Fully probabilistic seismic source inversion – Part 2: Modelling errors and station covariances

Simon C. Stähler[1,2,3] **and Karin Sigloch**[4,1]

[1]Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität (LMU), Theresienstr. 41, 80333 Munich, Germany
[2]Munich Centre of Advanced Computing, Department of Informatics, Technische Universität München, Munich, Germany
[3]Leibniz Institute for Baltic Sea Research (IOW), Seestr. 15, 18119 Rostock, Germany
[4]Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, UK

Correspondence to: Simon C. Stähler (mail@simonstaehler.com)

Abstract. Seismic source inversion, a central task in seismology, is concerned with the estimation of earthquake source parameters and their uncertainties. Estimating uncertainties is particularly challenging because source inversion is a non-linear problem. In a companion paper, Stähler and Sigloch (2014) developed a method of fully Bayesian inference for source parameters, based on measurements of waveform cross-correlation between broadband, teleseismic body-wave observations and their modelled counterparts. This approach yields not only depth and moment tensor estimates but also source time functions.

A prerequisite for Bayesian inference is the proper characterisation of the noise afflicting the measurements, a problem we address here. We show that, for realistic broadband body-wave seismograms, the systematic error due to an incomplete physical model affects waveform misfits more strongly than random, ambient background noise. In this situation, the waveform cross-correlation coefficient CC, or rather its decorrelation $D = 1 - \mathrm{CC}$, performs more robustly as a misfit criterion than ℓ^p norms, more commonly used as sample-by-sample measures of misfit based on distances between individual time samples.

From a set of over 900 user-supervised, deterministic earthquake source solutions treated as a quality-controlled reference, we derive the noise distribution on signal decorrelation $D = 1 - \mathrm{CC}$ of the broadband seismogram fits between observed and modelled waveforms. The noise on D is found to approximately follow a log-normal distribution, a fortunate fact that readily accommodates the formulation of an empirical likelihood function for D for our multivariate problem. The first and second moments of this multivariate distribution are shown to depend mostly on the signal-to-noise ratio (SNR) of the CC measurements and on the back-azimuthal distances of seismic stations. By identifying and quantifying this likelihood function, we make D and thus waveform cross-correlation measurements usable for fully probabilistic sampling strategies, in source inversion and related applications such as seismic tomography.

1 Introduction

The quantitative estimation of seismic source characteristics is one of the most important inverse problems in geophysics, from both scientific and societal points of views. Source parameters not only can be used to locate earthquakes and to understand earthquake mechanisms and their implications for tectonic settings and seismic hazard, but they are also important in seismic tomography, where accurate source information is a prerequisite for achieving optimal fits between observed and modelled (waveform) data.

Estimation of seismic source parameters includes an earthquake's location, depth, fault plane and temporal rupture evolution. The inverse problem is non-linear, and parameter correlations result in trade-offs and non-uniqueness, e.g. the correlation between dip and scalar moment that was discovered by Kanamori and Given (1981). Source depth is

Table 1. Symbols frequently used in this paper

$\boldsymbol{m}$	Model vector of earthquake source parameters (M-dimensional) as defined in Stähler and Sigloch (2014): earthquake depth, moment tensor and source time function
$\boldsymbol{d}$	Geophysical data vector (N-dimensional)
M	Number of model parameters
N	Number of data
$g(\boldsymbol{m})$	Forward operator acting on a model parameter vector $\boldsymbol{m}$
$\mathcal{L}(\boldsymbol{m}\|\boldsymbol{d})$	Likelihood of a model $\boldsymbol{m}$ given the data $\boldsymbol{d}$
$\mathcal{L}^*(\boldsymbol{m}\|\boldsymbol{d})$	Likelihood-equivalent function of a model $\boldsymbol{m}$ given the data $\boldsymbol{d}$, constructed from the distribution of misfit values. Termed "empirical likelihood" in this article.
$\mathbf{S}_D$	Data covariance matrix
Φ	Total misfit of one model $\boldsymbol{m}$ and data $\boldsymbol{d}$, $\Phi = -\ln\mathcal{L}$
Φ^W	Misfit between one recorded and predicted seismogram
$\boldsymbol{u}_{j,i,i=\{1,\dots,n_j\}}$	Time-discrete seismogram j in a time window around a phase of interest.
$\boldsymbol{u}^c_{j,i,i=\{1,\dots,n_j\}}$	Synthetic seismogram j predicted by a model $\boldsymbol{m}$ and a forward operator $g(\boldsymbol{m})$
j	Index of seismogram time window
i	Index of sample in seismogram time window
n_j	Number of samples in a time window j
n_S	Number of time windows. $n_S \equiv N$ if the decorrelation misfit is used.
$n_{\text{tot}} = \sum_{j=1}^{n_S} n_j$	Total number of samples in all time windows. $n_{\text{tot}} \equiv N$ if a ℓ^p misfit is used.
$CC_k^{\boldsymbol{u}_i,\boldsymbol{u}_i^c}$	(Normalised) cross-correlation function between time series $\boldsymbol{u}$ and $\boldsymbol{u}^c$ using a window function w_i: $CC_k^{\boldsymbol{u}_i,\boldsymbol{u}_i^c} = \dfrac{\sum_{i=1}^{n}\left(w_i u_{i-k}^c \cdot u_i\right)}{\sqrt{\sum_{i=1}^{n}(w_i u_{i-k}^c)^2 \cdot \sum_i^n (w_i u_i)^2}}$
$CC^{\boldsymbol{u}_i,\boldsymbol{u}_i^c}$	Maximum of $CC_k^{\boldsymbol{u}_i,\boldsymbol{u}_i^c}$ over k; the "correlation between $\boldsymbol{u}_i$ and $\boldsymbol{u}_i^c$"
$D^{\boldsymbol{u}_i,\boldsymbol{u}_i^c}$	Decorrelation, $D^{\boldsymbol{u}_i,\boldsymbol{u}_i^c} = 1 - \max_k\{CC_k^{\boldsymbol{u}_i,\boldsymbol{u}_i^c}\}$
α	Coefficient for the level of waveform perturbation in the synthetic tests described in Sect. 2.4
β	Coefficient for the level of background noise in said test

a particularly challenging parameter; for example Sigloch and Nolet (2006) often find multiple local minima in waveform data misfits as a function of depth, even when source time functions (STFs) are explicitly estimated. This makes global search methods and ensemble sampling particularly attractive if the associated computational hurdles can be surmounted. For finite-fault inversion of large earthquakes, Bayesian methods have been developed in recent years (Duputel et al., 2012, 2014; Dettmer et al., 2014), as they also have been for non-kinematic inversions of regional events (Mustać and Tkalčić, 2016), but we focus on the inversion of source time functions of intermediate-sized events (m_b 5.5 to 7.5) from broadband, teleseismic waveforms.

In a companion paper (Stähler and Sigloch, 2014), we developed the PRobabilistic Interference of Source Mechanisms (PRISM) algorithm, a fully probabilistic inversion for source depth, moment tensor and STF, via sampling by both stages of the neighbourhood algorithm (NA; Sambridge, 1999). Figure 1 sums up the procedure and its results.

The need for PRISM arose from our work in global-scale waveform tomography, which fits broadband body-wave seismograms of moderate to large earthquakes to modelled synthetics, up to the highest occurring frequencies

($\approx 1\,\text{Hz}$). This can only be achieved with good a priori estimates of source depth, which strongly shapes the synthetic Green's functions, and of source time functions, which convolve the Green's functions. At the time, no data centre delivered routine estimates of broadband STFs (by now, efforts other than ours are underway; Vallée et al., 2011; Vallée and Douet, 2016). Hence Sigloch and Nolet (2006) developed a linearised, iterative approach that semi-automatically deconvolved broadband source time functions, source depths and moments tensors of more than 2000 earthquakes, which were subsequently used in several waveform tomographies (Sigloch et al., 2008; Sigloch, 2011; Sigloch and Mihalynuk, 2013; Hosseini and Sigloch, 2015).

The required human supervision time called for full automatisation, preferably in a Bayesian setting that would circumvent the occasional divergence of the non-linear optimisation and would automatically diagnose parameter trade-offs of the kind described. PRISM (Stähler and Sigloch, 2014) solved this problem, but we left the justification of its misfit criterion and the derivation of its noise model and likelihood function to the present study.

To render ensemble sampling with the NA computationally feasible, the dimensionality of the model parameter
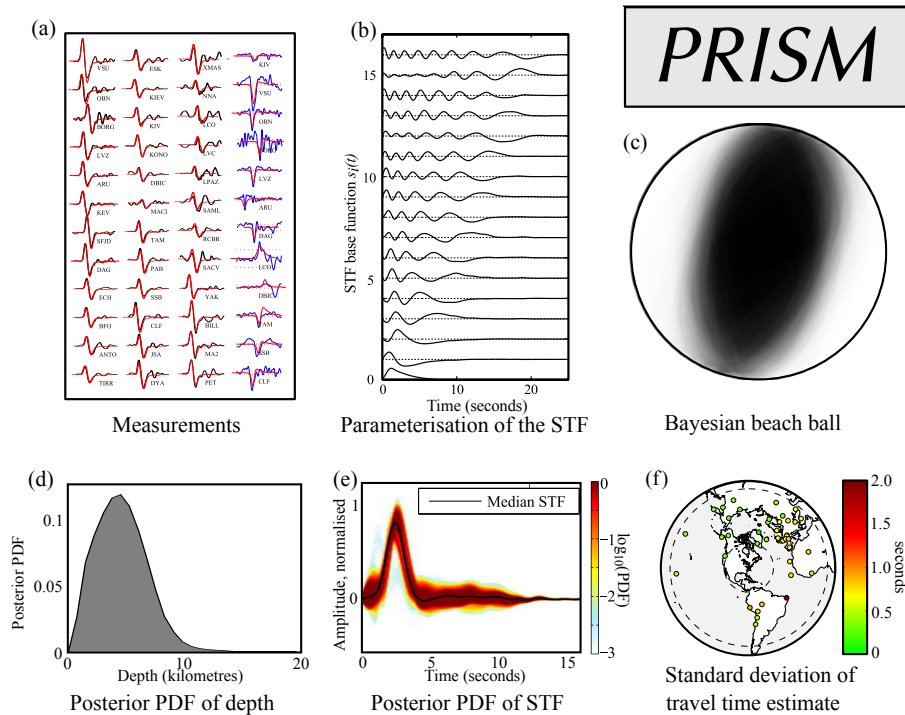
Figure 1. Visual summary of the fully probabilistic source inversion algorithm PRISM presented in the companion paper (Stähler and Sigloch, 2014), on the example of a magnitude-5.7 earthquake in the US state of Virginia on 23 August 2011. **(a)** Candidate source solutions are evaluated according to the cross-correlation fit they produce between observed broadband, teleseismic P waveforms (black) or SH waveforms (blue), and their modelled counterparts (red). The present study is concerned with quantifying the noise distribution on these cross-correlation measurements CC – one scalar per source–receiver pair, 48 in total for this earthquake. **(b)** To reduce the dimensionality of the model space to a number accessible to Bayesian sampling, the source time function (STF) is parameterised as a linear combination of 15 empirical orthogonal functions found to best span the space of a large set of 900 reference STFs (Sigloch and Nolet, 2006; Stähler and Sigloch, 2014). **(c)** The "Bayesian beach ball", a visual average of the posterior ensemble of well-fitting solutions, conveys not only the nature of the moment tensor but also the magnitude and nature of its uncertainties. **(d)** The marginal probability of the hypocentre depth. **(e)** Weighted average of STFs from the posterior ensemble of good solutions permits assessment of the uncertainties in STF shape. This STF is clearly unimodal and of less than 5 s duration. **(f)** As a secondary benefit, this procedure yields the uncertainties (standard deviations) of cross-correlation travel time measurements at all stations, and their inter-station correlations. Travel times are the primary input data for seismic tomography, and these insights into their uncertainties are not readily available from other methods.

