Accounting for theory errors with empirical Bayesian noise models in nonlinear centroid moment tensor estimation

H. Vasyura-Bathke[*,1,2], J. Dettmer[3], R. Dutta[1], P.M. Mai[1], S. Jónsson[1]

[1] *King Abdullah University of Science and Technology, Thuwal* 23955-6900*, Saudi Arabia*

[2] *Now at: University of Potsdam, D-14476 Potsdam, Germany*

[3] *University of Calgary, Calgary, Canada*

** Email: hvasbath@uni-potsdam.de*

Word Counts

SUMMARY

Centroid moment-tensor (CMT) parameters can be estimated from seismic waveforms. Since these data indirectly observe the deformation process, CMTs are inferred as solutions to inverse problems which are generally under-determined and require significant assumptions, including assumptions about data noise. Broadly speaking, we consider noise to include both theory and measurement errors, where theory errors are due to assumptions in the inverse problem and measurement errors are caused by the measurement process. While data errors are routinely included in parameter estimation for full CMTs, less attention has been paid to theory errors related to velocity-model uncertainties and how these affect the resulting moment-tensor (MT) uncertainties. Therefore, rigorous uncertainty quantification for CMTs may require theory-error estimation which becomes a problem of specifying *noise models*.

Various noise models have been proposed, and these rely on several assumptions. All approaches quantify theory errors by estimating the covariance matrix of data residuals. However, this estimation can be based on explicit modelling, empirical estimation, and/or ignore or include covariances. We quantitatively compare several approaches by presenting parameter and uncertainty estimates in non-linear full CMT estimation for several simulated data sets and regional field data of the M_l 4.4, 13 June 2015 Fox Creek, Canada, event. While our main focus is at regional distances, the tested approaches are general and implemented for arbitrary source model choice. These include known or unknown centroid locations, full MTs, deviatoric MTs, and double-couple MTs. We demonstrate that velocity-model uncertainties can profoundly affect parameter estimation and that their inclusion leads to more realistic parameter uncertainty quantification. However, not all approaches perform equally well. Including theory errors by estimating non-stationary (non-Toeplitz) error covariance matrices via iterative schemes during Monte Carlo sampling performs best and is computationally most efficient. In general, including velocity-model uncertainties is most important in cases where velocity structure is poorly known.

Key words: Bayesian inference, seismic data, velocity model uncertainties, moment tensors

1 INTRODUCTION

Earthquakes are routinely monitored by broadband seismic networks. Initial source analysis is often based on solving a weighted time-domain least squares inverse problem to obtain seismic moment tensors that assume a point source with fixed location, fixed source-time-function (STF) and simple velocity structure (e.g., Sipkin 1982; Koch 1991; Tocheport et al. 2007). These simplifying assumptions can result in erroneous estimates of the parameters of the moment tensor (MT) (e.g., Šílený et al. 1992; Kravanja et al. 1999). Thus, a more comprehensive approach is to determine the location, the STF and the MT parameters simultaneously (e.g., Kravanja et al. 1999; Wéber 2006; Sigloch & Nolet 2006; Ekström 2006; Ekström et al. 2012; Stähler & Sigloch 2014, 2016). In addition, these source parameters should be quantified not only in terms of their optimal parameter values, but also in terms of their uncertainties. Uncertainty quantification can be accomplished by formulating the problem via Bayes' Theorem (e.g., Tarantola 2005; Wéber 2006; Dębski 2008; Stähler & Sigloch 2014; Vackář et al. 2017).

The physical processes of earthquake deformation have significant non-linearities in source parameters (Cesca et al. 2016), especially for the origin in space and time, which causes numerical challenges in determining source location and mechanism. In addition, seismic data are contaminated by various noise sources of natural (e.g., meteorological and oceanic) and human origins (Bonnefoy-Claudet et al. 2006). The estimation of noise characteristics is important to obtain appropriate weights for the data in the parameter inference. A simple approach is to estimate the pre-event noise variance and to derive a diagonal weight matrix (e.g., Duputel et al. 2012). In addition, the covariances between seismogram components can be estimated and these can account for the directionality of seismic noise (Tarantola 2005; Vackář et al. 2017). Accounting for such dependence in noise and its rigorous quantification leads to better estimation of the MT parameters and their uncertainties.

For inverse problems, it has been shown that both data errors and and errors due to assumptions in the model formulation affect parameter uncertainty, theory errors in the following (Tarantola & Valette 1982). In source parameter estimation, significant assumptions are made about the Earth structure (e.g., Tarantola & Valette 1982; Duputel et al. 2014) and the parameterisation of the deformation source (e.g., Dettmer et al. 2014; Pugh et al.

2016). For example, theory error can be due to a pre-defined earthquake hypocentre location, but this location is inconsistent with the centroid location (Duputel et al. 2012; Ragon et al. 2018). Another example is assuming the STF to be of particular shape (e.g., triangular) that is not sufficiently general to describe the moment release of the source (Stähler & Sigloch 2014). Yet another important source of theory error is the Earth structure (Minson et al. 2013). While actual structure is 3D, anisotropic and heterogeneous, it is often approximated by isotropic, horizontally stratified half-spaces. Errors due to these assumptions have mostly been ignored in source studies (e.g., Hofstetter et al. 2003; Fukuda & Johnson 2008; Baer et al. 2008; Bathke et al. 2013). In addition, trade-offs between source parameterisation and Earth-structure can cause poor assumptions about structure to be compensated by biased estimates of source parameters (e.g., Valentine & Woodhouse 2010).

Recent research incorporated uncertainties in the assumed Earth structure into distributed slip-estimates of extended sources through a prediction covariance matrix. For instance, Yagi & Fukahata (2011) included a Gaussian noise term for teleseismic Green's functions (GFs) and iteratively estimated a prediction covariance matrix in an optimization scheme employing Akaike's Bayesian information criterion (ABIC). Similarly, Minson et al. (2013) estimated a scale factor for an identity matrix that treats the variance in GFs to account for uncertainty in the subsurface structure in Bayesian inference. With linear perturbations of the original GFs, a prediction covariance matrix including off-diagonal terms can be formulated (Duputel et al. 2014). This approach includes physical constraints to improve the robustness of finite-fault inversion (Yagi & Fukahata 2008, 2011; Minson et al. 2013; Duputel et al. 2014). Incorporating a prediction covariance matrix to resolve distributed kinematic rupture parameters for data computed from a synthetic dynamic rupture model, Razafindrakoto & Mai (2014) reported loss in resolution on the kinematic rupture parameters through Bayesian inference. However, they investigated only the variance effect in the prediction covariance matrix. In MT estimation, the MT components can be more robustly inferred by including the centroid uncertainty (Duputel et al. 2012). Hallo & Gallovic (2016) showed that including uncertainties in Earth structure in Bayesian linear MT estimation yields more reliable estimates and uncertainties. These developments mostly focused on improving the robustness of linear inversion under the premise that the source geometry and location was known *a-priori*. However, it remains unclear if improvements can be achieved when including non-linear parameters (e. g., source location) in the inference.

In this study, we compare various approaches to estimate covariance matrices with respect to uncertainties in Earth velocity models and we show how to include these in Bayesian inference. For simplicity, we approximate the STF as a delta function, which is a valid assumption if the source duration is shorter than the shortest periods in the waveforms (Aki & Richards 2002). In synthetic tests, we demonstrate the influence of various parameterisations of the covariance matrix on parameter estimates of full CMTs, deviatoric (DV) CMTs and double-couple (DC) CMTs. We apply the approach to regional seismic data for the 13th June 2015, Fox Creek (Canada) event.

2 METHODS

This section provides background information on source parameter estimation with Bayesian inference. In particular, we consider how uncertainties in Earth structure (i.e., layer depths and velocities) are propagated to source parameter uncertainties by estimating theory errors in terms of noise covariance matrices.

2.1 Bayesian Inference

Bayes' theorem (Bayes 1763) has been widely applied to study earthquake sources (e.g., Tarantola & Valette 1982; Wéber 2006; Monelli & Mai 2008; Monelli et al. 2009; Fukuda & Johnson 2008; Duputel et al. 2012; Minson et al. 2013; Dettmer et al. 2014; Razafindrakoto & Mai 2014; Vackář et al. 2017; Dutta et al. 2018). Recently, we introduced a flexible software (BEAT-Bayesian Earthquake Analysis Tool) for source estimations in layered elastic halfspaces with Bayesian inference (Vasyura-Bathke et al. 2019, 2020). Using this software, we estimate parameters $\mathbf{m}$ of nonlinear CMT parameterisations (see Appendix A) from seismic data $\mathbf{d}^{obs}$, i.e., seismic displacement waveforms.

Assuming Gaussian-distributed noise on the data, a likelihood function is straightforward to formulate. However, since data noise cannot generally be determined independently, residual errors $\mathbf{r}(\mathbf{m}) = \mathbf{d}^{obs} - \mathbf{d}(\mathbf{m})$ serve as a proxy. The posterior probability density (PPD) for residual errors of K datasets is given by (Tarantola & Valette 1982)

$$p(\mathbf{m}|\mathbf{d}^{obs}) \propto p(\mathbf{m}) \prod_{k=1}^{K} \frac{1}{(2\pi)^{N/2}|\mathbf{C}_k|^{1/2}} \exp\left[-\frac{1}{2}[\mathbf{d}_k^{obs} - \mathbf{d}_k(\mathbf{m})]^T \mathbf{C}_k^{-1}[\mathbf{d}_k^{obs} - \mathbf{d}_k(\mathbf{m})]\right], \quad (1)$$

where $\mathbf{d}_k(\mathbf{m})$ are predicted seismic data at seismic station k with N samples that depend on the MT parameters $\mathbf{m}$. This formulation assumes that for a seismic station k the noise at different components is independent. The covariance matrices $\mathbf{C}_k$ represent the noise statistics, and play an important role in the parameter estimation as well as in the uncertainty quantification.

2.2 Residual error covariance matrix

The residual covariance matrices include variances and covariances of the data residuals $\mathbf{r}_k$. Under the assumption that noise between stations is not correlated, one matrix is required for each station. We study five approaches of formulating parameterisations for the noise covariance matrix (Fig. 2). The *variance* approach estimates the noise standard deviation as a hierarchical noise model (Malinverno & Briggs 2004) and ignores covariances. Another hierarchical model is realised with a simple function describing off-diagonal components (*exponential* parameterisation). More sophisticated approaches model theory errors explicitly (*variance_cov* and *exponential_cov* parametrisations). Finally, the *non-Toeplitz* approach is empirical and non-parametric. In the following we use the terms: *variance*, *exponential*, *variance_cov*, *exponential_cov* and *non-Toeplitz* to distinguish between the different covariance parameterisations described below and listed in Tab. 1.

