



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

Simon C. Stähler<sup>1,2,3</sup> and Karin Sigloch<sup>4,1</sup>

<sup>1</sup>Dept. of Earth and Environmental Sciences, Ludwig-Maximilians-Universität (LMU), Theresienstr. 41, 80333 München

<sup>2</sup>Munich Centre of Advanced Computing, Department of Informatics, Technische Universität München

<sup>3</sup>Leibniz Institute for Baltic Sea Research (IOW), Seestr. 15, 18119 Rostock

<sup>4</sup>Dept. of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, UK

*Correspondence to:* Simon C. Stähler (staehler@geophysik.uni-muenchen.de)

**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 & 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 modeled counterparts. This approach yields not only depth and moment tensor estimates but also source time functions.

A prerequisite for Bayesian inference is the proper characterization of the noise distribution 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  $D = 1 - CC$ , performs more robustly as a misfit criterion than  $\ell^P$ -norms, more commonly used measures of misfit based on distances between individual time samples.

From a set of over 900 user-supervised, deterministic source solutions treated as a quality-controlled reference, we derive the noise distribution on signal decorrelation  $D = 1 - CC$  of the broadband seismogram fits. The noise on  $D$  is found to approximately follow a log-normal distribution, a fortunate fact that readily accommodates the formulation of a 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 of the  $CC$  measurements and on the back-azimuthal distances of seismic stations. By identifying and quantifying its 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.

seismic source inversion – Bayesian inference – waveform cross-correlation – covariance matrix

## 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 can be used to locate earthquakes, to understand earthquake mechanisms and their implications for tectonic settings and seismic hazard, but are also important in seismic tomography, where accurate source information is a prerequisite for achieving optimal fits between observed and modeled (waveform) data.



$\mathbf{m}$	model vector of earthquake source parameters ( $M$ -dimensional) as defined in Stähler and Sigloch (2014): earthquake depth, moment tensor and source time function
$\mathbf{d}$	geophysical data vector ( $N$ -dimensional)
$M$	number of model parameters
$N$	number of data
$g(\mathbf{m})$	forward operator acting on a model parameter vector $\mathbf{m}$
$\mathcal{L}(\mathbf{m} \mathbf{d})$	Likelihood of a model $\mathbf{m}$ given the data $\mathbf{d}$
$\mathbf{S}_D$	Data covariance matrix
$\Phi$	total misfit of one model $\mathbf{m}$ and data $\mathbf{d}$ , $\Phi = -\ln \mathcal{L}$
$\Phi^W$	misfit between one recorded and predicted seismogram
$\mathbf{u}_{j,i,i} = \{1, \dots, n_j\}$	time-discrete seismogram $j$
$\mathbf{u}_{j,i,i}^c = \{1, \dots, n_j\}$	synthetic seismogram $j$ predicted by a model $\mathbf{m}$ and a forward operator $g(\mathbf{m})$
$j$	index of seismogram
$i$	index of sample in a seismogram.
$n_j$	number of samples in a seismogram $j$
$n_S$	number of seismograms. $n_S \equiv N$ , if the decorrelation misfit is used.
$n_{\text{tot}} = \sum_{j=1}^{n_S} n_j$	total number of samples in all seismograms. $n_{\text{tot}} \equiv N$ , if a $\ell^p$ -misfit is used.
$CC_k^{\mathbf{u}_i, \mathbf{u}_i^c}$	(Normalized) cross-correlation function between time series $\mathbf{u}$ and $\mathbf{u}^c$ using a window function $w_i$ : $CC_k^{\mathbf{u}_i, \mathbf{u}_i^c} = \frac{\sum_{i=1}^n (w_i \mathbf{u}_{i-k} \cdot \mathbf{u}_i)}{\sqrt{\sum_{i=1}^n (w_i \mathbf{u}_{i-k})^2 \cdot \sum_{i=1}^n (w_i \mathbf{u}_i)^2}}$
$CC^{\mathbf{u}_i, \mathbf{u}_i^c}$	Maximum of $CC_k^{\mathbf{u}_i, \mathbf{u}_i^c}$ over $k$ ; the “correlation between $\mathbf{u}_i$ and $\mathbf{u}_i^c$ ”
$D^{\mathbf{u}_i, \mathbf{u}_i^c}$	Decorrelation, $D^{\mathbf{u}_i, \mathbf{u}_i^c} := 1 - \max_k \{CC_k^{\mathbf{u}_i, \mathbf{u}_i^c}\}$

**Table 1.** Symbols frequently used in this paper

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 tradeoffs and non-uniqueness. A correlation between dip and scalar moment was discovered by Kanamori and Given (1981). Source depth is 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 stfs are explicitly estimated. This makes global search methods and ensemble sampling particularly attractive if the associated computational hurdles can be surmounted.

In a companion paper (Stähler and Sigloch, 2014), we developed the PRISM algorithm, a fully probabilistic inversion for source depth, moment tensor and source time function (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 ( $\approx 1$  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 center 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 linearized, iterative approach that semi-automatically deconvolved broadband source time func-  
5 tions, source depths, and moments tensors of more than 2,000 earthquakes, which were subsequently used in several waveform tomographies (Sigloch et al., 2008; Sigloch, 2011; Sigloch and Mihalynuk, 2013).

The required human supervision time called for full automatization, preferably in a Bayesian setting that would circumvent the occasional divergence of the non-linear optimization and would automatically diagnose parameters tradeoffs of the kind described. PRISM (Stähler and Sigloch, 2014) solved this problem, but we left the justification of its misfit criterion and the  
10 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 space has to be as small as possible, preferably less than 20. Depth is one parameter and a normalized 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  
15 lateral location problem is adequately addressed by existing data centers (NEIC or Bondár and Storchak, 2011) and in any case, we would re-estimate all hypocenters at the time of tomographic inversion. The STF is a high-dimensional parameter vector, which Sigloch and Nolet (2006); Stähler et al. (2012) parameterised simply as a time series of 256 unknowns. To reduce its dimensionality for Bayesian sampling, Stähler and Sigloch (2014) selected the 900 best-constrained deterministic STF solutions from the total of  $> 2,000$  obtained by Sigloch and Nolet (2006), and decomposed them into empirical orthogonal  
20 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 figure 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  $\approx 20$ .

This space is sampled by both stages of the Neighbourhood Algorithm, resulting in an ensemble of source solutions  $\mathbf{m}$  (c.f. 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 are full uncertainties on traveltimes measurements  $\Delta T_j$  at stations  $j$ . These traveltimes delays are incidental in the context of source  
30 inversion (as the time shifts between observed and synthetic seismograms that maximize the cross-correlation coefficients  $CC_j$ , figure 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 are the  $CC_j$ . When parameter estimation is performed as a deterministic optimization 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  
35 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 misfit (the “noise model”). In 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”.

- 5 Section 2 argues for the adoption of the signal decorrelation  $D := 1 - CC$  as a robust measure of misfit, where  $CC$  is the normalized 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).
- 10 Section 2.2 shows that the popular  $\ell^2$  and  $\ell^1$  norms (Mahalanobis, 1936) would be sub-optimal misfit criteria because noise on seismic signals is not simply additive Gaussian or Laplacian, but 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 section 2.4 demonstrates that the signal decorrelation  $D$  performs more robustly than  $\ell^P$ -norms on realistic seismological waveform data.
- 15 To identify the likelihood function  $\mathcal{L}(\mathbf{m}|\mathbf{d})$  of misfit  $D$  in section 3, we draw once more on the prior knowledge contained in our set of 900 deterministic source solutions, 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  $D$ . Thus we can instruct the probabilistic inversion to explore subspaces of solutions  $\mathbf{m}$  that yield similarly low levels of misfit  $D$  as these best-fitting deterministic solutions.
- 20 Section 3.6 presents a worked example for the construction of a likelihood function  $\mathcal{L}(\mathbf{m}|\mathbf{d})$  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 section 4.