space has to be as small as possible, preferably less than 20. Depth is one parameter, and a normalised description of the moment tensor requires five more (a more rigorous and uniform parameterisation of the moment tensor has been derived by Tape and Tape, 2015, 2016). Although latitude and longitude could easily be added to this list, we do not consider them here, because the lateral location problem is adequately addressed by existing data centres (National Earthquake Information Center (NEIC) or Bondár and Storchak, 2011), and in any case we would re-estimate all hypocentres at the time of tomographic inversion. The STF is a high-dimensional parameter vector, which Sigloch and Nolet (2006) and Stähler et al. (2012) parameterised simply as a time series of 256 unknowns (10 Hz sampling rate, 25.6 s length). To reduce its dimensionality for Bayesian sampling, Stähler and Sigloch (2014) made use of a dataset of > 2000 deterministic earthquake source solutions (depth, moment tensor and

STF) obtained by Sigloch and Nolet (2006). We selected the 900 best-constrained STFs and composed this set into empirical orthogonal functions (EOFs), denoted $s_l(t)$. Any broadband STF $s(t)$ of events up to magnitudes of about 7.5 is well described by a linear combination of the first L EOFs, where $L \approx 15$ delivers sufficient accuracy for our purpose: $s(t) = \sum_{l=1}^{15} a_l s_l(t)$. These EOFs $s_l(t)$, shown in Fig. 1b, are the primary means by which we feed a priori expert knowledge into the Bayesian sampling problem. PRISM's STF parameterisation consists of the first L EOF weights a_l, bringing the total dimensionality of the parameter space to ≈ 20.

This space is sampled by both stages of the neighbourhood algorithm, resulting in an ensemble of source solutions $\boldsymbol{m}$ (cf. Table 1). From this ensemble, marginal probabilities for any model parameter can be estimated, e.g. for the depth (Fig. 1d) or the STF (Fig. 1e). As a visual means of conveying uncertainties in the moment tensor, we invented "Bayesian

beach ball" plots (Fig. 1c), a superposition of many beach ball representations in the a posteriori ensemble. A valuable side benefit is full uncertainties on travel time measurements ΔT_j at stations j. These travel time delays are incidental in the context of source inversion (as the time shifts between observed and synthetic seismograms that maximise the cross-correlation coefficients CC_j, Fig. 1f), but they represent the primary input data for our seismic waveform tomographies.

The primary measure of fit (or "input data") for PRISM's source inversions is the CC_j. When parameter estimation is performed as a deterministic optimisation problem, (only) a relative measure of fit or misfit is required: the optimal solution is the one that yields the smallest misfit between observations and model predictions, in our case the largest possible values of cross-correlation coefficients CC_j. By contrast, Bayesian parameter estimation requires not just a measure of misfit but also a likelihood function for it, which is derived from the probability distribution on the data (the "noise model"). In the absence of a noise model, the likelihood of a randomly drawn candidate solution cannot be evaluated. Obtaining a noise model for a misfit requires much more information about the measurement process and its statistics than the mere adoption of a misfit measure. This is the big challenge of Bayesian "inversion", which will be covered in this paper.

Section 2 argues for the adoption of the signal decorrelation $D = 1 - CC$ as a robust measure of misfit, where CC is the normalised cross-correlation coefficient (Table 1). To our knowledge, the decorrelation D of seismological waveforms has not been used as a misfit criterion in Bayesian inference (other than by Stähler and Sigloch, 2014) because its noise model and likelihood function were unknown – a shortcoming D shared with other deterministic misfit choices, such as the instantaneous phase coherence (Schimmel, 1999), time phase misfits (Kristekova et al., 2006) or multi-tapers (Tape et al., 2009).

Section 2.2 shows that the popular ℓ^2 and ℓ^1 norms (Mahalanobis, 1936) would be sub-optimal misfit criteria because noise in seismic signals is not simply additive Gaussian or Laplacian but rather partly signal-generated, i.e. highly correlated across time samples and stations, and better described by a transfer function. Figure 2 shows an example of this systematic noise "coda". Section 2.3 defines the general requirements of a good misfit criterion, and Sect. 2.4 demonstrates that the signal decorrelation D performs more robustly than sample-by-sample (ℓ^p) norms on realistic seismological waveform data.

To identify a likelihood function $\mathcal{L}(\boldsymbol{m}|\boldsymbol{d})$ of misfit D in Sect. 3, we draw once more on the prior knowledge contained in our set of deterministic source solutions for 900 earthquakes and on the 200 000 measurements of $CC = 1 - D$ made to obtain them. From this large, representative and highly quality-controlled dataset of confident source solutions, we obtain the statistics of the residual misfits D, which we use to construct an empirical likelihood $\mathcal{L}^*(\boldsymbol{m}|\boldsymbol{d})$. Thus
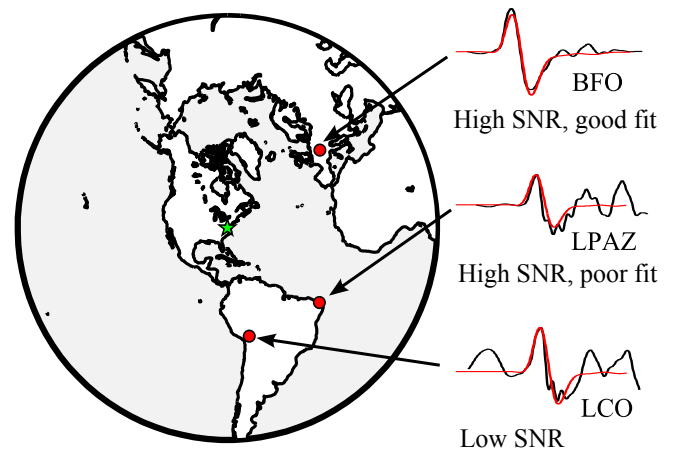

Figure 2. Three noise cases for compressional (P) waves in source inversion; the waveforms were produced by the M 5.7 earthquake in Virginia (23 August 2011). Station BFO has a high signal-to-noise ratio (no wiggles preceding the P pulse), and the waveform is fit well by a WKBJ synthetic using our best source solution for this earthquake. Station LPAZ has a high signal-to-noise ratio, but 3-D structure produces a strong coda following the P pulse, i.e. signal-generated, systematic "noise" not fit by the synthetic waveform. Station LCO has a low signal-to-noise ratio and a coda. Since the coda cannot be modelled, it must be considered noise, albeit of a systematic nature and correlated across time samples and across stations. By contrast, ambient noise is random and not correlated across stations, only across time samples (since the signal is bandlimited).

we can instruct the probabilistic inversion to explore subspaces of solutions $\boldsymbol{m}$ that yield similarly low levels of misfit D as these best-fitting deterministic solutions.

Section 3.6 presents a worked example for the construction of a likelihood function $\mathcal{L}(\boldsymbol{m}|\boldsymbol{d})$ from data of a typical earthquake, the 2011 Virginia event used throughout this paper and its companion Stähler and Sigloch (2014). We conclude with a discussion in Sect. 4.

2 Noise and misfit criteria

2.1 Bayesian inference

Bayesian inference estimates the posterior distribution $\pi(\boldsymbol{m})$ of the parameters $\boldsymbol{m}$ given $\boldsymbol{d}$, using the prior distribution $p(\boldsymbol{m})$ of the model parameters $\boldsymbol{m}$ and the likelihood $\mathcal{L}(\boldsymbol{m}|\boldsymbol{d})$ of the data $\boldsymbol{d}$, given the model $\boldsymbol{m}$, by applying Bayes' rule:

$$\pi(\boldsymbol{m}|\boldsymbol{d}) = \frac{1}{p(\boldsymbol{d})}\mathcal{L}(\boldsymbol{m}|\boldsymbol{d})\,p(\boldsymbol{m}). \tag{1}$$

$p(\boldsymbol{d})$ is the prior distribution of the data $\boldsymbol{d}$ and does not depend on the experiment. A likelihood function $\mathcal{L}(\boldsymbol{m}|\boldsymbol{d})$ is equivalent to the probability distribution $p(\boldsymbol{d}|\boldsymbol{m})$ of data $\boldsymbol{d}$ given the model parameters $\boldsymbol{m}$ (Gilks et al., 1996). It depends on the difference between measured data $\boldsymbol{d}$ and pre-

dicted data $g(\boldsymbol{m})$. This difference or misfit is defined, following convention, as

$$\Phi(\boldsymbol{d}, g(\boldsymbol{m})) = -\ln(\mathcal{L}(\boldsymbol{m}|\boldsymbol{d})), \tag{2}$$

so that a model with a high likelihood has a diminishing misfit. Since the likelihood of a model can vary by orders of magnitude, the logarithm brings the misfit back to natural scaling.

The exact formula for $\mathcal{L}(\boldsymbol{m}|\boldsymbol{d})$ depends on the assumed noise model and potential error sources in the forward model. Equation (2) requires that the misfit criterion take those into account as well. Next, we will show that this is straightforward only for specific assumptions about the noise, which are usually not realistic.

2.2 Metric-based misfit criteria

"Good" solutions $\boldsymbol{m}$ are associated with small misfits Φ, where the exact definition of Φ depends on the nature of the data $\boldsymbol{d}$, which may be hand-picked arrival times; dispersion curves; or, in our case, seismic displacement time series ("waveforms"). A waveform misfit is generally a functional $\Phi^W : \mathbb{R}^N \times \mathbb{R}^N \longmapsto [0, \infty)$ on $\boldsymbol{d}, g(\boldsymbol{m}) \in \mathbb{R}^N$.