The total noise covariance matrix $\mathbf{C}_k$ at station k is the sum of the data covariance matrix $\mathbf{C}_k^d$ that quantifies measurement errors and the model prediction covariance matrix $\mathbf{C}_k^t$, caused by physical and mathematical approximations in the forward model (theory errors),

$$\mathbf{C}_k = \mathbf{C}_k^d + \mathbf{C}_k^t. \quad (2)$$

Many MT studies ignore off-diagonal terms in $\mathbf{C}_k^d$ and the component $\mathbf{C}_k^t$ (e.g., Cesca et al. 2017; Ekström 2006; Ekström et al. 2012; Vackář et al. 2017). Consequently, only measurement errors are considered and assumed to be from a stationary, uncorrelated random Gaussian process (Fig. 1a, *variance*). This assumption is often unjustified when noise is serially correlated and/or non-stationary. For long-period data, it can be useful to estimate diagonal (Toeplitz) covariance matrices (Fig. 1b, *exponential*) with exponential decay depending on the shortest period t_0 of the data (Duputel et al. 2012, see Tab. 1). For $\mathbf{C}_k^d$ variances, σ^2 can be estimated from the recorded signal, filtered to the frequency band used in the inference, prior to the first arriving wave of the seismic event of interest at any given station. However, it must be ensured that there is no source of seismic signal other than background noise present in the estimation data; otherwise biases occur.

2.2.1 Explicit modeling of theory errors

Theory errors can result in source parameter uncertainties that are substantially larger than those due to measurement errors (Tarantola & Valette 1982). Here, we compare various approaches to accounting for velocity structure errors in the noise covariance matrix $\mathbf{C}_k$. First we consider a previously proposed strategy (Tarantola & Valette 1982; Yagi & Fukahata 2011; Duputel et al. 2014) to include theory error due to Earth-structure assumptions via the model prediction covariance matrix $\mathbf{C}_k^t$. We assume a horizontally stratified, elastic, isotropic half-space with uncertainties in the velocity-depth profile. One approach to estimate $\mathbf{C}_k^t$ in this case is to perturb the GFs that relate changes in velocity profile linearly to the displacements at the Earth's surface (Du et al. 1994; Duputel et al. 2014). Therefore, we calculate the GFs for various velocity models, where layer velocities and depths are varied in the crust by Gaussian perturbations with 10% standard deviation around the reference model (Mooney 1989) to generate an ensemble of Earth structures. From this ensemble, N_e sets of elementary GFs are computed and efficiently stored (Heimann et al. 2019). Each set of GFs is stored as a grid that covers all potential combinations of depths and distances in a source-receiver volume. If the source-receiver configuration falls between grid points during the sampling, GFs are linearly interpolated (Heimann et al. 2019).

Let i and j be indices for the rows and columns of the covariance matrix. Then, term $\bar{\mathbf{d}}_{k,i} = \frac{1}{N_e} \sum_{n=1}^{N_e} \mathbf{d}_{k,i}^n(\mathbf{m})$ is the sample mean over N_e predicted data vectors at station k (a similar term is defined for j) and the covariance matrix $\mathbf{C}_k^t$ is (Duputel et al. 2012)

$$\mathbf{C}^t_{k,ij}(\mathbf{m}) = \frac{1}{N_e}\sum_{n=1}^{N_e}(\mathbf{d}^n_{k,i}(\mathbf{m}) - \bar{\mathbf{d}}_{k,i})(\mathbf{d}^n_{k,j}(\mathbf{m}) - \bar{\mathbf{d}}_{k,j}). \quad (3)$$

This matrix is computed with respect to source parameters $\mathbf{m}$ while predicted data $\mathbf{d}^n_k$ are computed for each realization of Earth structure n (sets of GFs) and for each seismic station k. This covariance matrix $\mathbf{C}^t_k$ can be included in the likelihood function for inference following eqns. (1) and (2). Such formulation implies computing the synthetic seismic waveforms for each variation in the Earth structure (Fig. 2). As it is prohibitively expensive to calculate a realization of $\mathbf{C}^t_k$ for each iteration of a Monte Carlo algorithm, we assume that $\mathbf{C}^t_k$ changes less rapidly than the source parameters $\mathbf{m}$ in the sampling algorithm and we update it only periodically (Duputel et al. 2014). This approach accounts for errors in subsurface structure in addition to data errors in the estimation of source parameters and their uncertainties. Figure 1 (c, *variance_cov* & d, *exponential_cov*) demonstrates that theory errors due to Earth structure result in non-stationary covariance matrices with time-dependent error statistics. The computation of $\mathbf{C}^t_k$ is expensive and depends on the assumed variability of the Earth structure. If this variability is poorly known, the approach may result in over- or under-estimated parameter uncertainties.

2.2.2 Non-Toeplitz covariance matrix

A fast and non-parametric, alternative approach is to estimate non-stationary/non-Toeplitz covariance matrices $\mathbf{C}_k$ (Fig. 1e, *non-Toeplitz*) (Dettmer et al. 2007). These naturally include both data and theory errors as they are based on data residuals. Note that theory errors in this case are not limited to the explicitly modelled errors in layer velocities and layer depths (Sec. sec:explicitmod), but can also represent other sources of theory, e.g. anisotropy, errors in centroid location. An initial estimate $\tilde{\mathbf{m}}$ of the model parameters $\mathbf{m}$ is required to calculate data residuals $\mathbf{r}(\tilde{\mathbf{m}}) = \mathbf{d}^{obs} - \mathbf{d}(\tilde{\mathbf{m}})$ with number of samples N. Such $\tilde{\mathbf{m}}$ can be either obtained from prior information or from solving eq. 1 under the assumption of uncorrelated data errors. Standard deviations for the data residuals are estimated by a running average with a window of length M

$$\sigma_i = \sqrt{\frac{1}{M}\sum_{l=i-M/2}^{i+M/2} \mathbf{r}_l(\tilde{\mathbf{m}})^2}. \quad (4)$$

The vector $\boldsymbol{\sigma}$ containing the σ_i is used to standardize the data residuals $\mathbf{n}(\tilde{\mathbf{m}}) = \mathbf{r}(\tilde{\mathbf{m}})/\boldsymbol{\sigma}$, where division is element by element. The biased estimate of the autocovariance function of the scaled residuals is used to estimate correlation

$$c_i = \frac{1}{N}\sum_{j=0}^{N-i-1} (\mathbf{n}_{i+j}(\tilde{\mathbf{m}}) - \bar{\mathbf{n}}(\tilde{\mathbf{m}}))(\mathbf{n}_j(\tilde{\mathbf{m}}) - \bar{\mathbf{n}}(\tilde{\mathbf{m}})), \quad (5)$$

where i and j are data indices, and $\bar{\mathbf{n}}$ is the mean of $\mathbf{n}$. The c_i are used to fill the i-th diagonal of a square matrix ($N \times N$), yielding an unscaled covariance matrix $\tilde{\mathbf{C}}$. The *non-Toeplitz* covariance matrix estimate is obtained by scaling according to

$$\mathbf{C}_{ij} = \tilde{\mathbf{C}}_{ij}\sigma_i\sigma_j. \quad (6)$$

3 SIMULATION RESULTS

3.1 Simulated Data

To demonstrate the effect of the covariance matrix parameterisation and the influence of velocity-model uncertainties in earthquake source-parameter estimations, we present two simulated test cases. We generate two sets of simulated seismic displacement waveforms based on two different Earth structures (Tab. 2, Fig. 2a, blue and red lines) for a DC MT source (Tab. 3). We refer to these Earth structures as *reference structures* in the following. For each test case, we estimate the source parameters of a full CMT using the simulated data with the five different covariance matrix parameterisations (Tab. 1, Sec. 2.2).

In these test cases, we simulate theory errors due to unknown Earth structure by assuming a different Earth structure for source estimation than that of the reference model. We refer to this modified structure as the *estimation structure*. If no local Earth model is available in the study region, one would typically use a global model for the estimation. Here, we employ the AK135 velocity model (Kennett et al. 1995) for greater depth ($> 50\,km$) in combination with CRUST2 (Bassin et al. 2000) for shallow depth ($< 50\,km$) as the estimation structure for each test case (Fig. 2a). In case 1, the reference structure has the same number of layers as the estimation structure, but layer velocities and depths differ < 10% (Tab. 2, Fig. 2a). In case 2, the reference structure (Hofstetter et al. 2003) differs significantly from the estimation structure with a different number of layers, and different layer velocities and depth values (Fig. 2a).

Reference synthetic kinematic displacements for both cases are computed with frequencies up to 2 Hz for ten seismic stations at regional (up to 1000 km) epicentral distances (Tab. 3, Fig. 3). We added uncorrelated, Gaussian-distributed noise with a variance of 5% of the maximum waveform amplitude for each station. Data were filtered between 0.01 and 0.1 Hz, and rotated to radial (R), transverse (T) and vertical (Z) directions. MT parameters and the centroid location were estimated from the T and Z waveforms containing body- and surface-waves. For each test case, we estimated marginal distributions of source parameters while only changing the noise parameterisation (Fig. 1, Tab. 1), to demonstrate the influence of $\mathbf{C}_k$ on the results. Following the procedure in Sec. 2.2, the estimation structure was randomly perturbed 20 times to estimate $\mathbf{C}_k^t$ during sampling. The GFs are sampled at 1 Hz with 1 km grid spacing for depths from 0 to 15 km and distances from 0 to 1000 km using QSEIS (Wang 1999). The PPDs are estimated numerically with a sequential Monte Carlo (SMC) sampler (Moral et al. 2006; Vasyura-Bathke et al. 2020, Appendix B).

3.2 Results

For case 1 (small theory errors), estimation results are summarized in Fig. 4 in terms of posterior marginal probability densities. A notable observation is that when only applying $\mathbf{C}_k^d$ (i.e., ignoring theory error), the ranges of values obtained by the estimation do not include true parameter values. This result shows a significant limitation of applying only measurement errors in the estimation. In particular, the *exponential* noise parameterisation performs poorly and only the centroid location shows reasonable estimates. The *variance* parameterisation performs better, but marginals of the location parameters exhibit significant bias, while some MT components are resolved (e.g., m_{ee}, m_{ne}).

Including the $\mathbf{C}_k^t$ term leads to increased uncertainty, but more importantly, both noise parameterisation types (*variance_cov* and *exponential_cov*) resolve all MT parameters (Fig. 4).

However, centroid marginals are significantly wider than those observed for other noise parameterisations. In addition, the true value of north-shift is not recovered when using *variance_cov*. The *non-Toeplitz* parameterisation resolves all parameters, although in some instances, true parameter values are in the tail of the marginals (e.g., north-shift, m_{nd}, m_{ed}). Notably, centroid time is only recovered by the *non-Toeplitz* parameterisation.

The results for case 2 (large theory errors) are summarized in Fig. 5. Here, it is clear that only using $\mathbf{C}_k^d$ causes significant errors where true parameter values are rarely recovered (*variance* and *exponential* results in Fig. 5). The marginals exhibit even stronger biases with respect to the true values. While the location parameters (east-shift, north-shift and depth) are recovered by the *exponential* parameterisation in case 1, depth is biased. The MT components are not recovered in either case.