## 2 Noise and misfit criteria

### 2.1 Bayesian inference

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

$$\pi(\mathbf{m}|\mathbf{d}) = \frac{1}{p(\mathbf{d})} \mathcal{L}(\mathbf{m}|\mathbf{d}) p(\mathbf{m}). \quad (1)$$

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

30 between measured data  $\mathbf{d}$  and predicted data  $g(\mathbf{m})$ . This difference or *misfit* is defined, following convention as

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



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

## 2.2 Metric-based misfit criteria

“Good” solutions  $\mathbf{m}$  are associated with small misfits  $\Phi$ , where the exact definition of  $\Phi$  depends on the nature of the data  $\mathbf{d}$ , 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  $\Phi^W : \mathbb{R}^N \times \mathbb{R}^N \mapsto [0, \infty)$  on  $\mathbf{d}, g(\mathbf{m}) \in \mathbb{R}^N$ .

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

15 When the method of least squares is used to calculate the  $\ell^2$ -misfit

$$\Phi_{\ell^2}^W(\mathbf{m}|\mathbf{d}) = k' \left( \frac{1}{2} (\mathbf{d} - g(\mathbf{m}))^T \mathbf{S}_{\mathbf{D}}^{-1} (\mathbf{d} - g(\mathbf{m})) \right), \quad (3)$$

the assumption is that the noise  $\epsilon$  is additive and Gaussian distributed:

$$\mathbf{d} = g(\mathbf{m}) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{S}_{\mathbf{D}}) \quad (4)$$

The size  $[N \times N]$ -data covariance matrix  $\mathbf{S}_{\mathbf{D}} \in \text{Sym}_N$  describes the correlation between the error of individual measurements  
 20  $d_i$ .

In the case of a seismic waveform,  $\Phi^W$  is

$$\Phi_{\ell^2}^W(\mathbf{m}|\mathbf{d}) = k' \sum_{i=1}^n \sum_{i'=1}^n (u_i - u_i^c)^T (\mathbf{S}_{\mathbf{D}}^{-1})_{i,i'} (u_{i'} - u_{i'}^c), \quad (5)$$

and  $\mathbf{S}_{\mathbf{D}}$  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  
 25  $\mathbf{S}_{\mathbf{D}}$  from the autocorrelation function  $R_{\epsilon\epsilon}$  of the (discrete) noise time series  $\epsilon_i$ .  $\mathbf{S}_{\mathbf{D}}$  is a Toeplitz matrix, where the rows are shifted instances of the autocorrelation function  $R_{\epsilon\epsilon}$ .

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

See Bodin et al. (2012) for an example of how to construct under the assumption of an autoregressive noise model.

For the estimation of the parameters  $\mathbf{m}$  of one earthquake source, we would normally use  $n_S$  seismograms  $u_i$  measured at  
 30  $n_S$  stations with index  $j$ . The overall misfit  $\Phi(\mathbf{m})$  for a source solution will be comprised of the misfits of the single waveforms



$\Phi_{\ell^2,j}^W(\mathbf{m})$ . If the noise on each waveform  $j$  is assumed to be uncorrelated from the noise on all others, then it is legitimate to define the overall misfit as being simply additive:

$$\Phi(\mathbf{m}) = \sum_{j=1}^{n_S} \Phi_{\ell^2,j}^W(\mathbf{m}). \quad (7)$$

If the noise on the waveforms is correlated, the equation 3 has to be extended, such that  $\mathbf{d}$ ,  $\mathbf{m}$  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 sample  $u_i$  in a seismogram  $u_i$  is considered to be uncorrelated with the others and has a variance  $\sigma$ , then  $\mathbf{S}_D$  is a diagonal matrix with diagonal elements  $\sigma^2$  and eqn. 3 reduces to

$$\Phi_{\ell^2}^W(\mathbf{m}|\mathbf{d}) = k' \left( \frac{1}{2} \sum_i \frac{(d_i - g_i(\mathbf{m}))^2}{\sigma_i^2} \right), \quad (8)$$

and with  $n_S$  waveforms  $u_i$ , the total misfit defined in eqn. 7 becomes:

$$\Phi = k' \left( \frac{1}{2} \sum_j \sum_i \frac{u_i - u_i^c}{\sigma^2} \right), \quad (9)$$

the *weighted least squares criterion*.

If the noise can be described well by the normal distribution, the  $\ell^2$ -norm can be successfully applied. It is, however very sensitive to data  $d_i$  deviating strongly from the prediction  $g_i(\mathbf{m})$ . 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, Käuffl et al. (2013) have proposed to use the more outlier-resistant  $\ell^1$  norm as a misfit criterion of observed and modeled seismograms. They assume that noise on the time samples  $u_i$  is independently Laplace-distributed with width  $b_i$ , i.e. no temporal correlation:

$$\mathbf{d} = g(\mathbf{m}) + \boldsymbol{\epsilon}, \quad \epsilon_i \sim \text{Laplace}(0, b_i) \quad (10)$$

$$\Phi_{\ell^1}^W(\mathbf{m}|\mathbf{d}) = - \sum_i \frac{|d_i - g_i(\mathbf{m})|}{b_i} - \ln 2b_i. \quad (11)$$

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 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  $\ell^1$  norm can currently not be harnessed.



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

$$\Phi_{\ell^p}^W(\mathbf{m}|\mathbf{d}) = \left( \sum_{i=1}^n \frac{|d_i - g_i(\mathbf{m})|^p}{\sigma^p} \right)^{1/p} \quad (12)$$

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

In summary, it is tempting to chose  $\ell^p$  misfits based on the time-sample-wise distance between observed and modelled waveforms because the underlying noise models are straightforward to state (uncorrelated/correlated Gaussian, uncorrelated Laplace distribution) and to translate into corresponding likelihood functions. Unfortunately, these noise models are very crude approximations to 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  $\mathcal{L}(\mathbf{m}|\mathbf{d})$  can be constructed from the distribution  $p(F)$  of any functional  $F$  of the observed and predicted waveforms  $\mathbf{u}_i, \mathbf{u}_i^c \in \mathbb{R}$ :  $F: \mathbb{R} \times \mathbb{R} \mapsto [0, \infty)$ . In our attempt to move beyond  $F$  being a sample-wise distance between  $\mathbf{u}_i$  and  $\mathbf{u}_i^c$ , we generally want a candidate  $F$  to meet the following conditions:

1. For  $\mathbf{u}_i = \mathbf{u}_i^c$ ,  $F$  should take a fixed value, say 0.
2. With decreasing similarity of  $\mathbf{u}_i$  and  $\mathbf{u}_i^c$ ,  $F$  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  $\mathbf{u}_i$ , i.e.  $F(a \cdot \mathbf{u}_{i+k}, \mathbf{u}_i^c) \cong F(\mathbf{u}_i, \mathbf{u}_i^c)$  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  $\mathbf{m}$ , 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*, *instrument error*.