The misfit functional has similar properties to a metric on $\mathbb{R}^N$, but it should be noted that there is no natural choice; rather, its choice implies a strong assumption of prior knowledge about the statistical properties of the noise on $\boldsymbol{d}$. In the case of seismic waveform data, the data vector $\boldsymbol{d}$ is the measured time-sampled seismogram $\boldsymbol{u}_i$ and the separate data are the samples $u_i, i = \{1, \dots, n\}$ of this time series. The vector $g(\boldsymbol{m})$ is the synthetic seismogram $\boldsymbol{u}_i^c, i = \{1, \dots, n\}$ predicted by the forward operator g for the model $\boldsymbol{m}$.

When the method of least squares is used to calculate the ℓ^2 misfit,

$$\Phi^W_{\ell^2}(\boldsymbol{m}|\boldsymbol{d}) = k' \left(\frac{1}{2} (\boldsymbol{d} - g(\boldsymbol{m}))^T \mathbf{S}_D^{-1} (\boldsymbol{d} - g(\boldsymbol{m})) \right), \tag{3}$$

the assumption is that the noise $\boldsymbol{\epsilon}$ is additive and Gaussian-distributed:

$$\boldsymbol{d} = g(\boldsymbol{m}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{S}_D). \tag{4}$$

The size $[N \times N]$ data covariance matrix $\mathbf{S}_D \in \mathrm{Sym}_N$ describes the correlation between the error of individual measurements d_i. k' is a normalisation constant.

In the case of a seismic waveform u_i, Φ^W is

$$\Phi^W_{\ell^2}(\boldsymbol{m}|\boldsymbol{d}) = k' \sum_{i=1}^{n} \sum_{i'=1}^{n} (u_i - u_i^c)^T (\mathbf{S}_D^{-1})_{i,i'} (u_{i'} - u_{i'}^c), \tag{5}$$

and $\mathbf{S}_D$ describes mainly the band-limited spectrum of environmental noise. Since a simple time shifting of u_i or u_i^c will violate the assumption of Eq. (4), the u_i or u_i^c need to be aligned first. Because we assume this noise to be time-invariant, we can build $\mathbf{S}_D$ from the autocorrelation function

$R_{\epsilon\epsilon}$ of the (discrete) noise time series ϵ_i. $\mathbf{S}_D$ is a Toeplitz matrix, where the rows are shifted instances of the autocorrelation function $R_{\epsilon\epsilon}$.

$$\mathbf{S}_{\mathbf{D},k,k+l} = R_{\epsilon\epsilon}(l) = \sum_{i}^{n} \epsilon_i \epsilon_{i-l} \tag{6}$$

See Bodin et al. (2012) for an example of how to construct $\mathbf{S}_D$ under the assumption of an autoregressive (AR) noise model.

For the estimation of the parameters $\boldsymbol{m}$ of one earthquake source, we would normally use seismograms measured at different stations, cut into a total of n_S time windows u_i, counted with index j. The overall misfit $\Phi(\boldsymbol{m})$ for a source solution will be comprised of the misfits of the single waveforms $\Phi^W_{\ell^2, j}(\boldsymbol{m})$. If the noise on each waveform j is assumed to be uncorrelated with the noise on all others, then it is legitimate to define the overall misfit as being simply additive:

$$\Phi(\boldsymbol{m}) = \sum_{j=1}^{n_S} \Phi^W_{\ell^2, j}(\boldsymbol{m}). \tag{7}$$

If the noise on the waveforms is correlated, then Eq. (3) has to be extended, such that $\boldsymbol{d}$, $\boldsymbol{m}$ and $\mathbf{S}_D$ contain all time samples of all waveforms recorded at different stations. This effort has – to our best knowledge – not been made in seismic inverse problems.

If each measurement i is considered to be uncorrelated with the others and has a variance σ_i, then $\mathbf{S}_D$ is a diagonal matrix with diagonal elements σ_i^2 and Eq. (3) reduces to

$$\Phi^W_{\ell^2}(\boldsymbol{m}|\boldsymbol{d}) = \frac{k'}{2} \sum_{i=1}^{N} \frac{(d_i - g_i(\boldsymbol{m}))^2}{\sigma_i^2} \tag{8}$$

or, in the case of waveforms,

$$\Phi^W_{\ell^2}(\boldsymbol{m}|\boldsymbol{d}) = \frac{k'}{2} \sum_{i=1}^{N} \frac{(u_i - u_i^c)^2}{\sigma_i^2}. \tag{9}$$

With a set of n_S waveforms $u_{i,j}$, the total misfit defined in Eq. (7) becomes

$$\Phi = \frac{k'}{2} \sum_{j=1}^{n_S} \sum_{i=1}^{n_j} \frac{(u_{i,j} - u_{i,j}^c)^2}{\sigma_i^2}, \tag{10}$$

the weighted least-squares criterion.

If the noise can be described well by the normal distribution, the ℓ^2 norm can be successfully applied. It is, however, very sensitive to data d_i deviating strongly from the prediction $g_i(\boldsymbol{m})$. Outlier samples can dominate the whole inversion process, while the residual misfit of almost-fitting parts of the waveform has no influence. Experience shows that realistic noise on seismic waveforms usually has more outliers than predicted by Eq. (4).

Hence, Käufl et al. (2013) have proposed to use the more outlier-resistant ℓ^1 norm as a misfit criterion of observed and modelled seismograms. They assume that noise on the time samples u_i is independently Laplace-distributed with width b_i, i.e. no temporal correlation:

$$\boldsymbol{d} = g(\boldsymbol{m}) + \boldsymbol{\epsilon}, \quad \epsilon_i \sim \text{Laplace}(0, b_i), \tag{11}$$

$$\Phi^W_{\ell^1}(\boldsymbol{m}|\boldsymbol{d}) = -\sum_i \frac{|d_i - g_i(\boldsymbol{m})|}{b_i} - \ln 2b_i. \tag{12}$$

Time samples of realistic, band-limited seismograms are strongly correlated, which calls for the use of multivariate Laplace distributions. This is the subject of ongoing research (Kotz et al., 2001; Kozubowski et al., 2013), but the resulting probability density functions (PDFs) are still too complex to be used in ensemble inference. To make things worse, seismograms recorded at different stations j will generally also be correlated. Hence the simplicity of the univariate Laplace distribution is not applicable, and the robustness of the ℓ^1 norm currently cannot be harnessed.

Other authors have proposed to use misfits based on general ℓ^p norms (e.g. $p = 1.5$ in Sambridge and Kennett, 2001), which allow the robustness of the misfit to be tuned to the noise on the data.

$$\Phi^W_{\ell^p}(\boldsymbol{m}|\boldsymbol{d}) = \left(\sum_{i=1}^n \frac{|d_i - g_i(\boldsymbol{m})|^p}{\sigma^p}\right)^{1/p} \tag{13}$$

The underlying noise model is an exponential power distribution. However, all problems described for the ℓ^1 norm apply here as well, and no multivariate forms exist in general.

In summary, it is tempting to chose ℓ^p misfits based on the time-sample-wise distance between observed and modelled waveforms because the underlying noise models are straightforward to state (uncorrelated or correlated Gaussian, uncorrelated Laplace distribution) and to translate into corresponding likelihood functions. Unfortunately, these noise models are very crude approximations of the pervasive noise characteristics and correlation found in real time series.

These serious shortcomings motivate our proposal of alternate misfit criteria.

2.3 Noise-model-based misfit

In a Bayesian context, the likelihood $\mathcal{L}(\boldsymbol{m}|\boldsymbol{d})$ is a defined by the noise model on the data. An equivalent function $\mathcal{L}^*(\boldsymbol{m}|\boldsymbol{d})$ can be constructed from the distribution $p(F)$ of any functional F of the observed and predicted waveforms $u_i, u_i^c \in \mathbb{R}$: $F : \mathbb{R} \times \mathbb{R} \longmapsto [0, \infty)$. In our attempt to move beyond F being a sample-wise distance between u_i and u_i^c, we generally want a candidate F to meet the following conditions:

1. For $u_i = u_i^c$, F should take a fixed value, say 0.

2. With decreasing similarity of u_i and u_i^c, F should increase, irrespective of the exact definition of similarity (Sect. 3 will consider this further).

3. F should be robust against time shifts $\Delta t = k \cdot dt$ or amplitude errors a affecting the waveform u_i, i.e. $F\left(a \cdot u_{i+k}, u_i^c\right) \cong F\left(u_i, u_i^c\right)$ for any $a \in \mathbb{R}, k \in \mathbb{N}$, because such unknown time shifts will affect real-world seismograms.

4. F should have discriminative power with respect to the model parameters $\boldsymbol{m}$, combined with robustness against realistic noise and theoretical errors.

Concerning the noise, we need to be able to calculate the distribution of F for a waveform afflicted by the typical three error sources: background noise, waveform modelling error and instrument error.

1. *Ambient noise* $\boldsymbol{\epsilon}_{\text{noise}}$: this is noise from man-made or natural sources around the receiver. It can be described very well by an additional term, like $\boldsymbol{\epsilon}_{\text{noise}} \sim \mathcal{N}(0, \mathbf{S})$ (see Eq. 5).

2. *Waveform modelling error* $\boldsymbol{T}_{\text{model},i}$: the synthetic waveform u_i^c can never be identical to the observed u_i, even in the absence of ambient noise. In the context of source modelling, the earth's impulse response (Green's function) can be considered a linear, time-invariant operator that acts on the source time function. The calculation of this Green's function is not perfect (e.g. due to errors in the earth model or imperfect computational methods). Tarantola and Valette (1982) called this the theoretical density function and proposed to model this systematic error by an additive term on u_i^c, but we think that it should rather take the form of a transfer function $\boldsymbol{T}_{\text{model},i}$, between u_i and u_i^c, which will hopefully be Dirac-like in character. However, $\boldsymbol{T}_{\text{model},i}$ will include the site response (receiver side reverberations), which can create strong waveform coda; see Fig. 2. Hence, $\boldsymbol{T}_{\text{model},i}$ could in practice be rather oscillatory.

3. *Instrument error* $\boldsymbol{T}_{\text{inst},i}$: a displacement seismogram u_i is assumed to have been corrected for the instrument response of its seismic sensor. In practice, this correction may be imperfect (Bogert, 1962), e.g. due to erroneous sensor metadata. We model this systematic error by another (hopefully Dirac-like) transfer function $\boldsymbol{T}_{\text{inst},i}$ convolving u_i.

In summary, the difference between a modelled u_i^c and observed waveform u_i is

$$u_i = u_i^c * \boldsymbol{T}_{\text{model},i} * \boldsymbol{T}_{\text{inst},i} + \boldsymbol{\epsilon}_{\text{noise},i}. \tag{14}$$

It is this complex mixture of noises that misfit criterion F should be robust against while retaining discriminatory power toward source model parameters $\boldsymbol{m}$.

Next, we will test the signal decorrelation D as an alternative to ℓ^p norms against these four criteria.