Including $\mathbf{C}_k^t$ substantially widens marginals (*exponential_cov* and *variance_cov* results in Fig. 5). Only some of the marginals include the true value for these parameterisations (e.g., m_{nn}, m_{ee}), while other marginals are biased and the true values are not recovered. In contrast, the *non-Toeplitz* parameterisation recovers true values appropriately and with low uncertainty for most parameters. The centroid time is poorly recovered for all parameterisations, but magnitude is well recovered with most parameterisations, except for the *variance*, which underestimates.

3.3 Residual Analysis

To increase confidence in the results, we analyze the statistics of the data residuals. Since we assume Gaussian-distributed residuals with some covariance matrix eq. (1), both Gaussianity and randomness of standardized residuals should be tested. Standardized residuals are obtained by scaling raw residuals with their covariance matrix. That is to say, $\hat{\mathbf{r}}_k = \mathbf{L}_k^{-1}\mathbf{r}_k$, where $\mathbf{L}_k$ is the lower triangle of the Cholesky decomposition of the total covariance matrix, $\mathbf{C}_k = \mathbf{L}_k\mathbf{L}_k^T$. If the covariance matrix that was applied in the estimation agrees well with the actual correlations, the standardized residuals are uncorrelated Gaussian distributed with unit variance. That is to say, standardized residuals should be from an uncorrelated random process, which can be assessed by considering their auto-correlations and histograms. Ideally, the auto-correlation functions should exhibit a sharp central peak with no or small side-lobes.

Histrograms should agree closely with a Gaussian probability density function (PDF) with unit variance (Dettmer et al. 2007).

Histograms of standardized residuals for cases 1 and 2 (Fig. 6, station-individual histograms Supplemental Figs. S1-S5) show that for the parameterisations of *variance* and *exponential* the assumption of Gaussianity of residuals is not met in the estimation. These distributions are more heavy-tailed and peaked than Gaussian distributions. Including, $\mathbf{C}_k^t$ vastly improves this issue and the standardized residuals are more Gaussian. In particular, peak height is reduced (i.e., reduced overfitting of data). However, the distributions exhibit extensive tails with large standard deviations. The *variance_cov* performed better than the *exponential_cov* in this case, while the *non-Toeplitz* parameterisation shows standardized residuals with satisfactory Gaussianity.

The station-individual autocorrelations show that parameterisations *variance* and *exponential* have long-wavelength sidelobes (Supplemental Figs. S6, S8). This means that residuals contain significant residual correlations that the covariance model in the estimation could not capture. Including $\mathbf{C}_k^t$ reduces the residual correlation for both parameterisations (Supplemental Figs. S7, S9). The *non-Toeplitz* covariance accounts for most correlations and standardized residuals appear close to random white noise (Supplemental Fig. S10). This result suggests that non-Toeplitz covariance matrices produce results that are most consistent with the assumptions made in the estimation and, from the tested parameterisations they can best address problems with significant theory error .

The results when *non-Toeplitz* covariance matrices have been applied in the estimation and can best address problems with significant theory error.

3.4 Moment tensor decompositions

To evaluate the focal-mechanism representation of the sampled MT components, MTs can be decomposed into isotropic and deviatoric (DV) source components (Jost & Herrmann 1989). The DV component can be split further into the compensated linear vector dipole (CLVD) and double-couple (DC) components. We applied this MT decomposition to the results of both cases for each noise parameterisation. In general, the different percentages of the decomposed source components vary between different noise parameterisations.

For case 1, the differences are noticeable, e.g., *variance* and *exponential* show isotropic components between ∼5 and ∼10 %, respectively. Significant CLVD components of up to ∼20 and ∼25 % were estimated by using the *exponential* and *exponential_cov* noise parameterisations, respectively (Fig. 7a). For case 2, *exponential* and *exponential_cov* show noticable isotropic components, while the CLVD component of the *variance_cov*, *exponential* and *exponential_cov* noise parameterisations is significant (Fig. 7b).

Since the target source was a pure DC MT, it is obvious that theory errors cause erroneous CLVD and isotropic MT components if the noise parameterisation of the covariance matrix is inappropriate. In this regard, the *non-Toeplitz* noise parameterisation outperformed all the other parameterisations with overall the smallest errors in estimating isotropic and CLVD components for both cases. It is worth noting that the *variance* noise parameterisation is the second best.

3.5 Deviatoric and Double-Couple Moment Tensors

Sometimes, MT are estimated under the assumption of a DV or a DC model for earthquakes. Such assumptions remove the possibility of estimating isotropic or CLVD components that may be considered unphysical for earthquakes. Consequently, the estimation may be more successful as long as this assumption is consistent with the actual rupture mechanism.

For the DV case with small theory error, it is noteworthy that most parameterisations, except for *exponential*, estimated the MT components well, and most, except for *variance*, recovered the source location (Supplemental Fig. S11). The centroid time was not recovered by any of the parameterisations. However, most parameterisations, except for *exponential*, resolved magnitude, but in the tail of the distributions.

For DV with large theory errors, *non-Toeplitz* was the only parameterisation that could recover source location and MT components. Including $\mathbf{C}_k^t$ permitted recovery of source location, but *variance* and *exponential* poorly estimated MT components and centroid (Supplemental Fig. S12). In this case, centroid time and magnitude were similar to the small-error case. Decomposing the DV MT for the *exponential* parameterisation showed a large CLVD component for cases 1 and 2. The CLVD component was less when including $\mathbf{C}_k^t$ (Supplemental Fig. S13). Supplemental Figure S14 and Tab. 3 present results for assuming a DC model. For case 1, *variance* and *exponential* parameterisations do not recover true values (Supplemental Fig. S14). With $\mathbf{C}_k^t$, true parameters are resolved, but location and time parameters are not estimated well. While parameters are not fully recovered by the *exponential* parameterisation, there is a vast improvement when including $\mathbf{C}_k^t$ (e.g., rake, time, depth, magnitude). Only the *non-Toeplitz* parameterisation resolved the true source mechanism, magnitude and centroid location well. The true centroid time was recovered only by the *exponential_cov* parameterisation, but it resolved several other parameters poorly.

For large theory errors the source mechanism and location could only be recovered by the *non-Toeplitz* parameterisation (Fig. 8). Including $\mathbf{C}_k^t$ did not help to reliably recover the true parameter values. Only the source magnitude was recovered by most parameterisations, except for the *variance* parameterisation.

Our results show that under the assumption of a DC MT, source parameters can be biased if correlated, non-stationary data errors are ignored in the noise parameterisation of the covariance matrix. Similar to the results for the full MT, for small theory errors, including $\mathbf{C}_k^t$ improved source parameter estimates. For large theory errors, only the *non-Toeplitz* parameterisation resolved the true source parameters successfully.

4 APPLICATION TO FOX CREEK EARTHQUAKE

In this section we apply the various approaches to theory-error estimation to a regional earthquake. Regional seismic data are considered for the M_l=4.4 earthquake occurring on 13 June 2015 near Fox Creek, Alberta, Canada (Wang et al. 2016) (Fig. 9). The event is related to hydraulic fracturing operations in this area, which was previously seismically relatively inactive (Schultz et al. 2015). Thus, the possibility of sizable non DC source components due to fluid effects could be expected, and hence it is justified to do a full MT estimation.

We use data from stations at epicentral distances of up to 300 km, based on the location from the NEIC catalog (54.102° N and 116.95° W). We convert the data to displacement waveforms, downsample them to 1.0 Hz and rotate them to radial (R), transverse (T) and vertical (Z) components. We then estimate parameters (location, MT components, centroid time) of a full MT using body waves (band-pass filtered to 0.08-0.3Hz on the Z component) and surface waves (band-pass filtered to 0.04-0.1Hz on the T component) for each noise parameterisation

(Tab. 2). With such station configuration and filter settings, we try to resemble the setup of Wang et al. (2016) for comparison, although data of some stations are not publicly available.

To test our method, we use two reference subsurface structures, a regional structure (Wang et al. 2016) and the global AK135 earth structure (Kennett et al. 1995) (Supplemental Fig. S15). Following our procedure from Sec. 2.2, we vary these reference structures 20 times each with standard deviations of 15% and 35% for velocity and layer depth values for the regional structure and 15% and 10% for the global structure (Supplemental Fig. S15). The GFs are computed with QSEIS (Wang 1999) with 1 Hz sampling on a grid with 200-m and 1000-m spacing for depths from 0 to 8 km and distances from 0 to 400 km, respectively.

4.1 Results

For the regional subsurface structure, estimation results are summarized in Fig. 10 in terms of marginal probability densities. It is most striking that *variance*, *exponential* and *non-Toeplitz* parameterisation show similar results across all parameters. This observation implies that it is not necessary to account for non-stationary correlated noise and that the theory error is small. Including $\mathbf{C}_k^t$ into estimation significantly widens the marginals and results in shifts of the marginals (e.g., magnitude, depth, m_{ne}). By artificially introducing theory error through $\mathbf{C}_k^t$, the *variance_cov* and *exponential_cov* marginals resemble uncertainty, which in reality may not be significant, and correspondingly we likely overestimated the errors in the regional structure (Supplemental Fig. S15a). In this case, the results become worse since the subsurface structure appears to be well known.

For the global subsurface structure, estimation results of *variance* and *exponential* parameterisations show higher magnitude estimates, earlier centroid times as well as shallower source depth (Fig. 11). Results become more consistent including $\mathbf{C}_k^t$ and *variance_cov* and *exponential_cov* marginals mostly contain the *non-Toeplitz* marginals. The *exponential_cov* and *variance_cov* parameterisations lose the source depth resolution. This indicates that the global structure contains significant theory error for data of the study area and accounting for it through $\mathbf{C}_k^t$ better characterises uncertainties.

We note that published solutions (e.g., Wang et al. 2016) are close to the marginals for *variance*, *exponential* and *non-Toeplitz* when employing the regional velocity model. The wider (more uncertain) marginals for *variance_cov* and *exponential_cov* also include the published solutions when employing a global velocity model. However, published solutions fix centroid location after an initial grid-search. Therefore, trade-offs between location and MT components are not investigated.

The fit to the transverse data (surface waves) is better (weighted variance reduction of 75% to 99.5%) than for Z components (body and surface waves, -400% to 40%). The difference is likely due to the lower frequency content for transverse data (Fig. 12). Including $\mathbf{C}_k^t$ predominantly leads to larger variations in amplitude of predicted waveforms for the higher-frequency Z components (Supplemental Fig. S16). Note that $\mathbf{C}_k^t$ results depend on reasonable assumptions for velocity uncertainty. Expectedly, data fits are better with regional velocity models, rather than global models (Supplemental Fig. S17).

Particularly interesting is how noise parameterisations affect variance reductions at stations with high noise. Accounting for correlated noise results in significantly higher weighted variance reductions compared to when ignoring correlations (Supplemental Figs. S16-S24). For example, TD.TD010.Z has a weighted variance reduction of -160 to -10% for *variance* compared to 20 to 40% for *non-Toeplitz* (Supplemental Figs. S17, S24).