1. **Ambient noise**  $\epsilon_{\text{noise}}$ : This is noise from man-made or natural sources around the receiver. It can be described very well by an additional term of like  $\epsilon_{\text{noise}} \sim \mathcal{N}(0, S)$  (see eq. 5).
2. **Waveform modelling error**  $T_{\text{model},i}$ : The synthetic waveform  $\mathbf{u}_i^c$  can never be identical to the observed  $\mathbf{u}_i$ , 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  $\mathbf{u}_i^c$ , but we think that it should rather take the form of a transfer function  $\mathbf{T}_{\text{model},i}$ , between  $\mathbf{u}_i$  and  $\mathbf{u}_i^c$ , which will hopefully be Dirac-like in character. However,  $\mathbf{T}_{\text{model},i}$  will include the site response (receiver side reverberations), which can create strong waveform coda, see Fig. 2. Hence,  $\mathbf{T}_{\text{model},i}$  could in practice be rather oscillatory.

3. **Instrument error**  $\mathbf{T}_{\text{inst},i}$ : A displacement seismogram  $\mathbf{u}_i$  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  $\mathbf{T}_{\text{inst},i}$  convolving  $\mathbf{u}_i$ .

In summary, the difference between a modeled  $\mathbf{u}_i^c$  and observed waveform  $\mathbf{u}_i$  is:

$$10 \quad \mathbf{u}_i = \mathbf{u}_i^c * \mathbf{T}_{\text{model},i} * \mathbf{T}_{\text{inst},i} + \epsilon_{\text{noise},i}. \quad (13)$$

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

Next, we will test the signal decorrelation  $D$  against these four criteria.

## 2.4 Signal decorrelation coefficient as a misfit

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

$$D^{\mathbf{u}_i, \mathbf{u}_i^c} := 1 - \max_k \{CC_k^{\mathbf{u}_i, \mathbf{u}_i^c}\}, \quad (14)$$

where

$$CC_k^{\mathbf{u}_i, \mathbf{u}_i^c} = \frac{\sum_{i=1}^n (w_i u_{i-k}^c \cdot u_i)}{\sqrt{\sum_{i=1}^n (w_i u_{i-k}^c)^2 \cdot \sum_{i=1}^n (w_i u_i)^2}} \quad (15)$$

is the normalized cross-correlation coefficient and  $k$  is the time delay between  $\mathbf{u}_i^c$  and  $\mathbf{u}_i$  for which the normalized cross-correlation function  $CC_k^{\mathbf{u}_i, \mathbf{u}_i^c}$  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.  $D^{\mathbf{u}_i, \mathbf{u}_i^c}$  takes the value 0 for identical signals  $\mathbf{u}_i^c \equiv \mathbf{u}_i$ , since  $CC_{k=0}^{\mathbf{u}_i, \mathbf{u}_i^c} = 1$ .
2. For  $\mathbf{u}_i \neq \mathbf{u}_i^c$ ,  $0 < D^{\mathbf{u}_i, \mathbf{u}_i^c} < 2$ , i.e.  $D$  values larger than for the case  $\mathbf{u}_i^c \equiv \mathbf{u}_i$ , and  $D^{\mathbf{u}_i, \mathbf{u}_i^c}$  increases with decreasing similarity of  $\mathbf{u}_i$  and  $\mathbf{u}_i^c$ . (The value of  $D = 2$  is not strictly reached for a real signal, but rather the maximum value of  $D$  is 1 minus the minimum of the autocorrelation function.)
3. If a time shift  $k'$  is small compared to the window length, we have:

$$CC_k^{\mathbf{u}_i, \mathbf{u}_i^c} \approx CC_{k+k'}^{\mathbf{u}_i, \mathbf{u}_i^c} \quad \text{and thus} \quad D^{\mathbf{u}_i, \mathbf{u}_i^c} \approx D^{\mathbf{u}_i, \mathbf{u}_i^c}$$



Due to the normalization in eq. 15,  $D$  is amplitude-independent:

$$CC^{u_i, u_i^c} = CC^{u_i, A \cdot u_i^c} \text{ and thus } D^{u_i, u_i^c} = D^{u_i, A \cdot u_i^c}$$

The fourth criterion, discriminative power and robustness against noise is less straightforward to demonstrate. We proceed empirically by showing its superior performance over the  $\ell^2$  and  $\ell^1$  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-wavetrain, recorded at 40° epicentral distance. The seismograms were modeled using the WKB method of Chapman (1978) in the IASP91 velocity model (Kennett and Engdahl, 1991). Since the chosen source depth is shallow (10 km), the P pulse is within seconds followed by depth phases pP, sP, which effectively permits to invert 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: 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 to have a unit amplitude spectrum and a random phase spectrum between 0 and  $\alpha \cdot \pi/2$ .

$$\mathbf{u}_{\text{m.e.}} = \mathbf{u}_i^c * T_{\text{error},i} \quad (16)$$

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

2. By adding a band-limited noise term

$$\mathbf{u}_{\text{pert}} = \mathbf{u}_{\text{m.e.}} + \beta \boldsymbol{\epsilon}, \text{ where } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{S}_{\mathbf{D}}) \quad (17)$$

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

Fig. 3 shows the resulting reference waveform (left) and perturbed waveforms for  $\alpha = 0.9$  and  $\beta = 0.5$ , i.e. moderate background noise and strong perturbation of the signal. The unperturbed waveform  $\mathbf{u}_i$  is plotted in solid thin black, the waveform perturbed with modelling error  $\mathbf{u}_{\text{m.e.}}$  in dotted blue and the resulting reference trace in solid red. It bears little resemblance with the unperturbed waveform.

The right plot shows the value of the three waveforms misfits  $\ell^1$ ,  $\ell^2$  and  $D$  between  $\mathbf{u}_i^c$  and  $\mathbf{u}_{\text{pert}}$  over varying source depths. It simulates an inversion for the depth of an earthquake using seismic waveforms. The waveform contains the  $P$ ,  $pP$  and  $sP$ -arrival. The depth is mainly constrained by the relative arrival time of the three and the resulting waveform of the whole  $P$ - $pP$ - $sP$  wave train. The perturbation of eq. 17 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 10km than for any of the others. To take into account the stochastic nature of these perturbation, 500 realisations of  $\mathbf{u}_{\text{pert}}$  were calculated for the same parameters  $\alpha, \beta$ , but with different random numbers. The coloured shades mark the 95%/2 sigma quantiles of the misfit values, the solid line marks the median.

5 The  $\ell^2$  misfit could not recognize  $\mathbf{u}_i^c$  in  $\mathbf{u}_{\text{pert}}$  anymore and assigns the lowest misfit to a depth of 3 km. An analysis of different noise and perturbation levels shows that the  $\ell^2$  norm is relatively robust against background noise, but not against perturbations from a modelling error. This seems reasonable given the underlying noise model of this misfit.

The  $\ell^1$ -norm does better, in that it has a minimum at 9 km depth, close to the true value. The zig-zag shape suggests that the value of 9 km is stochastic. However, the median value at 9-10 km reaches slightly below the lower quartile for other depths, 10 meaning that in reality, the resolution power of the  $\ell^1$ -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  $\ell^2$ -norm, but less 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  $\ell^1$  norm. More values of  $\alpha, \beta$  are shown in fig. S1 in the 15 electronic supplement. The analysis of the confidence intervals shows that the values for  $CC$  scatter slightly more than the ones for  $\ell^2$  and much more than for  $\ell^1$ . 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 stronger 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. It 20 shows the difference between the misfit value for the true depth 10 km and the average misfit value for a depth of 20-30 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  $\ell^1$  and  $\ell^2$  norm cannot recognize the true depth with more than two sigma, even for high signal to noise ratios, while the decorrelation  $D$  stays well above three sigma, even for SNRs of six.