2.4 Signal decorrelation coefficient as a misfit

We choose the signal decorrelation D as a misfit criterion, defined as

$$D^{u_i, u_i^c} = 1 - \max_k \{CC_k^{u_i, u_i^c}\}, \qquad (15)$$

where

$$CC_k^{u_i, u_i^c} = \frac{\sum_{i=1}^n (w_i u_{i-k}^c \cdot u_i)}{\sqrt{\sum_{i=1}^n (w_i u_{i-k}^c)^2 \cdot \sum_{i=1}^n (w_i u_i)^2}} \qquad (16)$$

is the normalised cross-correlation coefficient and k is the time delay between u_i^c and u_i for which the normalised cross-correlation function $CC_k^{u_i, u_i^c}$ takes its maximum value. w_i is a window function that allows to select a time window for the cross-correlation measurement. D satisfies three of the four criteria that we desired of a misfit in the last section:

1. D^{u_i, u_i^c} takes the value 0 for identical signals $u_i^c \equiv u_i$, since $CC_{k=0}^{u_i, u_i^c} = 1$.

2. For $u_i \neq u_i^c$, $0 < D^{u_i, u_i^c} < 2$, i.e. D values larger than for the case $u_i^c \equiv u_i$, and D^{u_i, u_i^c} increases with decreasing similarity of u_i and u_i^c.

3. If a time shift k' is small compared to the window length, we have
   $CC_k^{u_i, u_i^c} \approx CC_{k+k'}^{u_i, u_{i+k'}^c}$ and thus $D^{u_i, u_i^c} \approx D^{u_i, u_{i+k'}^c}$.
   Due to the normalisation in Eq. (16), D is amplitude-independent:
   $CC^{u_i, u_i^c} = CC^{u_i, a \cdot u_i^c}$ and thus $D^{u_i, u_i^c} = D^{u_i, a \cdot u_i^c}$

The fourth criterion, discriminative power and robustness against noise is less straightforward to demonstrate. We proceed empirically by showing its superior performance over the ℓ^2 and ℓ^1 misfits on an example of the kind of waveforms we typically use for source inversion. Figure 3 shows in black a simulated, broadband, noise-free P wave train, recorded at $40°$ epicentral distance. The seismograms were modelled using the WKBJ method of Chapman (1978) in the IASP91 velocity model (Kennett and Engdahl, 1991), assuming an explosion source with $M_0 = 10^{20}$ Nm. Since the chosen source depth is shallow (10 km), the P pulse is followed within seconds by depth phases like pP, which effectively permits inversion for source depth. However, once this waveform gets perturbed by realistic modelling error (convolutive) and additive noise, resulting in the red waveform, the fit to the unperturbed original becomes tedious. A meaningful robustness test is as follows: if the perturbed (red) waveform is modelled for different candidate source depths, will the smallest misfit be achieved for the perturbed wave simulated at the correct depth of 10 km? This is a meaningful test of robustness, because source depth tends to be the most challenging parameter to retrieve in source inversions. Algorithmically, the perturbation is done in two steps:

1. Perturbation by convolution with a "modelling error function" $T_{\text{error},i}$, which encompasses effects of $T_{\text{model},i}$ and $T_{\text{inst},i}$. It is defined as having a unit amplitude spectrum and a random phase spectrum between 0 and $\alpha \cdot \pi/2$.

   $$u_{\text{m.e.}} = u_i^c * T_{\text{error},i} \qquad (17)$$

   This method adds realistic coda to the waveform, which simulates the effects of structure, that was not included in the forward simulation. The parameter α regulates the perturbing effect of the modelling error function.

2. By adding a band-limited noise term

   $$u_{\text{pert}} = u_{\text{m.e.}} + \beta \epsilon, \text{ where } \epsilon \sim \mathcal{N}(0, \mathbf{S}_D), \qquad (18)$$

   the covariance matrix $\mathbf{S}_D$ is set to model a band-limited noise with corner frequencies of $(1/15, 1/6\,\text{Hz})$, similar to microseismic background noise at the seismic station. The peak amplitude is normalised to that of u_i^c, so that the parameter β controls the relative amplitude of this noise term.

Figure 3 shows the resulting reference waveform (left) and perturbed waveforms for $\alpha = 0.4$ and $\beta = 0.8$, i.e. moderate perturbation of the signal and strong background noise. The unperturbed waveform u_i is plotted in solid, thin black, the waveform perturbed with modelling error $u_{\text{m.e.}}$ in dotted blue and the resulting reference trace in solid red. It bears little resemblance to the unperturbed waveform.

The right plot shows the value of the three waveform misfits ℓ^1, ℓ^2 and D between u_i^c and u_{pert} over varying source depths. It simulates an inversion for the depth of an earthquake using seismic waveforms. The waveform contains the P and pP arrival. The depth is mainly constrained by the relative arrival time of the three and the resulting waveform of the whole $P - pP$ wave train. The perturbation of Eq. (18) adds artificial coda with additional arrivals to the waveform, which a good waveform misfit should be robust against. The misfit should have a distinctively lower value for the "true" depth of 10 km than for any of the others. To take into account the stochastic nature of these perturbations, 500 realisations of u_{pert} were calculated for the same parameters, α and β, but with different random numbers. The coloured shades mark the 95 % (2σ) quantiles of the misfit values; the solid line marks the median.

The ℓ^2 misfit could not recognise u_i^c in u_{pert} anymore and assigns the lowest misfit to a depth of 3 km. An analysis of different noise and perturbation levels shows that the ℓ^2 norm is relatively robust against background noise, but not against perturbations from a modelling error; see Fig. S1 in the Supplement. This seems reasonable given the underlying noise model of this misfit.

The ℓ^1 norm does better, in that it has a minimum at 9 km depth, close to the true value. The zigzag shape however suggests that the value of 9 km is stochastic. The median value
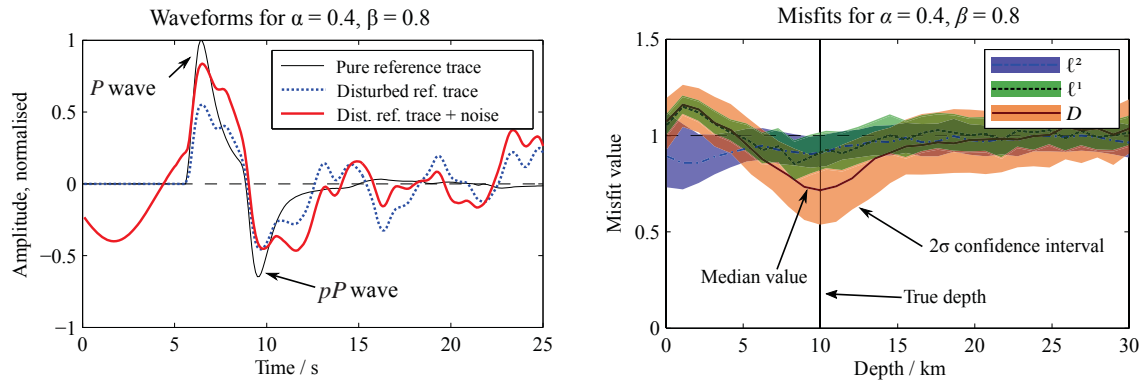
Figure 3. Comparison of the ℓ^1, ℓ^2 norm and the signal decorrelationComparison of the ℓ^1, ℓ^2 norm and the signal decorrelation $D = 1 - \mathrm{CC}$ as misfit criteria in noisy signals. A perturbed synthetic waveform u_{pert}^c for a 10 km deep explosion source, measured at a station at $40°$ epicentral distance, was compared to synthetic seismograms u^c for other depths, using the three misfit criteria. The shaded colours mark the 95 % quantiles of the misfit values, calculated by perturbing the reference waveform with different random seeds. The figure shows the relatively high robustness of the cross-correlation coefficient in recognising reference signals in perturbed measurements. For better visualisation, all misfit values have been normalised separately to have an average values of 1 between 20 and 30 km.
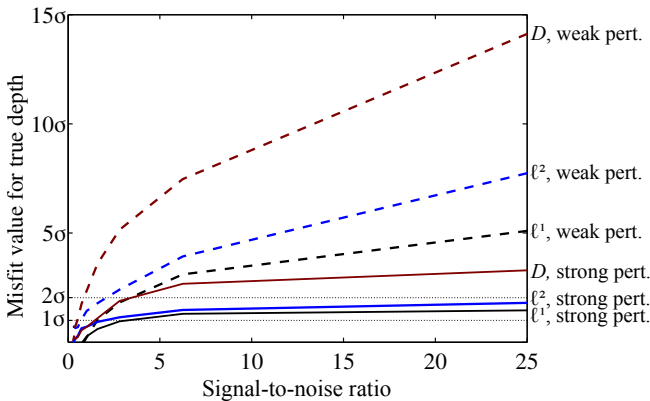

Figure 4. Distance between misfit value for the true source depth vs. the plateau for depths 20–30 km in standard deviations. See Fig. 3 for waveforms and misfit curves. The "weak-perturbation" curve is calculated with perturbation factor $\alpha = 0.1$, and the "strong-perturbation" curve with $\alpha = 0.9$ (see Eq. 17). For all SNR values, the decorrelation has a higher discriminative power than ℓ^1 or ℓ^2.

ℓ^2 and much more than for ℓ^1. To employ it in Bayesian inference, a detailed analysis of the statistical properties will be necessary. The analysis also shows that the actual values of D are influenced more strongly by the background noise level than by the modelling error. We will use that observation in Sect. 3.3.

Figure 4 compares the resolution power of the three misfits for different perturbation levels and signal-to-noise ratios (SNRs). It shows the difference between the misfit value for the true depth 10 km and the average misfit value for the depths between 20 and 30 km. The difference is expressed in numbers of standard deviations (sigmas) from the 500 separate noise realisations. The dashed line shows the result for weak perturbation ($\alpha = 0.1$), and the solid line for strong perturbation ($\alpha = 0.9$). It can be seen that, for strongly perturbed waveforms, the ℓ^1 and ℓ^2 norm cannot recognise the true depth with more than 2σ, even for high signal-to-noise ratios, while the decorrelation D stays well above 3σ, even for SNRs of 6.