To better visualize and interpret MT PPDs, we apply MT decomposition (also see Sec. 3.4) (Fig. 13). Notably, poor noise parameterisation choices lead to erroneous, large isotropic components (e.g., *variance_cov* and *exponential_cov* for the regional model, and *variance* and *exponential* for the global model). This inherent compensation of theory errors by biasing source parameters is well known and caused by the fundamental trade-offs between source parameterisation and Earth-structure (e.g., Valentine & Woodhouse 2010; Hejrani & Tkalčić 2020).

Based on sensitivity analysis, Wang et al. (2016) report a CLVD component of $\sim 23 \pm 17\%$ which is lower and more uncertain than our estimates obtained with the regional velocity model. With the global velocity model, the CLVD component is poorly constrained. We infer a nearly vertically dipping fault, striking N-S or E-W. In this case, the CLVD component may be due to fault complexity where the rupture is not occurring on a single planar fault but may include multiple segments that are offset in the vertical plane. Such complex faulting can occur in the presence of basement flower structures (Eaton et al. 2018).

5 CONCLUSIONS

We investigated the influence of noise parameterisation on estimates of CMTs in the presence of theory errors due to a mismatch between the Earth and a velocity model employed for GF computation. In particular, we compare five approaches to noise covariance estimation to account for these theory errors. For the comparison, Bayesian inference was applied to estimate CMT solutions and noise parameters for synthetic and field seismic data at regional distances. Several of these approaches were previously applied to MT inversion but are considered here for the nonlinear case with unknown centroids. In addition, we adopt the non-parametric iterative approach of estimating *non-Toeplitz* matrices from another field and demonstrate that it has significant advantages in situations of practical importance.

We demonstrated results for regional simulations with distances $< 1000\,km$ and for a field example with distances $< 300\,km$. The GFs from the synthetic test with 8 km source depth are mostly sensitive to mid and lower crust. Thus, if the centroid was located shallower, a wider bandpass filter to higher frequencies above 0.1 Hz would be required to resolve the source parameters (Hejrani & Tkalčić 2020). However, the BEAT software (Vasyura-Bathke et al. 2019, 2020) employs a general Bayesian framework for uncertainty quantification with extensive choices for noise models and has been successfully applied to data with arbitrary frequency content, centroid location, station distances, and source parameterisations. Users are free to change BEAT or apply as is for local, regional and global data using DC, DV, or full CMTs, rectangular faults, and multi-segment finite faults.

The five approaches we studied either ignore covariances or include them in the inverse problem. The *variance* approach ignores covariances and estimates the noise standard deviation as part of the inversion, i.e., a hierarchical noise model (Malinverno & Briggs 2004). The approaches with covariances include: (1) A hierarchical model with a simple function to describe off-diagonal terms (*exponential* parameterisation), (2) explicit modelling approaches (*variance_cov* and *exponential_cov* parameterisations), and (3) an empirical non-parametric approach (*non-Toeplitz*).

The hierarchical models were mostly inadequate to address theory errors. The explicit modelling approaches compute $\mathbf{C}_k^t$ to include the effects of theory errors. The computation is expensive, although more efficient computation may be possible (Hallo & Gallovic 2016). In addition, the computation requires specification of velocity-model uncertainties which are generally assumed, since no such knowledge is independently available. Including $\mathbf{C}_k^t$ improved results significantly over the simple hierarchical models. However, the dependence on specifying adequate velocity-model uncertainties is a significant disadvantage and led to some erroneous results. Also the inferred uncertainties depend mostly on the specified velocity-model uncertainty and are therefore subjective (Fox Creek). Note that both velocity and layer-depth errors need to be chosen which can pose a non trivial task. In particular, if the true velocity model is not included in the variations specified *a priori*, this approach leads to poor parameter estimates (case 2).

The *non-Toeplitz* parameterisation performed best overall. The formulation is non-parametric and therefore fast to compute during sampling. Importantly, it intrinsically accounts theory and measurement errors and does not differentiate between theory-error sources, by including but not limited to errors due to Earth-structure mismatch and centroid-location mismatch. Even when significant theory error exists (case 2), the covariance estimation procedure based on data residuals produced robust parameter and uncertainty estimates. A disadvantage is the iterative nature and that it may require the initial assumption of uncorrelated noise of unknown standard deviation. In conclusion, our results suggest that applying the *non-Toeplitz* covariance matrix parameterisation provides a reliable and, straightforward approach to account for correlated errors due to theory error in source parameter estimation.

6 ACKNOWLEDGMENTS

Seismic waveforms were originally downloaded from Incorporated Research Institutions for Seismology (IRIS) https://ds.iris.edu. We thank Babak Hejrani, Simon Stähler and the editor Andrew Valentine for their comments that helped to improve the quality of this manuscript. Special thanks go to Sebastian Heimann, Frank Krüger, Mehdi Nikkhoo and Olaf Zielke for valuable discussions. Plots were produced with *Matplotlib* and the Generic Mapping Tools (Hunter 2007; Wessel et al. 2013). This work employed the open source library *py-*

rocko (www.pyrocko.org) (Heimann et al. 2017) and the Bayesian Earthquake Analysis Tool (https://github.com/hvasbath/beat). The research was supported by King Abdullah University of Science and Technology (KAUST), under award numbers BAS/1/1353-01-01 and BAS/1/1339-01-1. H.V-B was partially supported by Geo.X, the Research Network for Geosciences in Berlin and Potsdam under the project number SO_087_GeoX.

REFERENCES

Aki, K. & Richards, P. G., 2002. *Quantitative seismology*, University Science Books, 2nd edn.

Baer, G., Funning, G. J., Shamir, G., & Wright, T. J., 2008. The 1995 November 22, M w 7.2 Gulf of Elat earthquake cycle revisited, *Geophys, J. Int.*, **175**(3), 1040–1054.

Bassin, C., Laske, G., & Masters, G., 2000. The current limits of resolution for surface wave tomography in North America, *EOS Trans. AGU*, **81**(F897).

Bathke, H., Sudhaus, H., Holohan, E., Walter, T. R., & Shirzaei, M., 2013. An active ring fault detected at Tendürek volcano by using InSAR, *J. Geophys. Res. Solid Earth*, **118**(8), 4488–4502.

Bayes, T., 1763. An Essay towards Solving a Problem in the Doctrine of Chances, *Phil. Trans.*, **53**, 370–418.

Bonnefoy-Claudet, S., Cotton, F., & Bard, P. Y., 2006. The nature of noise wavefield and its applications for site effects studies. A literature review, *Earth Sci. Rev.*, **79**(3-4), 205–227.

Cesca, S., Grigoli, F., Heimann, S., Dahm, T., Kriegerowski, M., Sobiesiak, M., Tassara, C., Olcay, M., Potsdam, G., Potsdam, D., & E-mail, G., 2016. The Mw 8.1 2014 Iquique, Chile, seismic sequence: a tale of foreshocks and aftershocks, *Geophys, J. Int.*, **204**, 1766–1780.

Cesca, S., Heimann, S., Kriegerowski, M., Saul, J., & Dahm, T., 2017. Moment Tensor Inversion for Nuclear Explosions : What Can We Learn from the 6 January and 9 September 2016 Nuclear Tests , North Korea ?, *Seismol. Res. Lett.*, **88**(2A).

Ching, J. & Chen, Y.-C., 2007. Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, *J. Eng. Mech.*, **133**(7), 816–832.

Dębski, W., 2008. Estimating the earthquake source time function by Markov Chain Monte Carlo sampling, *Pure appl. geophys.*, **165**(7), 1263–1287.

Dettmer, J., Dosso, S. E., & Holland, C. W., 2007. Uncertainty estimation in seismo-acoustic reflection travel time inversion, *J. Acoust. Soc. Am.*, **122**(1), 161–176.

Dettmer, J., Benavente, R., Cummins, P. R., & Sambridge, M., 2014. Trans-dimensional finite-fault inversion, *Geophys, J. Int.*, **199**(2), 735–751.

Du, Y., Segall, P., & Gao, H., 1994. Dislocations in inhomogeneous media via a moduli perturbation approach : General formulation and two-dimensional solutions, *J. Geophys. Res.*, **99**, 13767–13779.

Duputel, Z., Rivera, L., Fukahata, Y., & Kanamori, H., 2012. Uncertainty estimations for seismic source inversions, *Geophys, J. Int.*, **190**(2), 1243–1256.

Duputel, Z., Agram, P. S., Simons, M., Minson, S. E., & Beck, J. L., 2014. Accounting for prediction uncertainty when inferring subsurface fault slip, *Geophys, J. Int.*, **197**(1), 464–482.

Dutta, R., Jónsson, S., Wang, T., & Vasyura-Bathke, H., 2018. Bayesian estimation of source parameters and associated Coulomb failure stress changes for the 2005 Fukuoka (Japan) earthquake, *Geophys, J. Int.*, **213**, 261–277.

Eaton, D. W., Igonin, N., Poulin, A., Weir, R., Zhang, H., Pellegrino, S., & Rodriguez, G., 2018. Induced seismicity characterization during hydraulic-fracture monitoring with a shallow-wellbore geophone array and broadband sensors, *Seismol. Res. Lett.*, **89**(5), 1641–1651.

Ekström, G., 2006. Global detection and location of seismic sources by using surface waves, *Bull. Seismol. Soc. Am.*, **96**(4 A), 1201–1212.

Ekström, G., Nettles, M., & Dziewoński, A. M., 2012. The global CMT project 2004-2010: Centroid-moment tensors for 13,017 earthquakes, *Phys. Earth Planet. Inter.*, **200-201**, 1–9.

Fukuda, J. & Johnson, K. M., 2008. A fully Bayesian inversion for spatial distribution of fault slip with objective smoothing, *Bull. Seismol. Soc. Am.*, **98**(3), 1128–1146.

Goldstein, H., Poole, C. P., & Safko, J. L., 2001. *Classical Mechanics*, Addison Wesley.

Hallo, M. & Gallovic, F., 2016. Fast and cheap approximation of Green function uncertainty for waveform-based earthquake source inversions, *Geophys, J. Int.*, **207**, 1012–1029.

Hanks, T. C. & Kanamori, H., 1979. A moment magnitude scale, *J. Geophys. Res. Solid Earth*, **84**(B5), 2348–2350.

Heimann, S., Kriegerowski, M., Isken, M., Cesca, S., Daout, S., Grigoli, F., Juretzek, C., Megies, T., Nooshiri, N., Steinberg, A., Sudhaus, H., Vasyura-Bathke, H., Willey, T., & Dahm, T., 2017. Pyrocko - An open-source seismology toolbox and library, *GFZ Data Services*, **v. 0.3**.