## 25 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  $CC = 1 - D$  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, 30 1999; Kummerow, 2010). It has rarely been used as a misfit criterion in source inversion – we are only aware of Kikuchi and Kanamori (1991) and Marson-Pidgeon and Kennett (2000).  $CC$  or  $D = 1 - CC$  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 has 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 on the measurements  $\mathbf{d}$  these estimates  $\mathbf{m}$  are based upon. The measurements  $\mathbf{d}$  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  $\mathbf{m}_j$  approach the true source parameters  $\mathbf{m}_{\text{true},j}$ , the histogram of the  $CC$  (or  $D = 1 - CC$ ) values approximates the pdf of  $CC$  (or  $D$ ). Thus we can obtain an empirical likelihood function even in absence of an analytically describable noise model.

Summing up our reasoning and procedure:

- We can consider the measurements of misfit functional  $\Phi_j(\mathbf{m}_0|\mathbf{d})$  for one earthquake at  $j = 1, \dots, n_S$  recording receivers as realisations of a random process that follows a yet unknown probability density function  $p(x)$ .  $\mathbf{m}_0$  are the true source parameters, and any misfit  $\Phi_j$  is therefore due to ambient noise and modelling errors on the seismograms, as described in section 2.3.
- In practice we never get to know  $\mathbf{m}_0$  but only a (hopefully close) estimate  $\mathbf{m}_{\text{est}}$ , the result of a deterministic source inversion procedure. Hence all we can actually observe is  $\Phi(\mathbf{m}_{\text{est}}|\mathbf{d})$ , some of which is due to the estimation error  $\mathbf{m}_{\text{est}} - \mathbf{m}_0$ . However, by estimating  $\mathbf{m}_{\text{est}}$  carefully and repeatedly (900 times), and by considering the resulting 900 sets of misfits  $\Phi$  (at 200,000 source-receiver pairs) jointly, the histogram of their 200,000  $D$ -values should approximate a histogram of the true  $\Phi(\mathbf{m}_0|\mathbf{d})$  as closely as we can hope to get. Figure 5a shows this empirically obtained histogram  $\Phi_{\text{cumulative}}$  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  $\Phi'$  encountered in a future (Bayesian) inversion, we could in principle compare it to this empirical histogram  $\Phi_{\text{cumulative}}$ . It would however be more convenient and computationally efficient to identify an analytical expression for the  $p(x)$  that produced this histogram  $\Phi_{\text{cumulative}}$ , and to evaluate any  $\Phi'$  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 figure 5a, thus obtaining a pdf  $p_{\text{fit}}(x)$  as our best estimate for the true  $p(x)$ .
- The likelihood of a data vector  $\mathbf{d}$  given model  $\mathbf{m}$  is then considered to be

$$\mathcal{L}(\mathbf{m}|\mathbf{d}) = p_{\text{fit}}(\Phi(\mathbf{d}|\mathbf{m})). \quad (18)$$



### 3.2 Approximate log-normal distribution of decorrelation $D$

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

5 The beta and the exponential distributions are seen to overestimate 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-  
 10 quantile plot of figure 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  $\mu$  and  $\sigma$

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right). \quad (19)$$

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

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

$$\mathcal{L}_D = \frac{\exp\left(-\frac{1}{2}(\ln(\mathbf{D}) - \boldsymbol{\mu})^T \mathbf{S}_D^{-1}(\ln(\mathbf{D}) - \boldsymbol{\mu})\right)}{(2\pi)^{\frac{n_S}{2}} \sqrt{|\det(\mathbf{S}_D)|}} \quad (20)$$

and the misfit:

$$20 \quad \Phi = \frac{1}{2} \left( \sum_j^n \sum_j^n (\ln(D_j) - \mu_j)^T (\mathbf{S}_D^{-1})_{jk} (\ln(D_k) - \mu_k) \right) + \frac{1}{2} \ln((2\pi)^{n_S} |\det(\mathbf{S}_D)|) \quad (21)$$

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

25 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 case of the  $\ell^1$  or  $\ell^2$  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 likelihood function  $\mathcal{L}$ .



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

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

This may be an oversimplification since ambient noise levels  $\epsilon_{\text{noise}}$  show significant diurnal and seasonal variations, and are elevated at stations close to coastlines or cities (Peterson, 1993; Stutzmann et al., 2000; McNamara and Buland, 2004; 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 backazimuth, so we might even require different scale parameters for each source-receiver pair.

To avoid this level of complexity, recall the investigation of section 2.3 that revealed the distribution of  $D$  to be most sensitive to the level of ambient noise  $\epsilon_{\text{noise}}$ . Hence we bin our 200,000 source-receiver pairs by signal-to-noise ratio ( $SNR$ ), and estimate only one pair of  $(\mu, \sigma)$  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 second noise window prior to the arrival of the first body wave energy. Signal time windows  $u_i, i = 1, \dots, N_{\text{signal}}$  are: 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  $n_i, i = 1, \dots, N_{\text{noise}}$  are: for both  $P$ - and  $SH$ -phases, -150 s to -30 s before theoretical arrival time. We calculate signal-to-noise ratios ( $SNR$ ) for  $P$ - and  $SH$ -waves as:

$$SNR = \frac{N_{\text{noise}} \sum_{i=1}^{N_{\text{signal}}} u_i^2}{N_{\text{signal}} \sum_{i=1}^{N_{\text{noise}}} n_i^2}. \quad (22)$$

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  $p_{\text{fit}}(D)$ , as a function of  $SNR$ . Under low-noise conditions (high  $SNR$ ), the log-normal distributions are narrower and centered on smaller  $D$  misfit values, which seems plausible.

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

Hence the log-normal distribution  $p_{\text{fit}}(D)$  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:

$$\mu_i = a_{\mu,1} + a_{\mu,2} \cdot \exp(a_{\mu,3} \cdot SNR_i) \quad (23)$$

$$\sigma_i = a_{\sigma,1} + a_{\sigma,2} \cdot \exp(a_{\sigma,3} \cdot SNR_i) \quad (24)$$



### 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  $\vartheta$  as follows. For each earthquake, we calculated the azimuthal distances  $\vartheta_{jk}$  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  $\vartheta$  (in bins of  $5^\circ$  width).

We need to adjust for the fact that stations  $j$  and  $k$  usually have different  $SNR$  and hence different  $\mu_j$  and  $\sigma_j$  in their log-normal distributions of  $D$ . Hence we calculate the standard score of each station  $j$  as  $z_j = (\ln(D_j) - \mu_j) / \sigma_j$  and from this the Pearson correlation coefficient of a  $\vartheta$ -bin  $\{j, k\}_\vartheta$ , using all  $n_\vartheta$  station pairs in that bin:

$$r(\vartheta) = \frac{1}{n_\vartheta - 1} \sum_{\{j,k\}_\vartheta} z_j z_k \quad (25)$$

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

$$g(\vartheta) = b_1 + b_2 \cdot \exp(-b_3 \vartheta^2) \quad (26)$$

(see fig. 7).