3 Empirical likelihood function for the signal decorrelation

3.1 Empirical likelihood function obtained from high-quality, deterministic source estimates

In seismology, the cross-correlation coefficient $\mathrm{CC} = 1 - D$ has been used as a measure of goodness of fit to detect predicted waveforms in noisy signals (Sigloch and Nolet, 2006; Houser et al., 2008), to filter bad recordings, to detect temporal changes in repeating signals (e.g. Larose et al., 2010) and to estimate the spatial extents of earthquake clusters (Menke et al., 1990; Menke, 1999; Kummerow, 2010). It has rarely been used as a misfit criterion in source inversion – we are
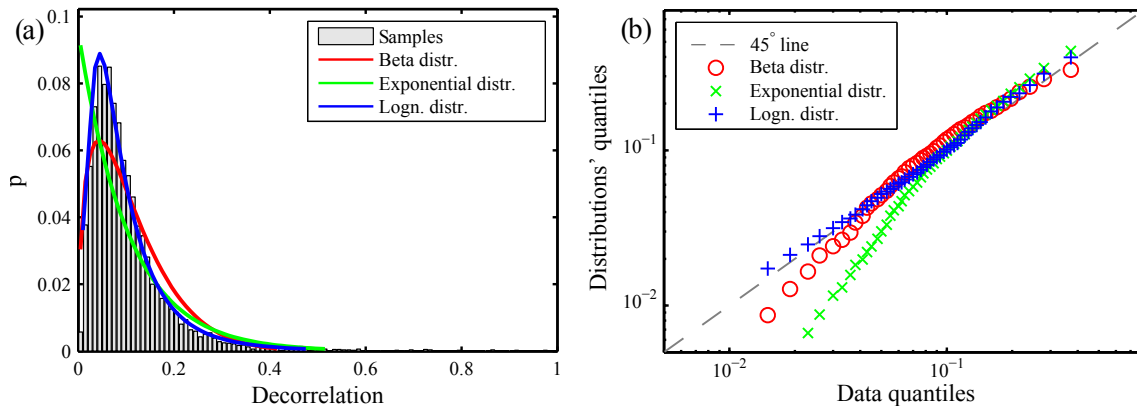
at 9 to 10 km reaches only slightly below the lower quartile for other depths, meaning that in reality the resolution power of the ℓ^1 norm for this kind of problem will be very limited. The studies for different noise and perturbation levels show that it is generally more robust against background noise and modelling error than the ℓ^2 norm but less so than the cross-correlation coefficient.

The cross-correlation misfit has the strongest difference between the plateau of wrong depth solutions and the true one. For low noise levels, the minimum is slightly wider than the one for the ℓ^1 norm. More values of α and β are shown in Fig. S1. The analysis of the confidence intervals shows that the values for CC scatter slightly more than the ones for

Figure 5. Probability distribution of D, the decorrelation of measured and synthetic P waveforms used for deterministic source inversions. **(a)** Empirical histogram of D is shown as grey bars. From 200 000 broadband, teleseismic P waveforms for 900 earthquakes, only waveforms with signal-to-noise ratios between 20.0 and 21.0 were considered for this figure (because the scaling parameters of analytic fitting functions depend mainly on SNR). Coloured lines show best-fitting realisations of three analytic probability density distributions: beta (red), exponential (green) and log-normal (blue). The log-normal distribution yields the best fit to data. **(b)** Quantile–quantile plot for the three candidate distributions of **(a)** confirms that the log-normal distribution best fits the empirical histogram of D. The values on the x axis are percentiles of the cumulative histogram of D in our dataset. The y axis shows the percentiles of the best-fitting distribution of each class. The closer the percentiles are to the line $y = x$, the better the fit of the distribution to the underlying data over the entire range of values. Both subfigures indicate that a log-normal distribution best fits the values of $D = 1 - \mathrm{CC}$.

only aware of Kikuchi and Kanamori (1991) and Marson-Pidgeon and Kennett (2000). CC and $D = 1 - \mathrm{CC}$ have not been used in probabilistic inversion, and the main obstacle would have been their unknown statistics.

We present an empirical solution to this problem by drawing on a large, pre-existing database of cross-correlation measurements that we assembled in the context of deterministic source inversions, as described in Section 1. Essentially we assert that our human expert knowledge and extensive experience have generated a large, representative and highly quality-controlled set of 900 teleseismic source parameter estimates that are sufficiently close to the true source parameters to reveal the statistics of the noise in the measurements $\boldsymbol{d}$ these estimates $\boldsymbol{m}$ are based upon. The measurements $\boldsymbol{d}$ consisted of 200 000 cross-correlation coefficients CC obtained from 200 000 broadband fits of observed seismograms to WKBJ synthetics. The synthetic waveforms were calculated using the WKBJ method (Chapman, 1978) in velocity model IASP91 (Kennett and Engdahl, 1991), with attenuation and density taken from PREM (Dziewoński and Anderson, 1981). To the extent that our source solutions $\boldsymbol{m}_j$ approach the true source parameters $\boldsymbol{m}_{0,j}$, the histogram of the CC (or $D = 1 - \mathrm{CC}$) values approximate the probability density function of CC (or D) in the presence of noise and modelling errors. Thus we can obtain an "empirical likelihood function" $\mathcal{L}^*(\boldsymbol{m}|\boldsymbol{d})$ even in the absence of an analytically describable noise model. We preface the term "likelihood" by "empirical" because strictly speaking the likelihood would be associated with the noise model on the raw samples i, rather than with the noise on the composite measure D. A similar approach has been adopted independently

and recently by Bodin et al. (2016) in the context of receiver-function inversion. Note that the term "empirical likelihood" has been used differently in statistics (Owen, 1988).

Our reasoning and procedure can be summed up as follows:

– We can consider the measurements of misfit functional $\Phi_j(\boldsymbol{m}_0|\boldsymbol{d})$ for one earthquake at $j = 1, \ldots, n_S$ recording receivers as realisations of a random process that follows a yet unknown probability density function $p(x)$. $\boldsymbol{m}_0$ are the true source parameters, and any misfit Φ_j is therefore due to ambient noise and modelling errors in the seismograms, as described in section 2.3.

– In practice we never get to know $\boldsymbol{m}_0$ but only a (hopefully close) estimate $\boldsymbol{m}_{\mathrm{est}}$, the result of a deterministic source inversion procedure. Hence all we can actually observe is $\Phi(\boldsymbol{m}_{\mathrm{est}}|\boldsymbol{d})$, some of which is due to the estimation error $\boldsymbol{m}_{\mathrm{est}} - \boldsymbol{m}_0$. However, by estimating $\boldsymbol{m}_{\mathrm{est}}$ carefully and repeatedly (for 900 different earthquakes), and by considering the resulting 900 sets of misfits Φ (at 200 000 source–receiver pairs) jointly, the histogram of their 200 000 D values should approximate a histogram of the true $\Phi(\boldsymbol{m}_0|\boldsymbol{d})$ as closely as we can hope to get. Figure 5a shows this empirically obtained histogram $\Phi_{\mathrm{cumulative}}$ of D in grey (for the subset of P seismograms that had a SNR of 20; reason to be discussed).

– To evaluate the likelihood of a misfit value Φ' encountered in a future (Bayesian) inversion, we could in principle compare it to this empirical histogram $\Phi_{\mathrm{cumulative}}$. It would however be more convenient and computationally efficient to identify an analytic expression for

the $p(x)$ that produced this histogram $\Phi_{\text{cumulative}}$ and to evaluate any Φ' against this $p(x)$.

– The best we can do is to identify a suitable type of distribution and fit its parameters to the empirical histogram $\Phi_{\text{cumulative}}$ of Fig. 5a, thus obtaining a PDF $p_{\text{fit}}(x)$ as our best estimate for the true $p(x)$.

– The likelihood of a data vector $\boldsymbol{d}$ given model $\boldsymbol{m}$ is then considered to be

$$\mathcal{L}^*(\boldsymbol{m}|\boldsymbol{d}) = p_{\text{fit}}(\Phi(\boldsymbol{d}|\boldsymbol{m})). \tag{19}$$

3.2 Approximate log-normal distribution of decorrelation D

We will consider three candidate distributions for fitting an analytic $p_{\text{fit}}(x)$: beta, exponential and log-normal. They are all positive one-sided (defined only for $D > 0$) and can take negligible values for $D > 2$, where strictly they should be 0. Figure 5a shows their fits to the empirical histogram after determining the best-fitting scale parameters for each.

The beta and the exponential distributions are seen to over-estimate the number of very small D values (i.e. values of $CC \approx 1$). Hence these distributions would predict more excellent waveform fits than observed. The likelihood of actually well-fitting waveforms would be estimated too low; i.e. we would be too pessimistic about the achievability of good waveform fits.

The log-normal distribution clearly yields the best approximation of the D histogram. This is confirmed by the quantile–quantile plot of Fig. 5b. Hence we choose the log-normal distribution to express our likelihood function.

The (univariate) log-normal distribution function is defined by two scale parameters μ and σ:

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right). \tag{20}$$

The log-normal distribution also yields the best fit to our synthetic data from Sect. 2.4, as calculated with the perturbations in Eqs. (17) and (18). See Fig. S4 for a corresponding quantile–quantile plot.

If random variable x in Eq. (20) is equated with the decorrelation D_j of one waveform j, the logarithm $\ln(D_j)$ is normally distributed with mean μ and standard deviation σ. This fortunate link of our empirical D histogram to the Gaussian distribution makes it trivial to express the joint, multivariate distribution of all n_S waveform measurements of an earthquake, collecting the D_j in vector $\boldsymbol{D}$ and the inter-station covariances in $n_S \times n_S$ covariance matrix $\mathbf{S}_D$.

The n_S-variate likelihood function for $\boldsymbol{D}$ becomes

$$\mathcal{L}_D^* = \frac{\exp\left(-\frac{1}{2}(\ln(\boldsymbol{D}) - \boldsymbol{\mu})^T \mathbf{S}_D^{-1} (\ln(\boldsymbol{D}) - \boldsymbol{\mu})\right)}{(2\pi)^{\frac{n}{2}}\sqrt{|\det(\mathbf{S}_D)|}}, \tag{21}$$
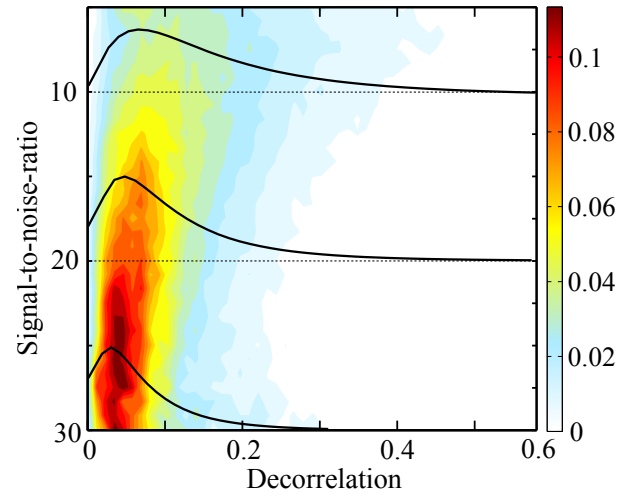

Figure 6. Colour shade map out a two-dimensional histogram of waveform decorrelation D, as a function of waveform SNR along the y axis. All 200 000 waveform measurements from our 900 deterministic source inversions entered this histogram. Black lines are the best-fitting log-normal distributions for SNRs of 10, 20 and 30. (The 1-D histogram for SNR = 20 was discussed in Fig. 5.) Toward smaller SNRs (high-noise conditions), the D distribution widens (more occurrences of poorly fitting waveforms).

and the misfit becomes

$$\begin{aligned}
\Phi &= \frac{1}{2}\left(\sum_{j=1}^n \sum_{k=1}^n (\ln(D_j) - \mu_j)^T (\mathbf{S}_D^{-1})_{jk} (\ln(D_k) - \mu_k)\right) \\
&+ \frac{1}{2}\ln\left((2\pi)^n |\det(\mathbf{S}_D)|\right).
\end{aligned} \tag{22}$$

This is the Mahalanobis distance, not between the individual samples of two waveforms $\boldsymbol{u}_i$ and $\boldsymbol{u}_i^c$ as in Eq. (3) but between the decorrelation D_j of these two waveforms and its expected value μ_j, taking into account correlated noise between two stations in $\mathbf{S}_D$.