Heimann, S., Vasyura-Bathke, H., Sudhaus, H., Isken, M. P., Kriegerowski, M., Steinberg, A., & Dahm, T., 2019. A Python framework for efficient use of pre-computed Green's functions in seismological and other physical forward and inverse source problems, *Solid Earth*, **10**(6), 1921–1935.

Hejrani, B. & Tkalčić, H., 2020. Resolvability of the Centroid-Moment-Tensors for Shallow Seismic Sources and Improvements From Modeling High-Frequency Waveforms, *J. Geophys. Res. Solid*

Earth, **125**(7), 1–31.

Hofstetter, A., Thio, H. K., & Shamir, G., 2003. Source mechanism of the 22 / 11 / 1995 Gulf of Aqaba earthquake and its aftershock sequence, *J. Seismolog.*, **7**, 99–114.

Hunter, J. D., 2007. Matplotlib : A 2D graphics environment, *Comput. Sci. Eng.*, **9**(3), 90–95.

Jost, M. L. & Herrmann, R. B., 1989. A Student's Guide to and Review of Moment Tensors, *Seismol. Res. Lett.*, **60**(2), 37–57.

Kennett, B. L. N., Engdahl, E. R., & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys, J. Int.*, **122**, 108–124.

Koch, K., 1991. Moment tensor inversion of local earthquake dataII. Application to aftershocks of the May 1980 Mammoth Lakes earthquakes, *Geophys, J. Int.*, **106**(2), 321–332.

Kravanja, S., Panza, G. F., & Šílený, J., 1999. Robust retrieval of a seismic point-source time function, *Geophys, J. Int.*, **136**(2), 385–394.

Malinverno, A. & Briggs, V. A., 2004. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes, *Geophysics*, **69**(4), 1005–1016.

Minson, S. E., Simons, M., & Beck, J. L., 2013. Bayesian inversion for finite fault earthquake source models I-theory and algorithm, *Geophys, J. Int.*, **194**(3), 1701–1726.

Monelli, D. & Mai, P. M., 2008. Bayesian inference of kinematic earthquake rupture parameters through fitting of strong motion data, *Geophys, J. Int.*, **173**(1), 220–232.

Monelli, D., Mai, P. M., Jónsson, S., & Giardini, D., 2009. Bayesian imaging of the 2000 Western Tottori (Japan) earthquake through fitting of strong motion and GPS data, *Geophysical Journal International*, **176**(1), 135–150.

Mooney, W. D., 1989. Seismic methods for determining parameters and lithospheric structure earthquake source, in *Geophysical framework of the United States: Boulder, Colorado*, pp. 71–110, eds Pakiser, L. C. & Mooney, W. D., Geological Society of America Memoir 172.

Moral, P. D., Doucet, A., & Jasra, A., 2006. Sequential Monte Carlo samplers, *J.R. Statist. Soc. B.*, **68**(3), 411–436.

Pugh, D. J., White, R. S., & Christie, P. A., 2016. A Bayesian method for microseismic source inversion, *Geophys, J. Int.*, **206**(2), 1009–1038.

Ragon, T., Sladen, A., & Simons, M., 2018. Accounting for uncertain fault geometry in earthquake source inversions - I: Theory and simplified application, *Geophys, J. Int.*, **214**(2), 1174–1190.

Razafindrakoto, H. N. T. & Mai, M. P., 2014. Uncertainty in earthquake source imaging due to variations in source time function and earth structure, *Bull. Seismol. Soc. Am.*, **104**(2), 855–874.

Sambridge, M. & Mosegaard, K., 2002. Monte Carlo Methods in Geophysical Inverse Problems, *Rev.*

Geophys., **40**(3), 1009.

Schultz, R., Stern, V., Novakovic, M., Atkinson, G., & Gu, Y. J., 2015. Hydraulic fracturing and the Crooked Lake Sequences: Insights gleaned from regional seismic networks, *Geophys. Res. Lett.*, **42**(8), 2750–2758.

Sigloch, K. & Nolet, G., 2006. Measuring finite-frequency body-wave amplitudes and traveltimes, *Geophys, J. Int.*, **167**(1), 271–287.

Šílenỳ, J., Panza, G. F., & Campus, P., 1992. Waveform inversion for point source moment tensor retrieval with variable hypocentral depth and structural model, *Geophys, J. Int.*, **109**(2), 259–274.

Silver, P. G. & Jordan, T. H., 1982. Optimal estimation of scalar seismic moment, *Geophys. J. R. astr. Soc.*, **70**(3), 755–787.

Sipkin, S. A., 1982. Estimation of earthquake source parameters by the inversion of waveform data: synthetic waveforms, *Phys. Earth Planet. Inter.*, **30**(2-3), 242–259.

Stähler, S. C. & Sigloch, K., 2014. Fully probabilistic seismic source inversion Part 1 : Efficient parameterisation, *Solid Earth*, **5**, 1055–1069.

Stähler, S. C. & Sigloch, K., 2016. Fully probabilistic seismic source inversion Part 2 : Modelling errors and station covariances, *Solid Earth*, (7), 1521–1536.

Tarantola, A., 2005. *Inverse Problem Theory and methods for model parameter estimation*, SIAM.

Tarantola, A. & Valette, B., 1982. Inverse Problems = Quest for Information, *J. Geophys.*, **50**, 159–170.

Tocheport, A., Rivera, L., & Chevrot, S., 2007. A systematic study of source time functions and moment tensors of intermediate and deep earthquakes, *J. Geophys. Res. Solid Earth*, **112**(7), 1–22.

Vackář, J., Burjánek, J., Gallovič, F., Zahradník, J., & Clinton, J., 2017. Bayesian ISOLA: New tool for automated centroid moment tensor inversion, *Geophysical Journal International*, **210**(2), 693–705.

Valentine, A. P. & Woodhouse, J. H., 2010. Reducing errors in seismic tomography: Combined inversion for sources and structure, *Geophys, J. Int.*, **109**(2), 259–274.

Vasyura-Bathke, H., Dettmer, J., Steinberg, A., Heimann, S., Isken, M., Zielke, O., Mai, P. M., Sudhaus, H., & Jónsson, S., 2019. BEAT - Bayesian Earthquake Analysis Tool, *GFZ Data Services*, **v.1.0**.

Vasyura-Bathke, H., Dettmer, J., Steinberg, A., Heimann, S., Isken, M. P., Zielke, O., Mai, P. M., Sudhaus, H., & Jónsson, S., 2020. The Bayesian Earthquake Analysis Tool, *Seismol. Res. Lett.*, **91**(2A), 1003–1018.

Wang, R., 1999. A simple orthonormalization method for stable and efficient computation of Green's functions, *Bull. Seismol. Soc. Am.*, **89**(3), 733–741.

Wang, R., Gu, Y. J., Schultz, R., Kim, A., & Atkinson, G., 2016. Source analysis of a potential hydraulic-fracturing-induced earthquake near Fox Creek, Alberta, *Geophys. Res. Lett.*, **43**(2), 564–573.

Wéber, Z., 2006. Probabilistic local waveform inversion for moment tensor and hypocentral location, *Geophys, J. Int.*, **165**(2), 607–621.

Wessel, P., Smith, W. H. F., Scharroo, R., Luis, J., & Wobbe, F., 2013. Generic Mapping Tools: Improved Version Released, *EOS Trans. AGU*, **94**(45), 409–410.

Yagi, Y. & Fukahata, Y., 2008. Importance of covariance components in inversion analyses of densely sampled observed data : an application to waveform data inversion for seismic source processes, *Geophys, J. Int.*, **175**, 215–221.

Yagi, Y. & Fukahata, Y., 2011. Introduction of uncertainty of Green's function into waveform inversion for seismic source processes, *Geophys, J. Int.*, **186**(2), 711–720.

7 TABLES

Table 1. Noise parameterisations used in this study. The data covariance matrix $\mathbf{C}_k^d$, can be estimated from waveform data at a station k before the arrival time of the event of interest.

Noise parameterisation	Covariance matrix components	Color coding	References
variance	$\mathbf{C}_k^d = \sigma^2 \mathbf{I}$	light yellow	
exponential	$\mathbf{C}_{k,ij}^d = \sigma^2 \exp -\lvert\Delta t^{ij}\rvert / t_0$	light blue	Duputel et al. (2012)
variance_cov	$\mathbf{C}_k^d + \mathbf{C}_k^t$	dark yellow	Tarantola & Valette (1982); Yagi & Fukahata (2011); Duputel et al. (2014)
exponential_cov	$\mathbf{C}_{k,ij}^d + \mathbf{C}_k^t$	dark blue	Tarantola & Valette (1982); Yagi & Fukahata (2011); Duputel et al. (2014)
non-Toeplitz	$\mathbf{C}_k$	red	Dettmer et al. (2007)

Table 2. Synthetic tests setup cases.

Setup case	velocity structures	
	reference	estimation
1.small theory error	blue	dark gray
2.large theory error	red	dark gray

Table 3. Target source parameters of the double-couple moment tensor.

Synthetic tests

moment tensor				
location	east-shift [km]	10.0		
	north-shift [km]	20.0		
	depth [km]	8.0		
strength	magnitude	4.8		
timing	centroid time [s]	-2.7		
mechanism	mnn	0.846	strike [deg]	150.0
	mee	-0.759	dip [deg]	75.0
	mdd	-0.087	rake [deg]	-10.0
	mne	0.513		
	mnd	0.146		
	med	-0.257		

