}$. The scatter plots to the right of the diagonal shows the MCMC samples, and also depict the pairwise correlations $\rho$ (with 95 per cent CI) of credible values of $\Omega_0$, $f_c$ and $n$ in $\phi$. Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra $\hat{S}_i = S_i(\phi)$ (in black) and 95 per cent HDI (in grey), where $\phi$ is the posterior mode. The plots in the right-hand most column are the observed spectra for closest sensor, middle sensor and furthest sensor coloured by distance, together with the estimate of the observed spectra $\hat{Y}$ (in black) and 95 per cent HDIs (in grey). Histograms for the statistical parameters $\nu$ and $\sigma^2_0$ are also provided left of the diagonal.
Figure D6. Bayesian inference of the event index corresponding to the 95th percentile with respect to $\Omega_0$ in the entire validation data set. Estimated using Model B, where $c_1 = 2$ and $c_2 = 1$ for the weighting of $E_{ij}$. The lower left-hand corner depicts the observed spectra $Y_{ij}$ coloured by distance $r_j$ to the estimated hypocentre $\hat{p}$. The scatter plots to the right of the diagonal shows the MCMC samples, and also depict the pairwise correlations $\rho$ (with 95 per cent CI) of credible values of $\Omega_0$, $f_c$ and $n$ in $\phi$. Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra $\hat{S}_i = S_i(\hat{\phi})$ (in black) and 95 per cent HDI (in grey), where $\phi$ is the posterior mode. The plots in the right-hand most column are the observed spectra for closest sensor, middle sensor and furthest sensor coloured by distance, together with the estimate of the observed spectra $\hat{Y}$ (in black) and 95 per cent HDIs (in grey). Histograms for the statistical parameters $\nu$ and $\sigma_0^2$ are also provided left of the diagonal.
Figure D7. Bayesian inference of the event index corresponding to the 99th percentile with respect to $\Omega_0$ in the entire validation data set. Estimated using Model B, where $c_1 = 2$ and $c_2 = 1$ for the weighting of $E_{ij}$. The lower left-hand corner depicts the observed spectra $Y_{ij}$ coloured by distance $r_j$ to the estimated hypocentre $\hat{p}$. The scatter plots to the right of the diagonal shows the MCMC samples, and also depict the pairwise correlations $\rho$ (with 95 per cent CI) of credible values of $\Omega_0$, $f_c$ and $n$ in $\phi$. Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra $\hat{S}_i = S_i(\hat{\phi})$ (in black) and 95 per cent HDI (in grey), where $\hat{\phi}$ is the posterior mode. The plots in the right-hand most column are the observed spectra for closest sensor, middle sensor and furthest sensor coloured by distance, together with the estimate of the observed spectra $\hat{Y}$ (in black) and 95 per cent HDIs (in grey). Histograms for the statistical parameters $\nu$ and $\sigma_0^2$ are also provided left of the diagonal.
Figure E1. Bayesian inference of one validation event with flat spectra. Estimated using Model B, where $c_1 = 2$ and $c_2 = 1$ for the weighting of $E_{ij}$. The lower left-hand corner depicts the observed observed spectra $Y_{ij}$ coloured by distance $r_j$ to the estimated hypocentre $\hat{p}$. The scatter plots to the right of the diagonal shows the MCMC samples, and also depict the pairwise correlations $\rho$ (with 95 per cent CI) of credible values of $\Omega_0$, $f_c$ and $n$ in $\phi$. Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra $\hat{S}_i = S_i(\hat{\phi})$ (in black) and 95 per cent HDI (in grey), where $\phi$ is the posterior mode. The plots in the right-hand most column are the observed spectra for closest sensor, middle sensor and furthest sensor coloured by distance, together with the estimate of the observed spectra $\hat{Y}$ (in black) and 95 per cent HDIs (in grey). Histograms for the statistical parameters $\nu$ and $\sigma_0^2$ are also provided left of the diagonal.
Figure E2. Bayesian inference of one validation event containing resonances for the closest sensor to source. Estimated using Model B, where $c_1 = 2$ and $c_2 = 1$ for the weighting of $E_{ij}$. The lower left-hand corner depicts the observed spectra $Y_{ij}$ coloured by distance $r_j$ to the estimated hypocentre $\hat{p}$. The scatter plots to the right of the diagonal shows the MCMC samples, and also depict the pairwise correlations $\rho$ (with 95 per cent CI) of credible values of $\Omega_0, f_c$ and $n$ in $\phi$. Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra $\hat{S}_i = S_i(\hat{\phi})$ (in black) and 95 per cent HDI (in grey), where $\hat{\phi}$ is the posterior mode. The plots in the right-hand most column are the observed spectra for closest sensor, middle sensor and furthest sensor coloured by distance, together with the estimate of the observed spectra $\hat{Y}$ (in black) and 95 per cent HDIs (in grey). Histograms for the statistical parameters $\nu$ and $\sigma_0^2$ are also provided left of the diagonal.
Figure E3. Bayesian inference of one large validation event showing poor high frequency behaviour. Estimated using Model B, where $c_1 = 2$ and $c_2 = 1$ for the weighting of $E_{ij}$. The lower left-hand corner depicts the observed observed spectra $Y_{ij}$ coloured by distance $r_j$ to the estimated hypocentre $\hat{p}$. The scatter plots to the right of the diagonal shows the MCMC samples, and also depict the pairwise correlations $\rho$ (with 95 per cent CI) of credible values of $\Omega_0, f_c$ and $n$ in $\phi$. Diagonal figures are histograms of posterior samples with corresponding 95 per cent HDI. The middle plot in the last row shows the estimated source spectra $\hat{S}_i = S_i(\hat{\phi})$ (in black) and 95 per cent HDI (in grey), where $\hat{\phi}$ is the posterior mode. The plots in the right-hand most column are the observed spectra for closest sensor, middle sensor and furthest sensor coloured by distance, together with the estimate of the observed spectra $\hat{Y}$ (in black) and 95 per cent HDIs (in grey). Histograms for the statistical parameters $\nu$ and $\sigma_0^2$ are also provided left of the diagonal.