Thus the use of D as a misfit criterion reduces the number of misfit values to n_S per earthquake (the number of source receiver paths, or waveforms) compared to $\sum_{j=1}^{n_S} n_j$ in the case of the ℓ^1 or ℓ^2 norms (n_j is the number of samples on waveform j). In other words, D_j itself accounts for any correlations across time samples on seismogram j and subsumes them into a single number, leaving only spatial (inter-station) correlations to be dealt with in $\mathbf{S}_D$ and in the empirical likelihood function $\mathcal{L}^*$.

3.3 Distribution coefficients determined by signal-to-noise-ratio

Here we describe how $\boldsymbol{\mu}$ and $\mathbf{S}_D$ can be estimated for one earthquake. So far it was implicitly assumed that a single distribution p_{fit} might fit $\Phi_{\text{cumulative}}$ for all source–receiver paths.

This may be an oversimplification since ambient noise levels ϵ_{noise} show significant diurnal and seasonal variations, and are elevated at stations close to coastlines or cities (Peterson, 1993; Stutzmann et al., 2009). Hence we might expect goodness of fit to vary across stations, which could be modelled by adjusting the scale parameters of the log-normal distribution for each station. Goodness of fit is also influenced by earthquake magnitude, and by station distance and back azimuth, so we might even require different scale parameters for each source–receiver pair.

To avoid this level of complexity, recall the investigation of Sect. 2.3 that revealed the distribution of D to be most sensitive to the level of ambient noise ϵ_{noise}. Hence we bin our 200 000 source–receiver pairs by SNR and estimate only one pair of (μ, σ) distribution parameters per SNR bin. This hopefully subsumes all individual sources of random misfit.

SNR is defined as the integrated spectral energy in the signal time window, divided by that of a 120 s noise window prior to the arrival of the first body-wave energy. Signal time windows $u_i, i = 1, \ldots, N_{signal}$ are as follows: for P phase, 5 s before to 20.6 s after its theoretical arrival time in IASP91, on the Z component; for SH phase, 10 s before to 41.2 s after, on the T component. Noise time windows $n_i, i = 1, \ldots, N_{noise}$ are as follows: for both P and SH phases, -150 to -30 s before theoretical arrival time. We calculate SNRs for P and SH waves as

$$\mathrm{SNR} = \frac{N_{noise} \sum_{i=1}^{N_{signal}} u_i^2}{N_{signal} \sum_{i=1}^{N_{noise}} n_i^2}. \tag{23}$$

Note that this way the noise window of the P wave measurement contains only ambient noise, whereas the SH wave noise window is in addition afflicted by some signal-generated noise: P coda and phases like PP or PcP, which get scattered into the transverse component due to lateral heterogeneities and anisotropy in the real earth.

Figure 6 shows the D histogram and three fitted probability densities $p_{fit}(D)$, as a function of SNR. Under low-noise conditions (high SNR), the log-normal distributions are narrower and centred on smaller D misfit values, which seems plausible.

By fitting functions of the form $h(\mathrm{SNR}) = a_1 + a_2 \cdot \exp(a_3 \cdot \mathrm{SNR})$ to the SNR-binned D histograms, we determined distribution parameters $\mu_P(\mathrm{SNR})$, $\mu_{SH}(\mathrm{SNR})$, $\sigma_P(\mathrm{SNR})$ and $\sigma_{SH}(\mathrm{SNR})$ for SNR ranging from 1 to 1000 for P waveforms and from 1 to 200 for SH waveforms (see Supplement for details).

Hence the log-normal distribution $p_{fit}(D)$ ascribed to a given source–receiver pair depends only on the ambient signal-to-noise ratio of the receiver i, and its scale parameters are given by

$$\mu_i = a_{\mu,1} + a_{\mu,2} \cdot \exp(a_{\mu,3} \cdot \mathrm{SNR}_i), \tag{24}$$

$$\sigma_i = a_{\sigma,1} + a_{\sigma,2} \cdot \exp(a_{\sigma,3} \cdot \mathrm{SNR}_i). \tag{25}$$
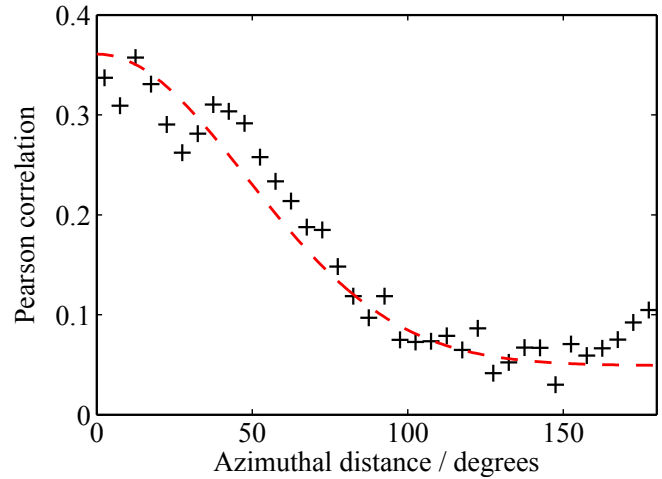

Figure 7. Correlation in misfit between neighbouring stations. The measured Pearson correlation (see Eq. 26) is plotted over the difference in azimuths between two station for the same earthquake. A fit function $g_{b_1,b_2,b_3}(\vartheta) = b_1 + b_2 \cdot \exp(-b_3 \vartheta^2)$ is plotted in dashed red lines.

The exact values for a_i depend on the velocity model and the solution method. Here, we used the WKBJ method, which results in a simplistic crustal response. Other methods, like the spectral-element method, in combination with a waveform database (as implemented in Instaseis by van Driel et al., 2015) may produce more realistic seismograms, resulting in higher average values of D. What matters is that the actual inversion uses exactly the same solver and velocity model as was used to determine the distributions of D.

3.4 Estimating inter-station covariances

Decorrelation values D measured at different stations cannot be expected to be uncorrelated, because systematic modelling errors (due to differences between assumed earth model and true earth, and to methodical inadequacies in the Green's function computations) will affect neighbouring stations in similar ways. A reasonable guess is that stations at similar azimuths from the source would show the strongest correlations because their wave paths have sampled similar parts of the sub-surface, in particular similar parts of the crust and upper mantle – regions to which the strongest modelling errors can be ascribed.

To check these systematics, we calculated the Pearson correlation coefficient $r(\vartheta)$ as a function of azimuthal distance ϑ as follows. For each earthquake, we calculated the azimuthal distances ϑ_{jk} between all station pairs (j, k) and binned those. A set $\{j, k\}_\vartheta$ then contains all stations pairs for one event that have the same azimuthal distance ϑ (in bins of $5°$ width).

We need to adjust for the fact that stations j and k usually have different SNR and hence different μ_j and σ_j in their log-normal distributions of D. Hence we calculate the

standard score of each station j as $z_j = \left(\ln(D_j) - \mu_j\right)/\sigma_j$ and from this the Pearson correlation coefficient of a ϑ bin $\{j, k\}_\vartheta$, using all n_ϑ station pairs in that bin:

$$r(\vartheta) = \frac{1}{n_\vartheta - 1} \sum_{\{j,k\}_\vartheta} z_j z_k. \tag{26}$$

The use of standard scores permits comparison of stations of different SNR and hence log-normal distribution parameters. The values for $r(\vartheta)$ are then fit by a function (see Fig. 7)

$$g(\vartheta) = b_1 + b_2 \cdot \exp(-b_3 \vartheta^2). \tag{27}$$

This permits comparison of D_j for stations with different SNR and distributions of D_j. Then the correlation coefficient was calculated for each azimuthal bin ϑ using all n_ϑ pairs $\{i, j\}_\vartheta$ in this bin.

This azimuth-dependent correlation coefficient $g(\vartheta)$ can be used to fill the elements of covariance matrix $\mathbf{S}_D$ in Eq. (21):

$$\mathbf{S}_{\mathrm{D},i,j} = \begin{cases} \sigma_i \sigma_j \cdot \left(b_1 + b_2 \cdot \exp(-b_3 \vartheta^2)\right), & i \neq j \\ \sigma_i^2, & i = j. \end{cases} \tag{28}$$

An example of such a covariance matrix is shown in Fig. 8. It is for the 2011 earthquake in the US state of Virginia that was used as a detailed working example of Bayesian source inversion in the companion paper (Stähler and Sigloch, 2014).