8 FIGURES

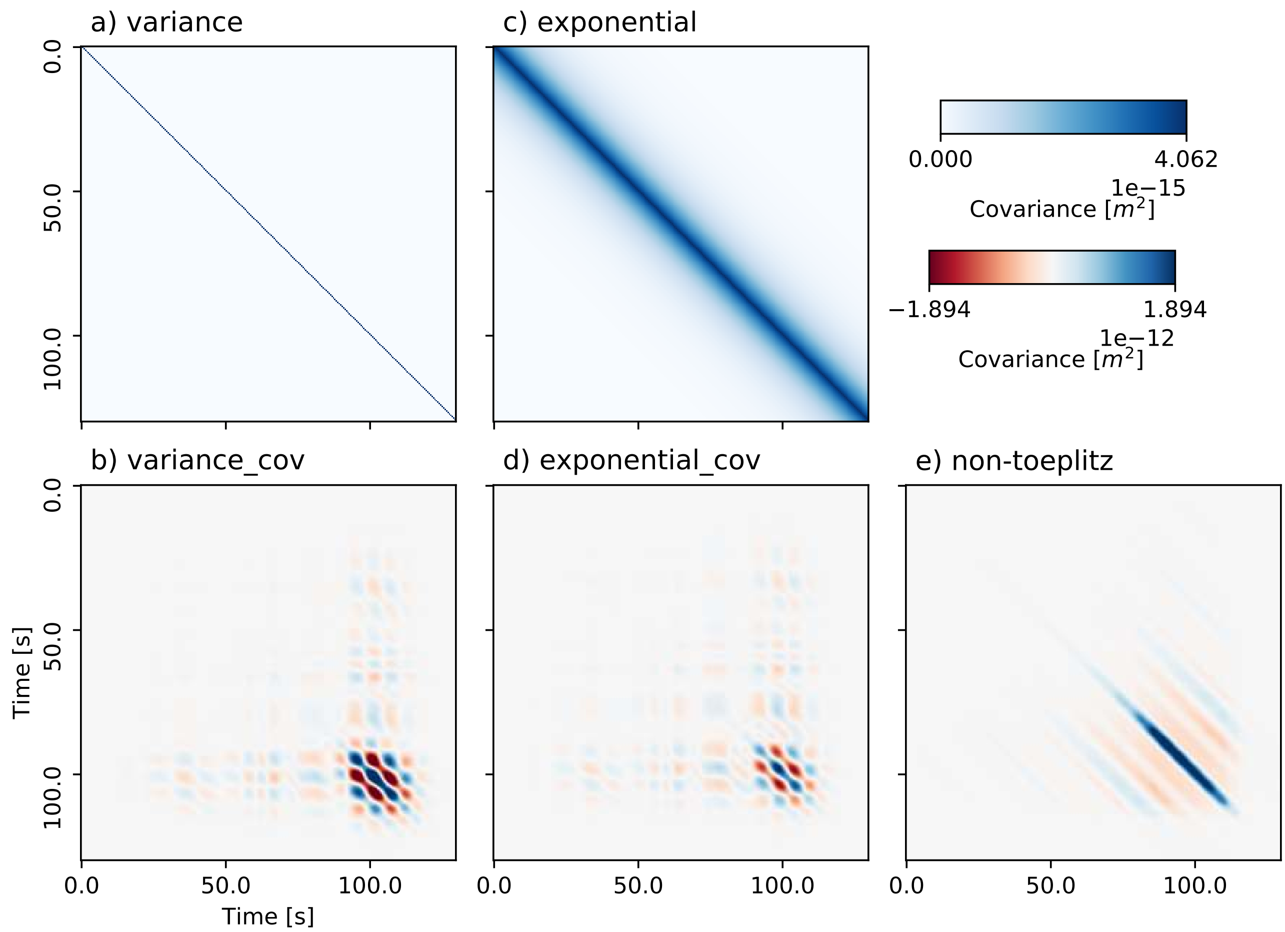

Figure 1. Covariance matrixes **C** with different noise parameterisations (Tab. 1). The parameterisations in a) and c) comprise only $\mathbf{C}_k^d$ while b), d) and e) also include $\mathbf{C}_k^t$, thus the ranges of covariance matrix values vary significantly. These covariance matrices are computed in a frequency band of 0.01 to 0.1 Hz at a sampling rate of 1Hz for a station at 167° azimuth and 350 km epicentral distance.

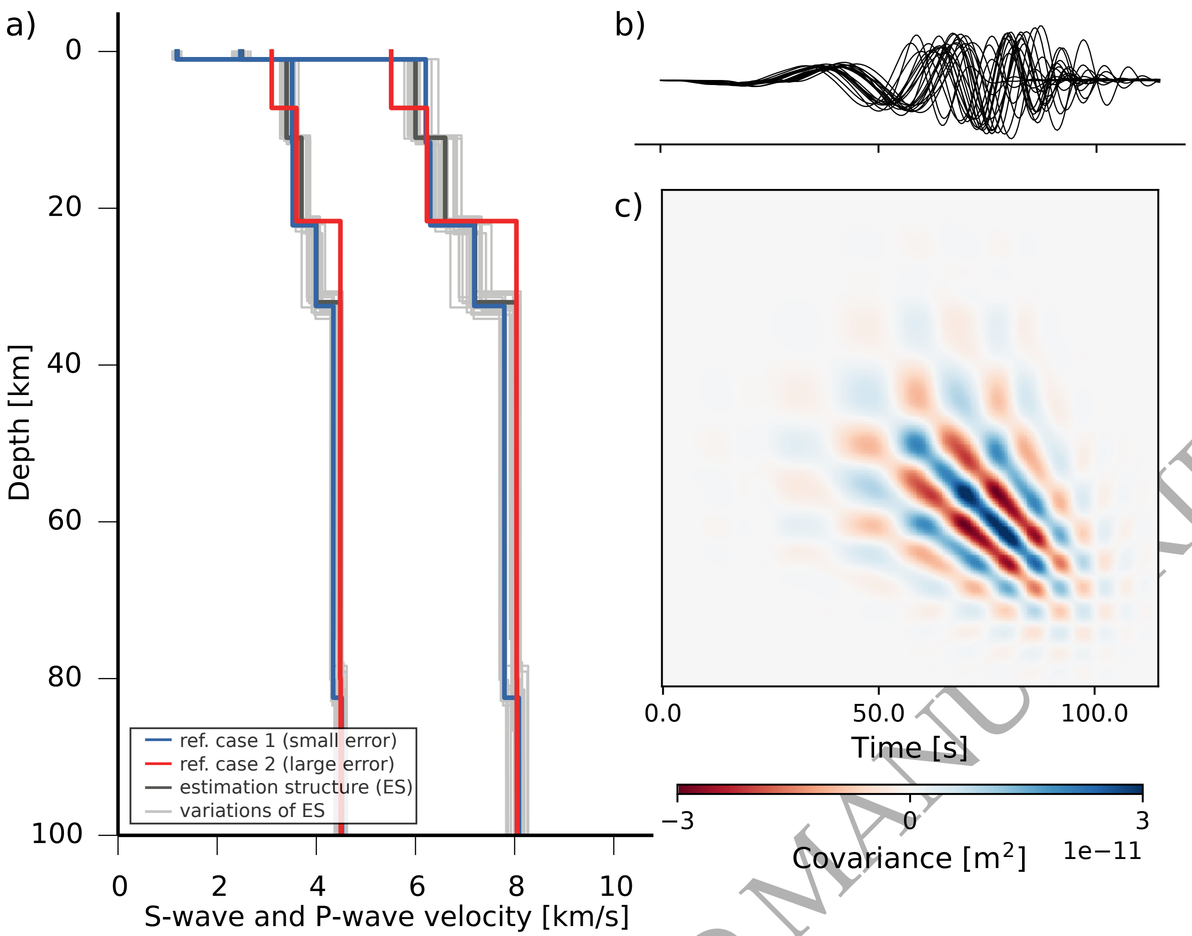

Figure 2. Steps to calculate the model prediction covariance; a) velocity model profiles: b) synthetic waveforms (vertical component) for the reference source simulated for each realization of the Earth structures; c) Covariance matrix $\mathbf{C}_k^t$ of seismic traces from b) following eq. 3.

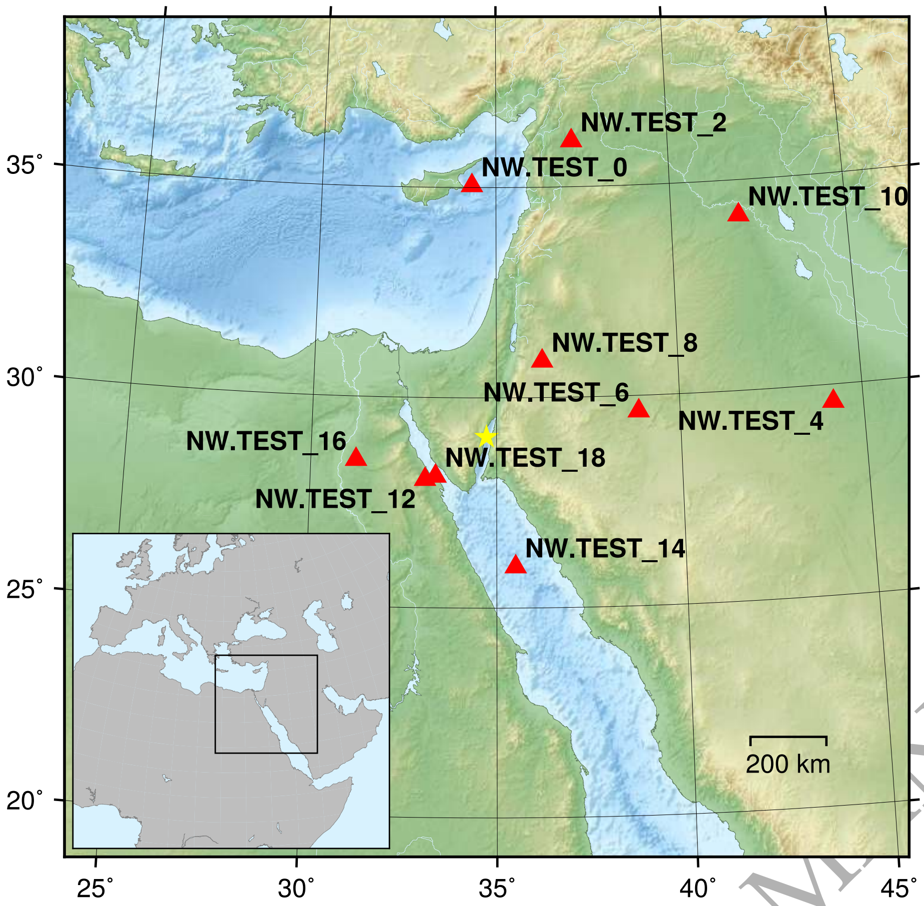

Figure 3. Stations (red triangles) used in the synthetic test that simulates a moment tensor optimization at regional distances. Station locations are randomly chosen around the reference event marked by the yellow star. The black box in the inset marks the outline of the station map.