This permits to compare  $D_j$  for stations with different  $SNR$  and distributions of  $D_j$ .

Then the correlation coefficient was calculated for each azimuthal bin  $\vartheta$  using all  $n_\vartheta$  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. 20

$$\mathbf{S}_{D,i,j} = \begin{cases} \sigma_i \sigma_j \cdot (b_1 + b_2 \cdot \exp(-b_3 \vartheta^2)), & i \neq j \\ \sigma_i^2, & i = j \end{cases} \quad (27)$$

An example of such a covariance matrix is shown in figure 8. It is for the 2011 earthquake in the U.S. state of Virginia that was used for a detailed worked 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  $\log_{10}(A)$  in a 1-second time window around the peak  $i = i_1, \dots, i_2$  of the measured seismogram and its synthetic:

$$\Delta \log_{10}(A)_j = \log_{10} \left( \sum_{i=i_1}^{i_2} u_{j,i}^2 \right) - \log_{10} \left( \sum_{i=i_1}^{i_2} u_{j,i}^c{}^2 \right) \quad (28)$$



Again our goal it to approximate the distribution of this misfit in order to obtain an empirical likelihood function. The distribution of  $\Delta \log_{10}(A)$  is almost symmetric around zero, see fig. S2 in the electronic supplement. The amplitude misfit  $|\Delta \log_{10}(A)|$  approximately follows a Laplace distribution, where parameter  $k$  does not vary much with  $SNR$  (see Electronic Supplement). We construct the likelihood function

$$5 \quad \mathcal{L}_{\text{Amp}} = \sum_{j=1}^{n_S} \frac{1}{2k} \exp\left(-\frac{|\Delta \log_{10}(A)|}{k}\right), \quad (29)$$

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 localized site effects.

### 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:

- 10 1. For every new earthquake, download and archive a suitable selection of broadband, three-component, teleseismic seismograms ( $\Delta = 32^\circ$  to  $85^\circ$ ). A pragmatic approach 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).
- 15 2. Bandpass filter between 0.02 Hz and 1.0 Hz. Rotate horizontal components to *RTZ* system. Select signal time windows and noise time windows and calculate  $SNR$  as defined in eq. 22.
3. For each station, and for P and SH separately, use  $SNR$  to calculate distribution parameters  $\mu_i$  and  $\sigma_i$  from equation 24. Populate the diagonal of covariance matrix  $S_{\mathbf{D},ii}$  with the  $\sigma_i^2$ .
4. Estimate correlation coefficient  $r(\vartheta_{j,k})$  between two stations  $j, k$  using equation 27. Fill off-diagonal elements:

$$20 \quad S_{\mathbf{D},jk} = r(\vartheta_{j,k})\sigma_j\sigma_k \quad (30)$$

5. Insert  $\mu_i$  and  $S_{\mathbf{D}}$  in the likelihood equation 20 and combine with  $\mathcal{L}_{\text{Amp}}$  (29) to create the total likelihood function

$$\mathcal{L} = \mathcal{L}_{\mathbf{D}} + \mathcal{L}_{\text{Amp}}. \quad (31)$$

This enable the evaluation of randomly sampled candidate solutions for earthquake source parameters  $\mathbf{m}$ . Parametrization of  $\mathbf{m}$ , Bayesian sampling strategy, and construction of the posterior distribution of  $\mathbf{m}$  is described in the companion paper (Stähler and Sigloch, 2014).

25



#### 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  $\ell^2$ -norm misfit, and the Laplace distribution, associated with the  $\ell^1$ -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  $\ell^1$  or  $\ell^2$ -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 time 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 ( $\ell^1$ -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  $\ell^2$ -norm, outliers disproportionately 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  $\ell^1$  norm is more robust against outliers, and with the same motivation distance norms with non-integer exponents  $\ell^p$  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 because the idealized circumstances of the Central Limit Theorem and multiple, comparable noise sources rarely apply. If a measured seismogram of length  $N$  samples is considered

$$u_i = u_i^c + \epsilon_{i,\text{noise}}, \quad (32)$$

then an  $(N \times N)$  covariance matrix for  $\epsilon_{\text{noise}}$  needs to be estimated under the  $\ell^2$  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), 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 (“ $+\epsilon_{\text{noise}}$ ”). We have argued that our noise model needs to be

$$u_i = u_i^c * T_{\text{model}}(t) * T_{\text{inst}}(t) + \epsilon_{\text{noise}}, \quad (33)$$

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 analytical expression of which will be unknown.

Another cause for leaving the Gaussian/ $\ell^2$  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).

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 stations 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  $D = 1 - CC$  of seismological broadband waveforms as their measure of data (mis-)fit – including 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 Author contributions

S. C. S. conceived of the concept of the empirical likelihood and did the data analysis. K. S. wrote the original source inversion code and created the earthquake database for the correlation misfit statistics. Both authors shared in the writing of the paper.

*Acknowledgements.* We thank M. Sambridge, R. Zhang, H. Igel and B. L. N. Kennett for fruitful discussions in an earlier stage of the work. S. C. S. was supported by the Munich Centre of Advanced Computing (MAC) of the International Graduate School on Science and Engineering (IGSSE) at Technische Universität München. IGSSE also funded his research stay at the Research School for Earth Sciences at A. N. U. in Canberra, where part of this work was carried out. K.S. acknowledges funding by ERC Grant 639003 “DEEPTIME” and by the Leverhulme Trust.