3.5 Misfit distribution of waveform amplitude measurements

Waveform amplitudes have not been considered so far, even though they provide crucial constraints on focal mechanisms. Our amplitude measurement consists of a comparison of the logarithmic energy content $\ln(A)$ in a 1 s time window around the peak $i = i_1, \ldots, i_2$ of the measured seismogram and its synthetic:

$$\Delta \ln(A)_j = \ln\left(\sum_{i=i_1}^{i_2} u_{j,i}^2\right) - \ln\left(\sum_{i=i_1}^{i_2} u_{j,i}^{c}{}^2\right). \tag{29}$$

Again our goal it to approximate the distribution of this misfit in order to obtain an empirical likelihood function. The distribution of $\Delta \ln(A)$ is almost symmetric around 0; see Fig. S2. The amplitude misfit $|\Delta \ln(A)|$ approximately follows a Laplace distribution, where parameter k does not vary much with SNR (see Supplement). We construct the likelihood function

$$\mathcal{L}^*_{\mathrm{Amp}} = \sum_{j=1}^{n_S} \frac{1}{2k} \exp\left(-\frac{|\Delta \ln(A)|}{k}\right), \tag{30}$$

which assumes no correlation in amplitude misfit between two stations. This assumption is not without problems, but motivated by the fact that amplitude errors are often caused by localised site effects.
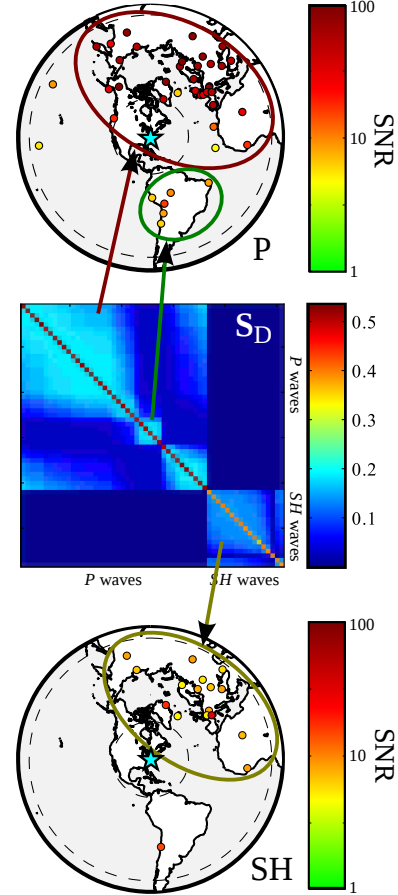

Figure 8. Visualisation of an inter-station covariance matrix $\mathbf{S}_D$ for misfit D (centre panel; cf. Eq. 21), on the example of an m_b 5.7 earthquake that occurred in the US state of Virginia in 2011. Two maps for P and SH data show the recording seismic stations as dots; colour fill indicates the SNR of each waveform measurement. Inter-station correlation depends directly on the azimuthal proximity of two stations. This results in a block-diagonal matrix structure for S_D, because we have sorted stations by azimuth from the source. Blocks correspond to groups of stations with an expected high correlation of errors: (1) a Northern Hemisphere cluster of P wave measurements (circled in dark red), (2) a South American cluster of P waveforms (green) and (3) a Northern Hemisphere cluster of SH waveforms measurement (olive). P and SH measurements are modelled as being uncorrelated. For the analysis, only stations between 32 and 85° epicentral distance have been used, as marked by the dashed lines.

3.6 Application in Bayesian source inversion

In practice these concepts are integrated with the Bayesian source inversion procedure of Stähler and Sigloch (2014) as follows:

1. For every new earthquake, download and archive a suitable selection of broadband, three-component, teleseismic seismograms ($\Delta = 32$ to $85°$). A pragmatic ap-

proach is to use stations from a handful of international, permanent networks (e.g. II, IU, G and GE) to ensure high quality, reliability and relatively even azimuthal coverage, avoiding station clustering in any particular region. This is easily automated using the freely available data management software ObsPyDMT (Scheingraber et al., 2013).

2. Bandpass filter between 0.02 and 1.0 Hz. Rotate horizontal components to the RTZ system. Select signal time windows and noise time windows, and calculate SNR as defined in Eq. (23).

3. For each station, and for P and SH separately, use SNR to calculate distribution parameters μ_i and σ_i from Eq. (25). Populate the diagonal of covariance matrix $\mathbf{S}_{D,ii}$ with the σ_i^2.

4. Estimate correlation coefficient $r(\vartheta_{j,k})$ between two stations (j, k) using Eq. (28). Fill off-diagonal elements:

$$\mathbf{S}_{D,jk} = r(\vartheta_{j,k})\sigma_j\sigma_k. \tag{31}$$

5. Insert μ_i and $\mathbf{S}_D$ in the likelihood equation (Eq. 21), and combine with $\mathcal{L}^*_{\mathrm{Amp}}$ (Fig. 30) to create the total likelihood function

$$\mathcal{L}^* = \mathcal{L}^*_{\mathrm{D}} + \mathcal{L}^*_{\mathrm{Amp}}. \tag{32}$$

6. For each source model $\boldsymbol{m}$ proposed by the sampling algorithm, calculate synthetic seismograms and pass them through the filters of step 2. Calculate the empirical likelihood $\mathcal{L}^*(\boldsymbol{m}|\boldsymbol{d})$ (Eq. 32), which is multiplied with a suitable prior to obtain a posterior probability for $\boldsymbol{m}$. Parameterisation of $\boldsymbol{m}$, Bayesian sampling strategy and construction of the posterior distribution of $\boldsymbol{m}$ are described in the companion paper (Stähler and Sigloch, 2014).

4 Discussion

The most common approach to Bayesian inversion is to assert a simple noise model for which an analytic likelihood function is known: this determines the measure of misfit. We have gone the opposite route in designing a misfit D based on considerations of robustness and dimensionality reduction. Since no noise model was known, we had to investigate the actual noise statistics and thus derive an empirical noise model and likelihood function from the data D. We were fortunate to find that the (multivariate) log-normal distribution provides the best fit to our decorrelation data because it can be evaluated almost as easily and cheaply as the most favourable of all distributions, the Gaussian (normal) distribution.

In fact, analytic probability densities are known for only a few misfit functionals. By far the most commonly used

are the Gaussian (normal) distribution, associated with the ℓ^2 norm misfit, and the Laplace distribution, associated with the ℓ^1 norm. Evaluating residuals of data fits against these analytic distributions is straightforward and fast, which is important in the computationally expensive Bayesian realm.

In practice however, the adoption of ℓ^1 or ℓ^2 misfits may be inappropriate or even impossible. Gauss and Laplace functions may be (too-)poor approximations of the actual distributions of data residuals. Even if they can be deemed adequate for some measurements (e.g. for the sample-wise distance of two times series), they may generate huge and non-sparse covariance matrices (because time samples are numerous and correlated), which are difficult to estimate from the data. Even worse in such multivariate scenarios, analytic expressions of the joint distribution functions may not exist – as is the case for the Laplace distribution (ℓ^1 norm). Effectively this often leaves as the only "choice" for a noise model the (multivariate) normal distribution – whether or not it fits the data at hand.

More often than not, real data contain many more outliers than expected by the normal distribution, certainly in the case of seismic data. Under the ℓ^2 norm, outliers disproportionally bias the solution (deterministic case) or posterior distribution (Bayesian case) and also affect convergence in the Bayesian case. The problem may be mitigated by manual removal of very poorly fitting waveforms, but this is usually time-intensive guesswork and likely to result in other biases.

The ℓ^1 norm is more robust against outliers, and with the same motivation distance norms with non-integer exponents ℓ^p have been proposed and successfully applied, including for source inversion (Marson-Pidgeon and Kennett, 2000). But all norms with $p \neq 2$ share the serious limitation that no analytic expressions are known for the multivariate case.

Samples of real-world, band-limited time series are correlated. If a measured seismogram of length N samples is considered,

$$\boldsymbol{u}_i = \boldsymbol{u}_i^{\mathrm{c}} + \boldsymbol{\epsilon}_{\mathrm{noise},i}, \tag{33}$$

then an $(N \times N)$ covariance matrix for $\epsilon_{\mathrm{noise}}$ needs to be estimated under the ℓ^2 norm. Hierarchical Bayesian methods can be applied to estimate the noise level and covariance from the data itself (see Malinverno and Briggs, 2004; Bodin, 2010; Mustać and Tkalčić, 2016)), but in many cases it may be more guessed than estimated.

The situation is further complicated if the noise model can no longer be purely additive ("$+\epsilon_{\mathrm{noise}}$"). We have argued that our noise model needs to be

$$\boldsymbol{u}_i = \boldsymbol{u}_i^{\mathrm{c}} * \boldsymbol{T}_{\mathrm{model}_i} * \boldsymbol{T}_{\mathrm{inst},i} + \boldsymbol{\epsilon}_{\mathrm{noise},i}, \tag{34}$$

where the convolving terms are systematic modelling error. In theory this type of error might be eliminated with computationally powerful waveform forward modelling and more research into detailed earth structure. But since those efforts would be tangential to the problem at hand (source inversion), the cost would seem prohibitive. Hence we do want the

option of treating the modelling error as "just another source of noise", to be accommodated by a more sophisticated noise model, the analytic expression of which will be unknown.

Another reason for leaving the Gaussian or ℓ^2 realm might be a change of measurement. In our case, the cross-correlation or decorrelation measurements collapse $N \times 2$ samples of two times series into a single scalar CC or D. Even if inter-sample correlations of the time series actually were multivariate Gaussian, the statistics of CC or D would be something more complicated. On the upside, the dimensionality of the multivariate problem is reduced by a factor of N, which helps substantially when forced to take the empirical path toward obtaining a likelihood function. Thus inter-station covariances are the only correlations to estimate, and the fact that they are simple covariances (second moments) is, again, owed to the fortunate fact that the log-normal distribution yielded the best fit to the misfit histogram.

We are not sure whether there is a theoretical reason that the log-normal distribution should be associated with the decorrelation misfit D, and thus effectively with CC. Whatever the case, this finding is highly relevant in that it also opens up the path to Bayesian sampling of other optimisation problems that have previously adopted the cross-correlation coefficient CC of seismograms as their misfit criterion, e.g. other flavours of seismic source inversion (Kikuchi and Kanamori, 1991; Marson-Pidgeon and Kennett, 2000), seismic tomography (Sigloch and Nolet, 2006; Tape et al., 2009) or the estimation of earthquake cluster sizes (Menke et al., 1990; Menke, 1999; Kummerow, 2010).

As noted, the proposed empirical likelihood function $\mathcal{L}^*(\boldsymbol{m}|\boldsymbol{d})$ is no likelihood function in a strict sense because it is not derived from the noise on the raw data samples but rather from the noise (i.e. residual) of misfit functional D. For other inverse problems, it has to be evaluated separately, whether or not a noise model exists that can describe the difference between modelled and measured seismograms completely as an additive term. If that is the case, a classical likelihood can be used, but many inverse problems in seismology are similar to the one presented here, and the proposed empirical likelihood offers a path to a more thorough Bayesian treatment. It is just important to remember that the distribution of D has to be determined from synthetic seismograms calculated with the same velocity model and forward solver as it is used for the actual inversion.

Other misfit criteria have been used in optimisation contexts in seismology. For the purpose of source parameter inversion, their noise properties could be investigated along the lines laid out by this work, and their empirical likelihood functions studied. But unless their noise distributions turn out to be as simple as for the D misfit (they would essentially have to follow the normal or log-normal distribution), these other misfit choices will be computationally more costly to sample. It is pleasing that the cross-correlation, long appreciated for its robust performance in deterministic optimisation,

is now also vindicated in a Bayesian context by the results of our study.

5 Conclusions

This paper presents an approach to Bayesian inference using the new misfit criterion of waveform (de)correlation. Decorrelation D greatly reduces the number of data uncertainties and correlations, by collapsing the temporal correlations of samples in a broadband seismogram into a single scalar D, or into n scalars per source estimate, where n is the number of time windows on different seismogram components used to estimate the source parameters of one earthquake. This leaves only n_S inter-station correlations to be determined, and we show how they depend on the SNR of the D measurements and on the azimuthal distances of seismic stations. The noise on D turns out to have simple characteristics, approximately following an n_S-variate log-normal distribution, a finding that renders the formulation of the likelihood function for D straightforward.