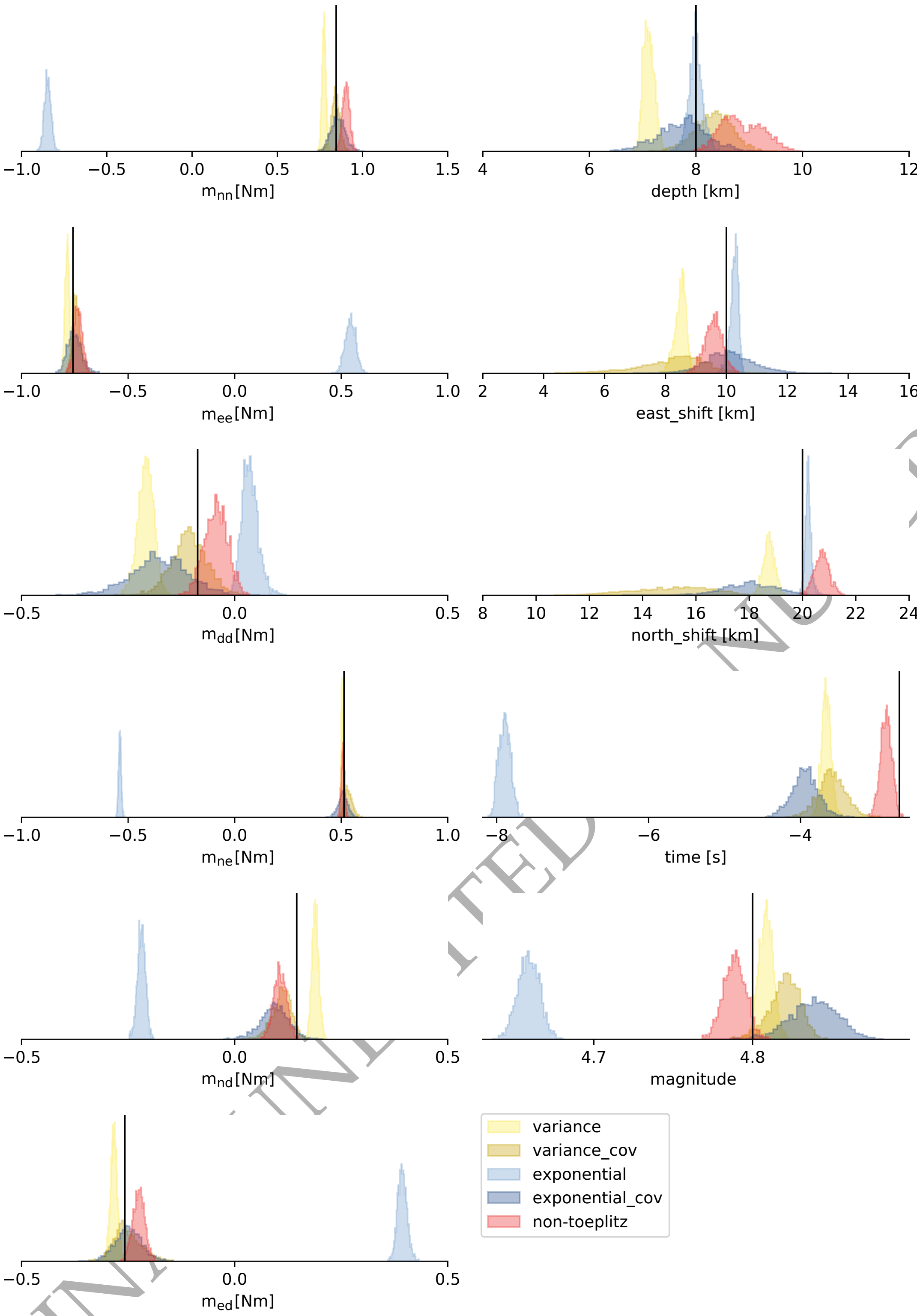

Figure 4. Case 1 with small theory error: Posterior marginal distributions for full CMT parameters. The black vertical lines mark the true input parameters. The different colors present results for different noise parameterisations (see legend and Table 2). Even small theory errors may lead to biased marginals when ignored.

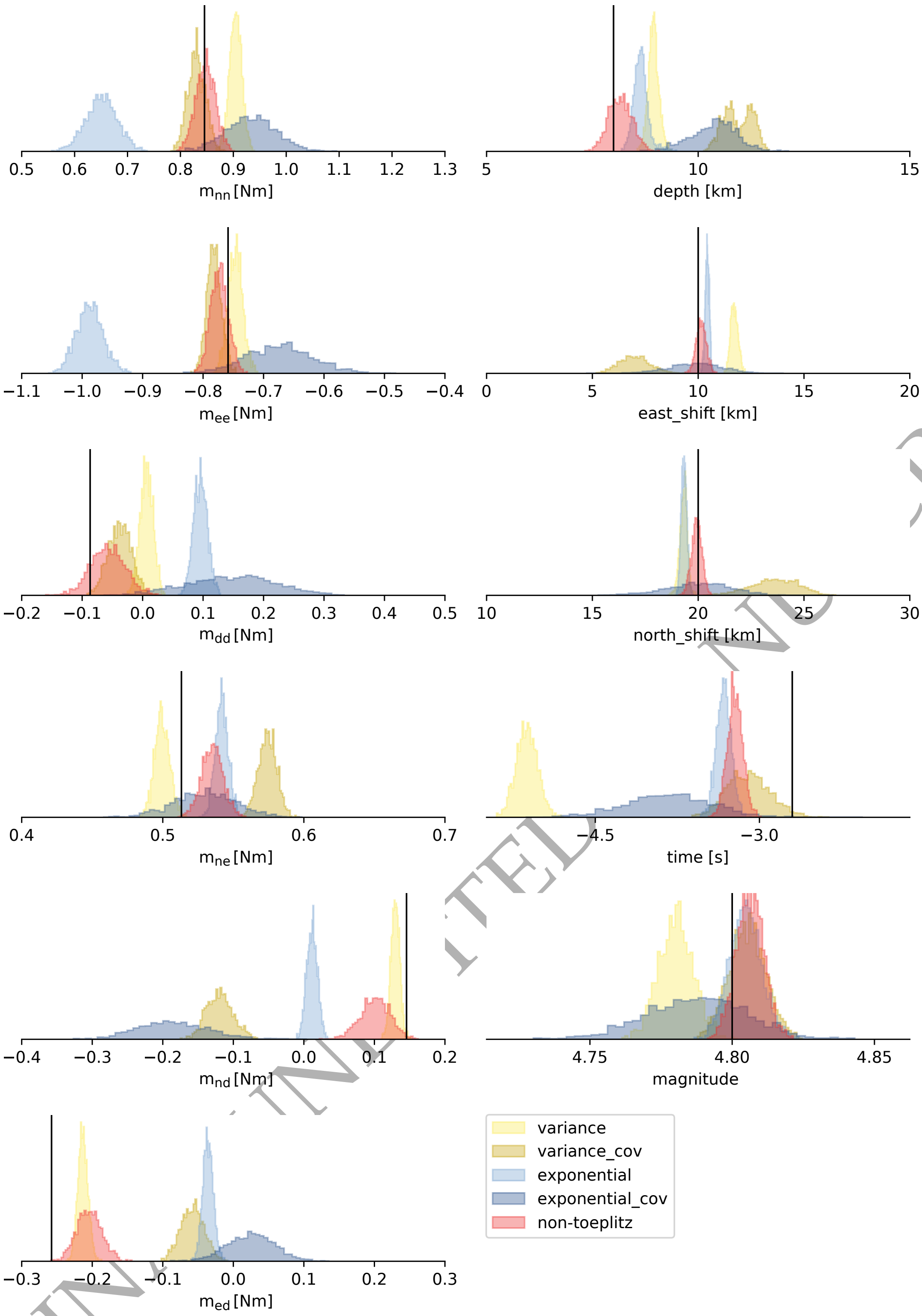

Figure 5. Case 2 with large theory error, otherwise same as Fig. 4. Only the *non-Toeplitz* parameterisation produces robust results overall. Modelling approaches suffer from the requirement to specify velocity uncertainties *a priori*.

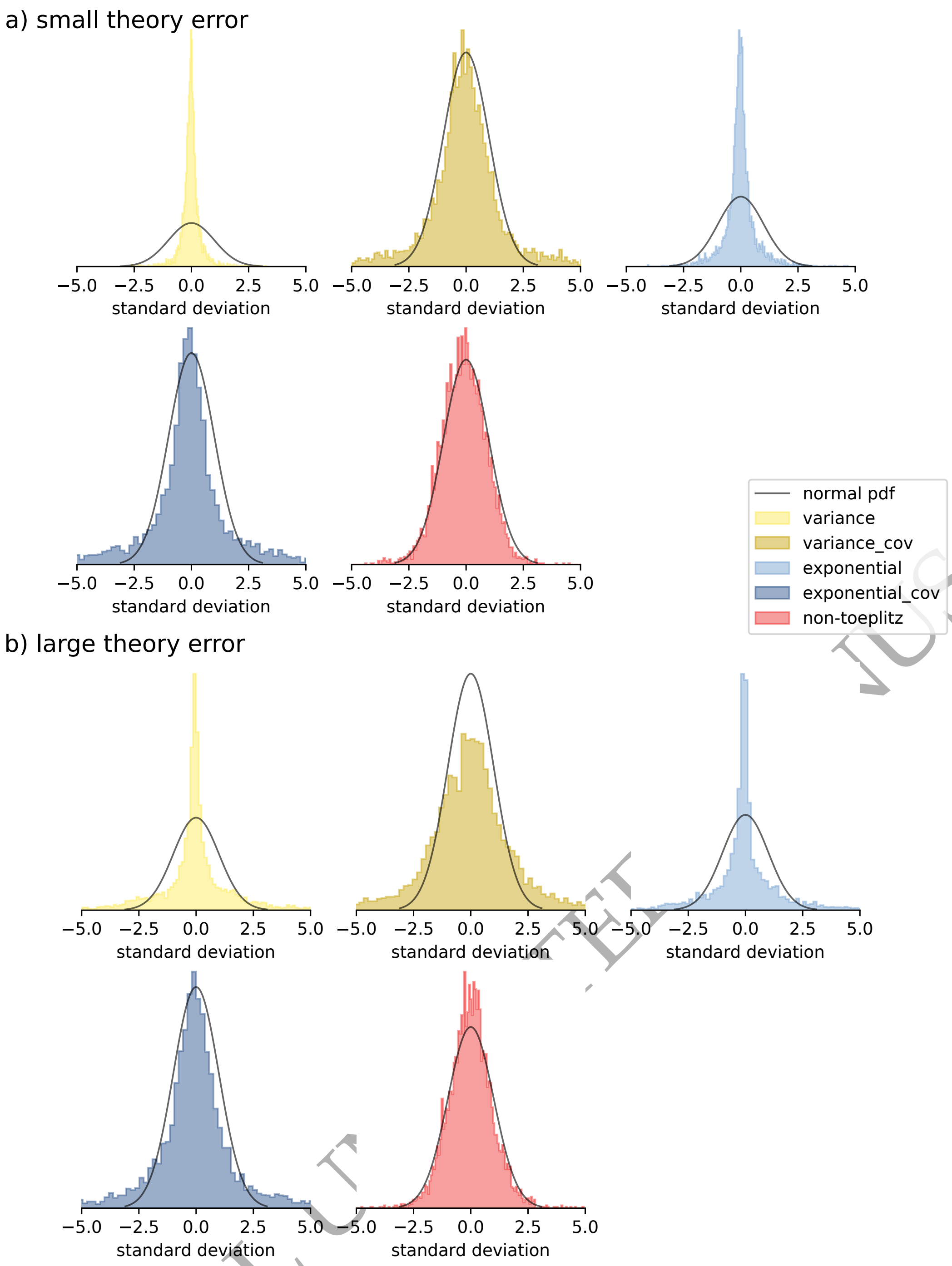

Figure 6. Standardized residuals for the different noise parameterisations for a) small theory error and b) large theory error. The black line marks the analytic normal distribution with zero mean and standard deviation of one. All histograms are normalized to unit area.