## References

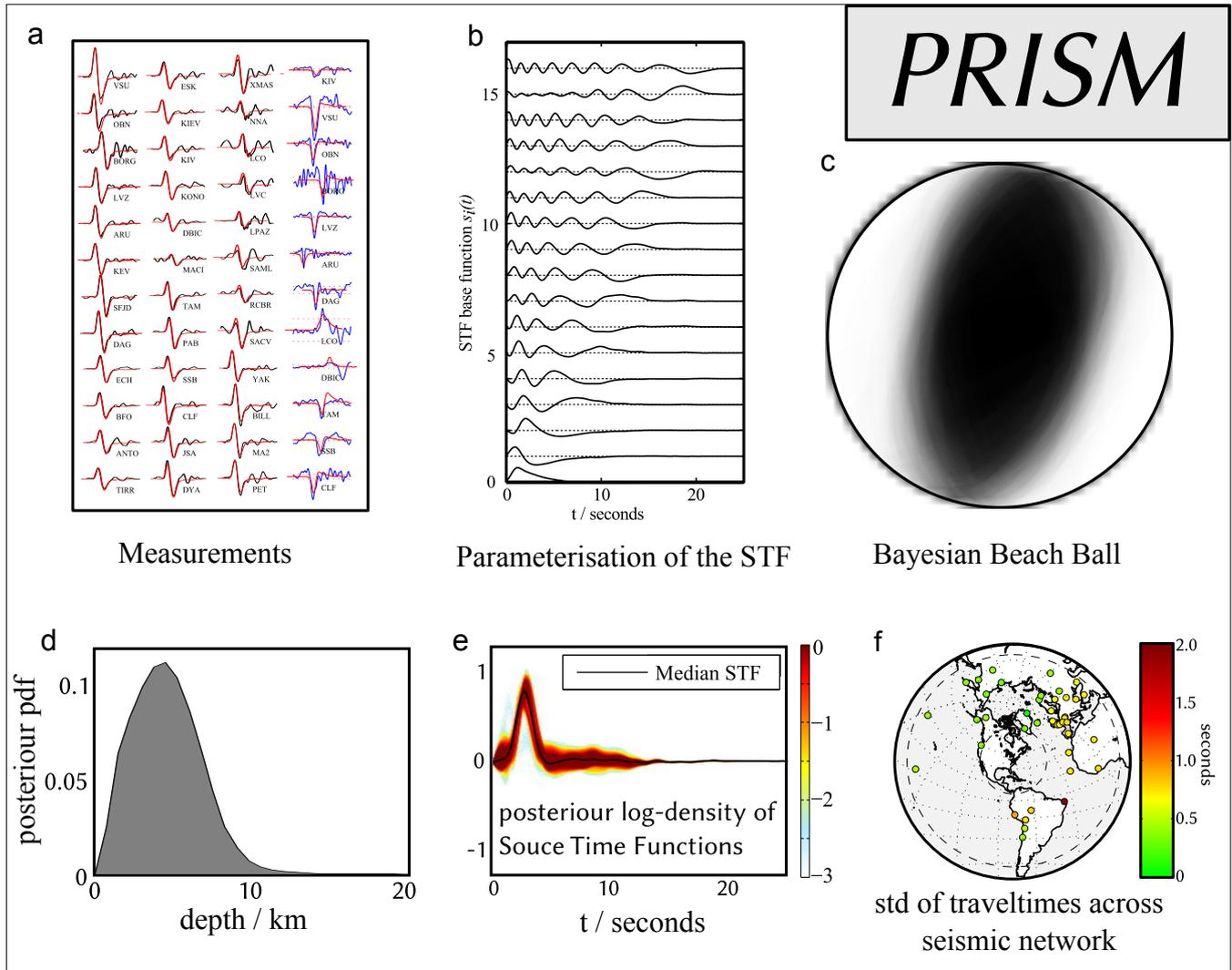
- Bodin, T.: Transdimensional Approaches to Geophysical Inverse Problems, Ph.D. thesis, Australian National University, 2010.
- Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P.: Transdimensional tomography with unknown data noise, *Geophys. J. Int.*, pp. no–no, doi:10.1111/j.1365-246X.2012.05414.x, <http://doi.wiley.com/10.1111/j.1365-246X.2012.05414.x>, 2012.
- 5 Bogert, B.: Correction of seismograms for the transfer function of the seismometer, *Bull. Seismol. Soc. Am.*, 52, 781–792, <http://www.bssaonline.org/content/52/4/781.short>, 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, <http://doi.wiley.com/10.1111/j.1365-246X.2011.05107.x><http://onlinelibrary.wiley.com/doi/10.1111/j.1365-246X.2011.05107.x/full>, 2011.
- 10 Chapman, C. H.: A new method for computing synthetic seismograms, *Geophys. J. R. Astron. Soc.*, 54, 481–518, 1978.
- 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, [http://dx.doi.org/10.1016/0031-9201\(81\)90046-7](http://dx.doi.org/10.1016/0031-9201(81)90046-7)<http://linkinghub.elsevier.com/retrieve/pii/0031920181900467>, 1981.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J.: *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC, London, [http://scholar.google.com/scholar?hl=en{%&btnG=Search{%&q=intitle:Markov+Chain+Monte+Carlo+in+Practice+\(Interdisciplinary+Statistics\){#}1](http://scholar.google.com/scholar?hl=en{%&btnG=Search{%&q=intitle:Markov+Chain+Monte+Carlo+in+Practice+(Interdisciplinary+Statistics){#}1), 1996.
- 15 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, <http://doi.wiley.com/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.
- 20 Käüfl, 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, <http://dx.doi.org/10.1111/j.1365-246X.1991.tb06724.x><http://www3.interscience.wiley.com/journal/119359470/abstracthttp://onlinelibrary.wiley.com/doi/10.1111/j.1365-246X.1991.tb06724.x/abstract>, 1991.
- 25 Kikuchi, B. Y. M. and Kanamori, H.: Inversion of complex body waves - III, *Bull. Seismol. Soc. Am.*, 81, 2235–2350, <http://bssa.geoscienceworld.org/content/81/6/2335.short>, 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.
- 30 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, <http://www.math.chalmers.se/Math/Research/Preprints/2010/47.pdf><http://www.sciencedirect.com/science/article/pii/S0047259X12000516><http://linkinghub.elsevier.com/retrieve/pii/S0047259X12000516>, 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, <http://www.bssaonline.org/cgi/doi/10.1785/0120060012>, 2006.
- 35 Kummerow, J.: Using the value of the crosscorrelation coefficient to locate microseismic events, *Geophysics*, 75, MA47, doi:10.1190/1.3463713, <http://geophysics.geoscienceworld.org/content/75/4/MA47.shorhttp://link.aip.org/link/GPYSA7/v75/i4/pMA47/s1{%&Agg=doi>, 2010.



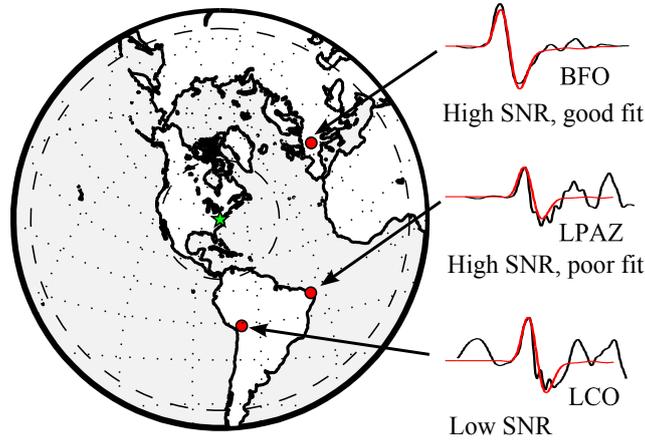
- Larose, E., Planès, T., Rossetto, V., and Margerin, L.: Locating a small change in a multiple scattering environment, *Appl. Phys. Lett.*, 96, 2010–2012, doi:10.1063/1.3431269, 2010.
- Mahalanobis, P. C.: On the generalized distance in statistics, *Proc. Natl. Inst. Sci. India*, 2, 49–55, <http://scholar.google.com/scholar?hl=en{%&}btnG=Search{%&}q=intitle:On+the+generalized+distance+in+statistics{#}0>, 1936.
- 5 Malinverno, A. and Briggs, V. A.: Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes, *Geophysics*, 69, 1005–10 016, doi:10.1190/1.1778243, <http://library.seg.org/doi/abs/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, <http://www.bssaonline.org/cgi/content/full/90/6/1369http://bssaonline.org/cgi/content/abstract/90/6/1369>, 2000.
- 10 McNamara, D. E. and Buland, R.: Ambient Noise Levels in the Continental United States, *Bull. Seismol. Soc. Am.*, 94, 1517–1527, doi:10.1785/012003001, <http://bssa.geoscienceworld.org/content/94/4/1517.fullhttp://bssa.geoscienceworld.org/cgi/doi/10.1785/012003001>, 2004.
- Menke, W.: Using waveform similarity to constrain earthquake locations, *Bull. Seismol. Soc. Am.*, 89, 1143–1146, <http://www.bssaonline.org/content/89/4/1143.short>, 1999.
- 15 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, <http://www.bssaonline.org/content/80/2/430.short>, 1990.
- 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, <http://onlinelibrary.wiley.com/doi/10.1046/j.1365-246x.1999.00900.x/full>, 1999.
- 20 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, <http://www.springerlink.com/index/LR4J5Y95YH57PF09.pdf>, 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, <http://srl.geoscienceworld.org/cgi/doi/10.1785/0220120103>, 2013.
- 25 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, <http://www.agu.org/pubs/crossref/2011/2010GC003421.shtmlhttp://doi.wiley.com/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, <http://www.nature.com/nature/journal/v496/n7443/abs/nature12019.htmlhttp://www.ncbi.nlm.nih.gov/pubmed/23552944>, 2013.
- 30 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, <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-246X.2006.03116.x/abstract>, 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, <http://dx.doi.org/10.1038/ngeo231http://www.nature.com/ngeo/journal/v1/n7/abs/ngeo231.html>, 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, <http://www.solid-earth.net/5/1055/2014/>, 2014.