This opens the way for the methodically correct Bayesian sampling of parameter estimation problems that use the cross-correlation CC or decorrelation $D = 1 - \text{CC}$ of seismological broadband waveforms as their measure of data (mis)fit – including not only our source inversion procedure PRISM but also certain flavours of waveform tomography or earthquake cluster analysis. In terms of data dimensionality reduction the present work complements its companion Stähler and Sigloch (2014), which focused on reducing the dimensionality of model parameters to a number amenable to Bayesian sampling. It can also serve as a template for the empirical derivation of noise models and likelihood functions for other misfit measures on broadband seismograms.

6 Data availability

The analysis has been performed on publicly available seismological data. All waveform data came from the IRIS and ORFEUS data management centres.

The Supplement related to this article is available online at doi:10.5194/se-7-1521-2016-supplement.

Edited by: C. Krawczyk
Reviewed by: T. Bodin and C. Tape

References

Bodin, T.: Transdimensional Approaches to Geophysical Inverse Problems, Ph.D. thesis, Australian National University, 2010.

Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P.: Transdimensional tomography with unknown data noise, Geophys. J. Int., 189, 1536–1556, doi:10.1111/j.1365-246X.2012.05414.x, 2012.

Bodin, T., Leiva, J., Romanowicz, B., Maupin, V., and Yuan, H.: Imaging anisotropic layering with Bayesian inversion of multiple data types, Geophys. J. Int., 206, 605–629, doi:10.1093/gji/ggw124, 2016.

Bogert, B.: Correction of seismograms for the transfer function of the seismometer, Bull. Seismol. Soc. Am., 52, 781–792, 1962.

Bondár, I. and Storchak, D. A.: Improved location procedures at the International Seismological Centre, Geophys. J. Int., 186, 1220–1244, doi:10.1111/j.1365-246X.2011.05107.x, 2011.

Chapman, C. H.: A new method for computing synthetic seismograms, Geophys. J. R. Astron. Soc., 54, 481–518, 1978.

Dettmer, J., Benavente, R., Cummins, P. R., and Sambridge, M.: Trans-dimensional finite-fault inversion, Geophys. J. Int., 199, 735–751, doi:10.1093/gji/ggu280, 2014.

Duputel, Z., Rivera, L., Fukahata, Y., and Kanamori, H.: Uncertainty estimations for seismic source inversions, Geophys. J. Int., 190, 1243–1256, doi:10.1111/j.1365-246X.2012.05554.x, 2012.

Duputel, Z., Agram, P. S., Simons, M., Minson, S. E., and Beck, J. L.: Accounting for prediction uncertainty when inferring subsurface fault slip, Geophys. J. Int., 197, 464–482, doi:10.1093/gji/ggt517, 2014.

Dziewoński, A. M. and Anderson, D. L.: Preliminary reference Earth model, Phys. Earth Planet. Inter., 25, 297–356, doi:10.1016/0031-9201(81)90046-7, 1981.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J.: Markov Chain Monte Carlo in Practice, Chapman & Hall/CRC, London, 1996.

Hosseini, K. and Sigloch, K.: Multifrequency measurements of core-diffracted P waves (Pdiff) for global waveform tomography, Geophys. J. Int., 203, 506–521, doi:10.1093/gji/ggv298, 2015.

Houser, C., Masters, G., Shearer, P. M., and Laske, G.: Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, Geophys. J. Int., 174, 195–212, doi:10.1111/j.1365-246X.2008.03763.x, 2008.

Kanamori, H. and Given, J. W.: Use of long-period surface waves for rapid determination of earthquake-source parameters, Phys. Earth Planet. Inter., 27, 8–31, doi:10.1016/0031-9201(81)90083-2, 1981.

Käufl, P., Fichtner, A., and Igel, H.: Probabilistic full waveform inversion based on tectonic regionalization – development and application to the Australian upper mantle, Geophys. J. Int., 193, 437–451, doi:10.1093/gji/ggs131, 2013.

Kennett, B. L. N. and Engdahl, E. R.: Traveltimes for global earthquake location and phase identification, Geophys. J. Int., 105, 429–465, doi:10.1111/j.1365-246X.1991.tb06724.x, 1991.

Kikuchi, B. Y. M. and Kanamori, H.: Inversion of complex body waves – III, Bull. Seismol. Soc. Am., 81, 2235–2350, 1991.

Kotz, S., Kozubowski, T. J., and Podgórski, K.: The Laplace Distribution and Generalizations: A Revisit With Applications to Communications, Economics, Engineering, and Finance, Springer, 2001.

Kozubowski, T. J., Podgórski, K., and Rychlik, I.: Multivariate generalized Laplace distribution and related random fields, J. Multivar. Anal., 113, 59–72, doi:10.1016/j.jmva.2012.02.010, 2013.

Kristekova, M., Kristek, J., Moczo, P., and Day, S. M.: Misfit Criteria for Quantitative Comparison of Seismograms, Bull. Seismol. Soc. Am., 96, 1836–1850, doi:10.1785/0120060012, 2006.

Kummerow, J.: Using the value of the crosscorrelation coefficient to locate microseismic events, Geophysics, 75, MA47, doi:10.1190/1.3463713, 2010.

Larose, E., Planès, T., Rossetto, V., and Margerin, L.: Locating a small change in a multiple scattering environment, Appl. Phys. Lett., 96, 2010–2012, doi:10.1063/1.3431269, 2010.

Mahalanobis, P. C.: On the generalized distance in statistics, Proc. Natl. Inst. Sci. India, 2, 49–55, 1936.

Malinverno, A. and Briggs, V. A.: Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes, Geophysics, 69, 1005–1016, doi:10.1190/1.1778243, 2004.

Marson-Pidgeon, K. and Kennett, B. L. N.: Source depth and mechanism inversion at teleseismic distances using a neighborhood algorithm, Bull. Seismol. Soc. Am., 90, 1369–1383, doi:10.1785/0120000020, 2000.

Menke, W.: Using waveform similarity to constrain earthquake locations, Bull. Seismol. Soc. Am., 89, 1143–1146, 1999.

Menke, W., Lerner-Lam, A. L., Dubendorff, B., and Pacheco, J. F.: Polarization and coherence of 5 to 30 Hz seismic wave fields at a hard-rock site and their relevance to velocity heterogeneities in the crust, Bull. Seismol. Soc. Am., 80, 430–449, 1990.

Mustać, M. and Tkalčić, H.: Point source moment tensor inversion through a Bayesian hierarchical model, Geophys. J. Int., 204, 311–323, doi:10.1093/gji/ggv458, 2016.

Owen, A. B.: Empirical likelihood ratio confidence intervals for a single functional, Biometrika, 75, 237–249, doi:10.1093/biomet/75.2.237, 1988.

Peterson, J.: Observations and Modeling of Seismic Background Noise, Tech. Rep., USGS, Albuquerque, New Mexico, 1993.

Sambridge, M.: Geophysical inversion with a neighbourhood algorithm – II. Appraising the ensemble, Geophys. J. Int., 138, 727–746, doi:10.1046/j.1365-246x.1999.00900.x, 1999.

Sambridge, M. and Kennett, B. L. N.: Seismic event location: nonlinear inversion using a neighbourhood algorithm, Pure Appl. Geophys., 158, 241–257, doi:10.1007/PL00001158, 2001.

Scheingraber, C., Hosseini, K., Barsch, R., and Sigloch, K.: ObsPyLoad: A Tool for Fully Automated Retrieval of Seismological Waveform Data, Seismol. Res. Lett., 84, 525–531, doi:10.1785/0220120103, 2013.

Schimmel, M.: Phase cross-correlations: design, comparisons and applications, Bull. Seismol. Soc. Am., 89, 1366–1378, 1999.

Sigloch, K.: Mantle provinces under North America from multifrequency P wave tomography, Geochemistry, Geophys. Geosystems, 12, Q02W08, doi:10.1029/2010GC003421, 2011.

Sigloch, K. and Mihalynuk, M. G.: Intra-oceanic subduction shaped the assembly of Cordilleran North America, Nature, 496, 50–56, doi:10.1038/nature12019, 2013.

Sigloch, K. and Nolet, G.: Measuring finite-frequency body-wave amplitudes and traveltimes, Geophys. J. Int., 167, 271–287, doi:10.1111/j.1365-246X.2006.03116.x, 2006.

Sigloch, K., McQuarrie, N., and Nolet, G.: Two-stage subduction history under North America inferred from multiple-frequency tomography, Nat. Geosci., 1, 458–462, doi:10.1038/ngeo231, 2008.

Stähler, S. C. and Sigloch, K.: Fully probabilistic seismic source inversion – Part 1: Efficient parameterisation, Solid Earth, 5, 1055–1069, doi:10.5194/se-5-1055-2014, 2014.

Stähler, S. C., Sigloch, K., and Nissen-Meyer, T.: Triplicated P-wave measurements for waveform tomography of the mantle transition zone, Solid Earth, 3, 339–354, doi:10.5194/se-3-339-2012, 2012.

Stutzmann, E., Schimmel, M., Patau, G., and Maggi, A.: Global climate imprint on seismic noise, Geochemistry, Geophys. Geosystems, 10, Q11016, doi:10.1029/2009GC002619, 2009.

Tape, C., Liu, Q., Maggi, A., and Tromp, J.: Adjoint tomography of the southern California crust, Science, 325, 988–92, doi:10.1126/science.1175298, 2009.

Tape, W. and Tape, C.: A uniform parametrization of moment tensors, Geophys. J. Int., 202, 2074–2081, doi:10.1093/gji/ggv262, 2015.

Tape, W. and Tape, C.: A confidence parameter for seismic moment tensors, Geophys. J. Int., 205, 938–953, doi:10.1093/gji/ggw057, 2016.

Tarantola, A. and Valette, B.: Inverse problems = quest for information, J. Geophys., 50, 159–170, doi:10.1016/j.pepi.2016.05.012, 1982.

Vallée, M. and Douet, V.: A new database of Source Time Functions (STFs) extracted from the SCARDEC method, Phys. Earth Planet. Int., 257, 149–157, 2016.

Vallée, M., Charléty, J., Ferreira, A. M. G., Delouis, B., and Vergoz, J.: SCARDEC: a new technique for the rapid determination of seismic moment magnitude, focal mechanism and source time functions for large earthquakes using body-wave deconvolution, Geophys. J. Int., 184, 338–358, doi:10.1111/j.1365-246X.2010.04836.x, 2011.

van Driel, M., Krischer, L., Stähler, S. C., Hosseini, K., and Nissen-Meyer, T.: Instaseis: instant global seismograms based on a broadband waveform database, Solid Earth, 6, 701–717, doi:10.5194/se-6-701-2015, 2015.