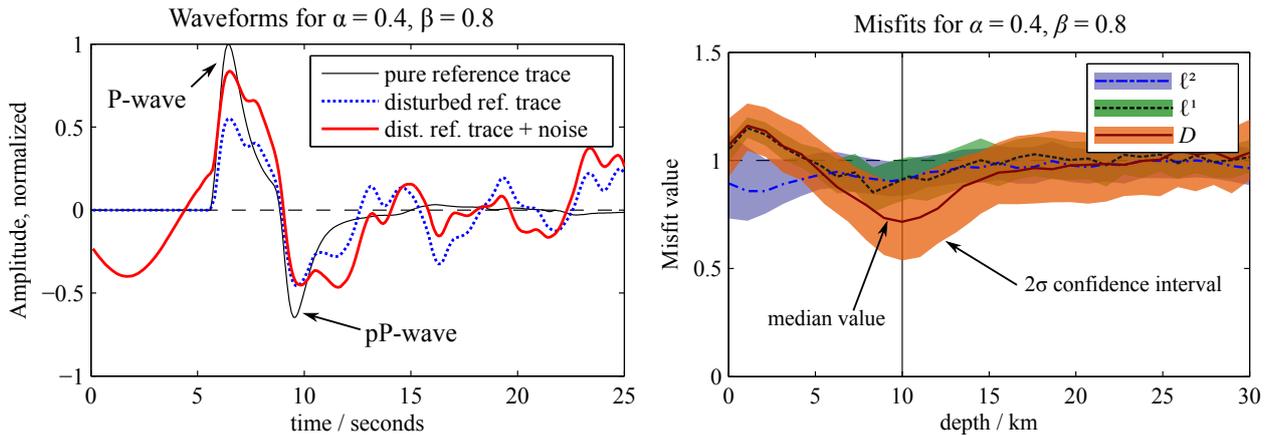
- Stähler, S. C., Sigloch, K., and Nissen-Meyer, T.: Triplicated P-wave measurements for waveform tomography of the mantle transition zone, *Solid Earth*, 3, 339–354, doi:10.5194/se-3-339-2012, <http://www.solid-earth.net/3/339/2012/>, 2012.
- Stutzmann, E., Roult, G., and Astiz, L.: GEOSCOPE station noise levels, *Bull. Seismol. Soc. Am.*, 3, 690–701, <http://bssa.geoscienceworld.org/content/90/3/690.short>, 2000.
- 5 Stutzmann, E., Schimmel, M., Patau, G., and Maggi, A.: Global climate imprint on seismic noise, *Geochemistry, Geophys. Geosystems*, 10, doi:10.1029/2009GC002619, 2009.
- Tape, C., Liu, Q., Maggi, A., and Tromp, J.: Adjoint tomography of the southern California crust., *Science* (80-. ), 325, 988–92, doi:10.1126/science.1175298, <http://www.sciencemag.org/content/325/5943/988.abstract>, 2009.
- Tape, W. and Tape, C.: A uniform parametrization of moment tensors, *Geophys. J. Int.*, 202, 2074–2081, doi:10.1093/gji/ggv262, <http://gji.oxfordjournals.org/lookup/doi/10.1093/gji/ggv262>, 2015.
- 10 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, <http://www.ipgp.fr/~tarantola/Files/Professional/Papers{ }PDF/IP{ }QI{ }latex.pdf>, 1982.
- Vallée, M. and Douet, V.: A new database of Source Time Functions (STFs) extracted from the SCARDEC method, *Phys. Earth Planet. Inter.*, in press, 2016.
- 15 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, <http://doi.wiley.com/10.1111/j.1365-246X.2010.04836.x>, 2011.



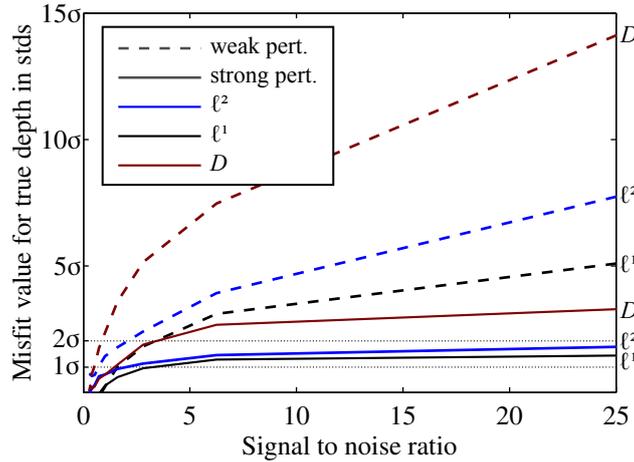
**Figure 1.** Visual summary of the fully probabilistic source inversion algorithm PRISM presented in the companion paper (Stähler and Sigloch, 2014), on the example of a magnitude 5.7 earthquake in the U.S. state of Virginia on 2011/08/23. a) Candidate source solutions are evaluated according to the cross-correlation fit they produce between observed broadband, teleseismic *P*-waveforms (black) or *SH*-waveforms (blue), and their modeled 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 hypocenter depth. e) Weighted average of STFs from the posterior ensemble of good solutions permits to assess 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 traveltimes measurements at all stations, and their inter-station correlations. Traveltimes are the primary input data for seismic tomography and these insights into their uncertainties are not readily available from other methods.



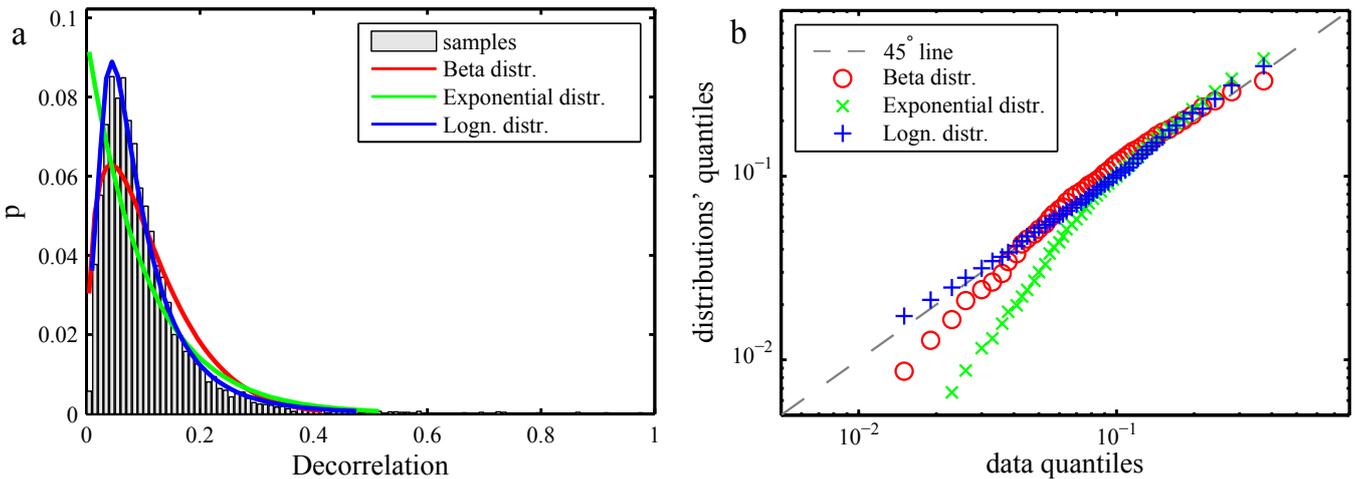
**Figure 2.** Three noise cases for compressional waves in source inversion, the waveforms were produced by the  $M5.7$  earthquake in Virginia (2011/08/23). Station BFO has a high signal-to-noise ratio (no wiggles preceding the P-pulse) and the waveform is fit well by a WKBJ synthetic using our best source solution for this earthquake. Station LPAZ has a high signal-to-noise ratio, but 3D 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, but 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).



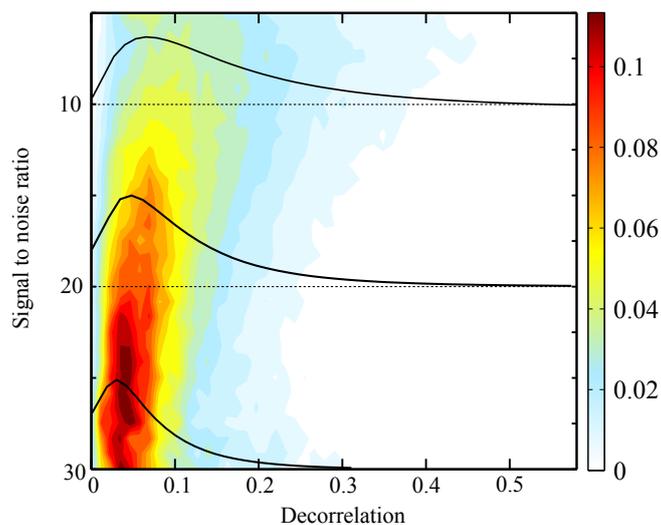
**Figure 3.** Comparison of the  $\ell^1$ ,  $\ell^2$  norm and the *signal decorrelation*  $D = 1 - CC$  as misfit criteria in noisy signals. A perturbed synthetic waveform  $u_{\text{pert}}^c$  for a 10 km deep explosion source, measured in  $40^\circ$  distance was compared to synthetic seismograms  $u^c$  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 recognizing reference signals in perturbed measurements. For better visualisation, all misfit values have been normalized separate to have an average values of one between 20 and 30 km. The normalization coefficients are plotted in the top right corner. See fig. S1 for more noise and perturbation levels.



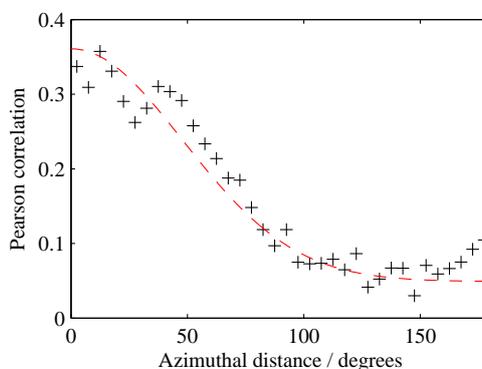
**Figure 4.** Distance between misfit value for the true source depth vs the plateau for depths 20-30 km in standard deviations. See fig. 3 for waveforms and misfit curves. The "weak perturbation" curve is calculated with perturbation factor  $\alpha = 0.1$  and the "strong perturbation" curve with  $\alpha = 0.9$  (see eq. 16). For all SNR values, the decorrelation has a higher discriminative value than  $\ell^1/\ell^2$



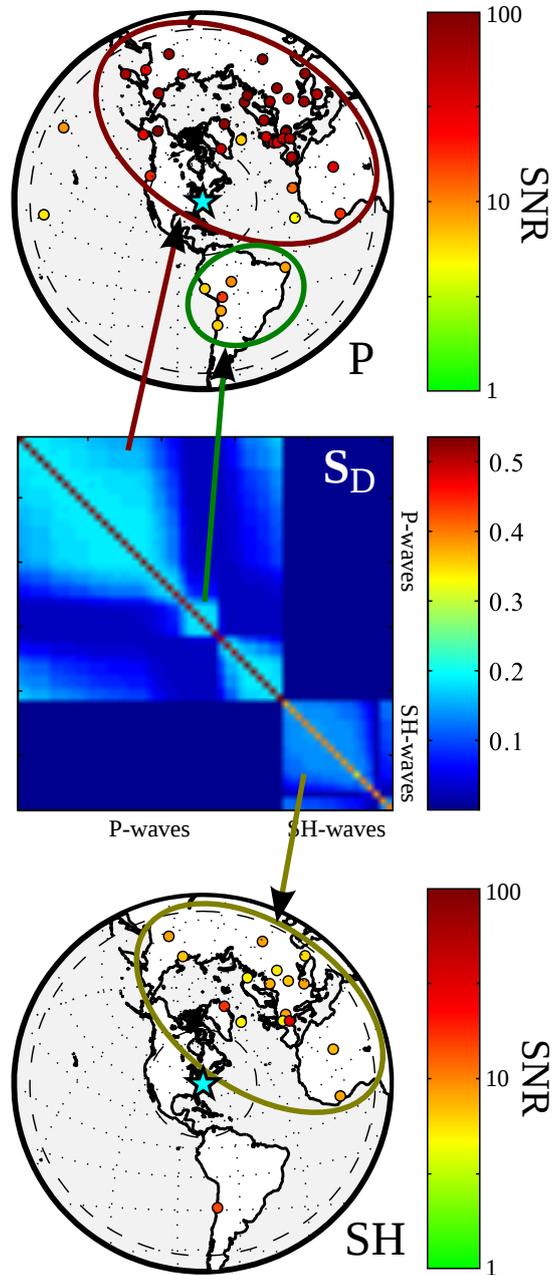
**Figure 5.** Probability distribution of  $D$ , the decorrelation of measured and synthetic  $P$ -waveforms used for deterministic source inversions. a) Empirical histogram of  $D$  is shown as grey bars. From 900 source inversions based on 200,000 broadbanded, teleseismic  $P$ -waveforms, only waveforms with signal-to-noise ratios between 20.0 and 21.0 were considered for this figure (because the scaling parameters of analytical fitting functions depend mainly on SNR). Coloured lines show best-fitting realisations of three analytical probability density distributions: beta (red), exponential (green), 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  $D$ .



**Figure 6.** Colour shades 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).



**Figure 7.** Correlation in misfit between neighbouring stations. The measured Pearson correlation (see eq. 25 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.



**Figure 8.** Visualisation of an inter-station covariance matrix  $S_D$  for misfit  $D$  (centre panel, c.f. equation 20), on the example of an  $m_b$  5.7 earthquake that occurred in the U.S. 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); 3. a northern hemisphere cluster of SH-waveforms measurement (olive). P- and SH measurements are modelled as being uncorrelated.