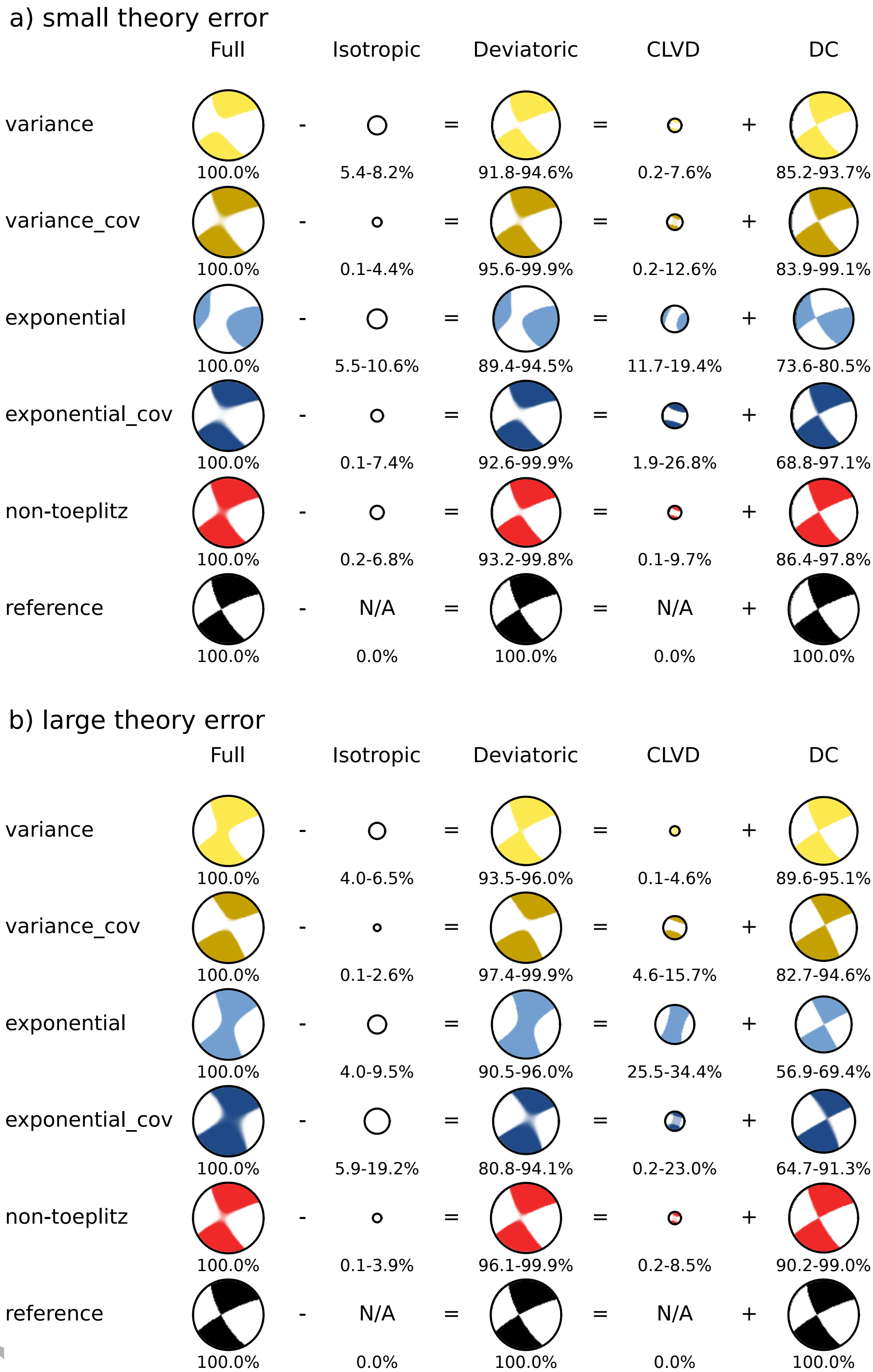

Figure 7. Moment tensor decompositions for a) case 1 with small theory error and for b) case 2 with large theory error. Each row shows the decomposition for a different noise parameterisation following the color-coding in Tab. 1 and Fig. 6. The sizes of the focal mechanisms are scaled with respect to MAP magnitudes. The numbers below each focal mechanism depict the percentage of scalar seismic moment.

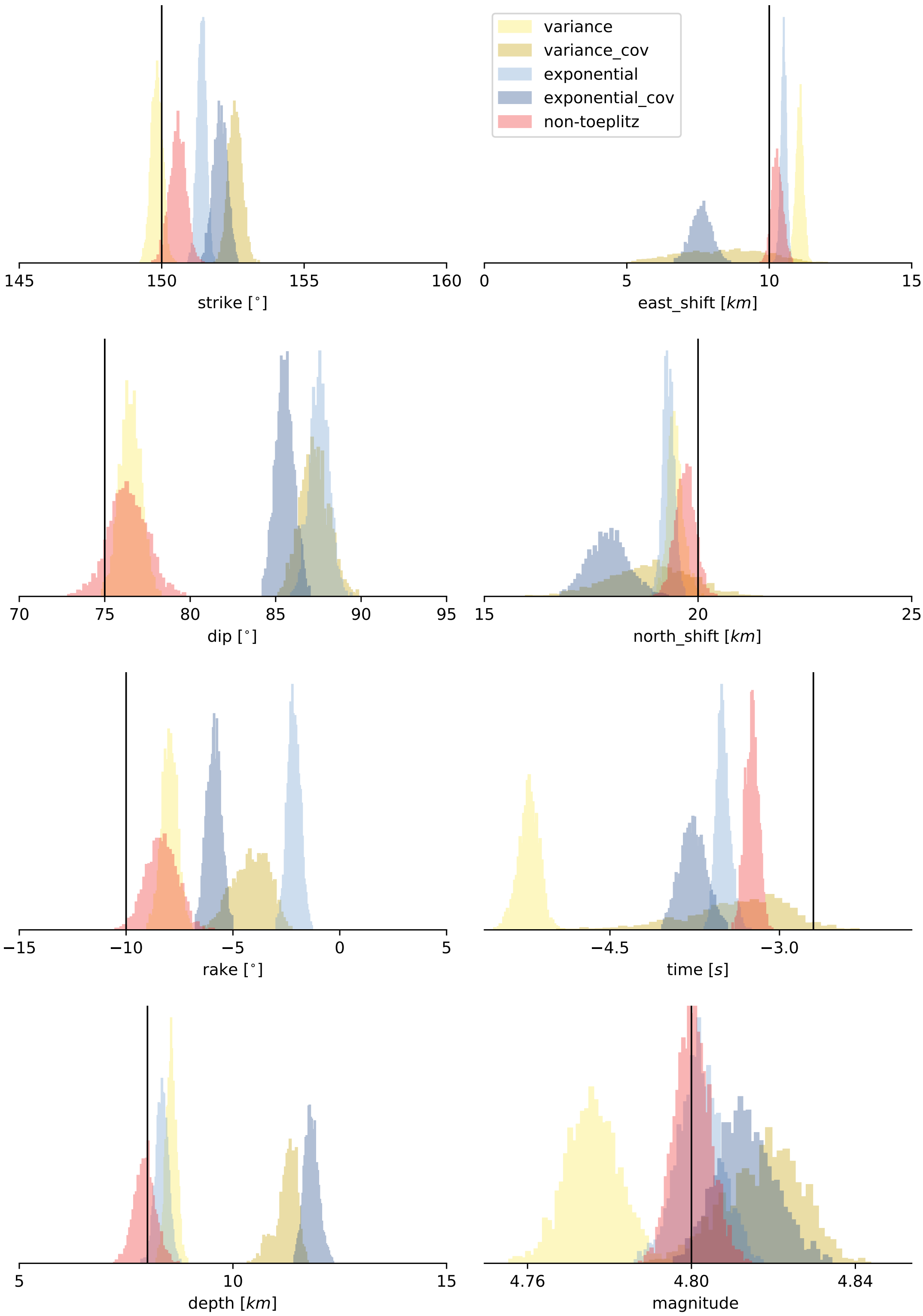

Figure 8. Posterior marginals for double-couple CMT results with large theory errors, otherwise same as Fig. 4. The *non-Toeplitz* parameterisation performed best, followed by*variance*.

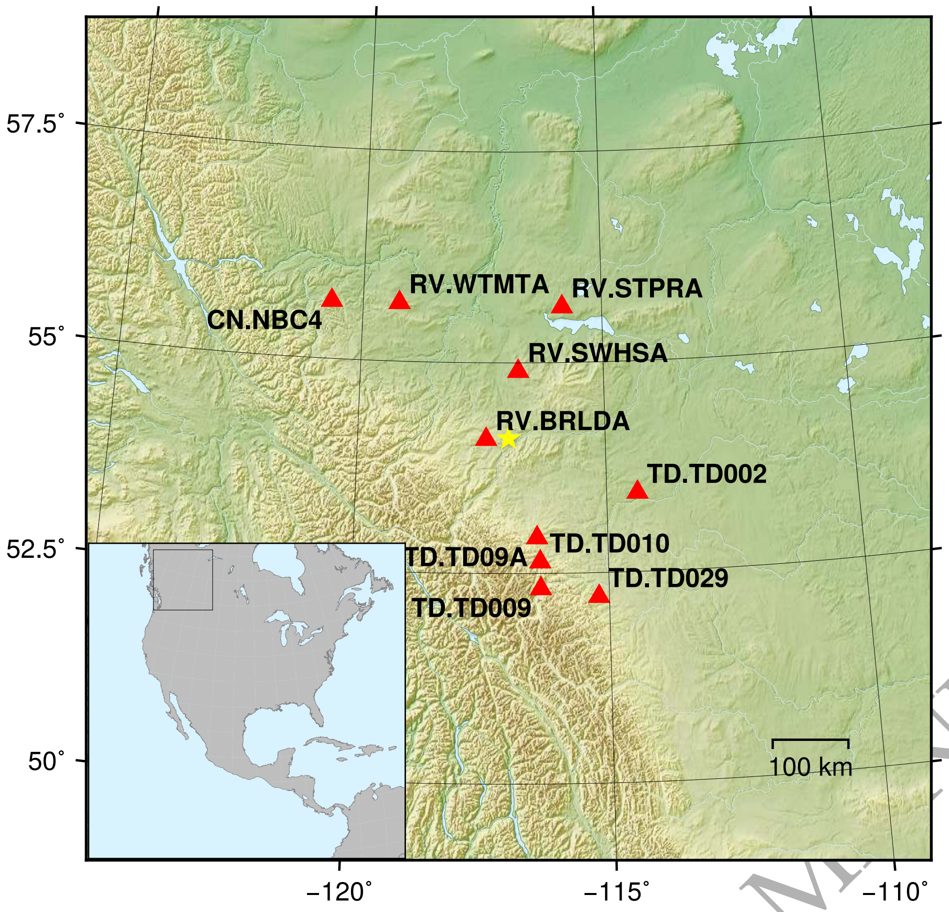

Figure 9. Stations (red triangles) used in the full moment tensor estimation at regional distances for the 13th June 2015 Fox Creek event (yellow star at 54.102° N and 116.95° W). The black box in the inset marks the outline of the station map.

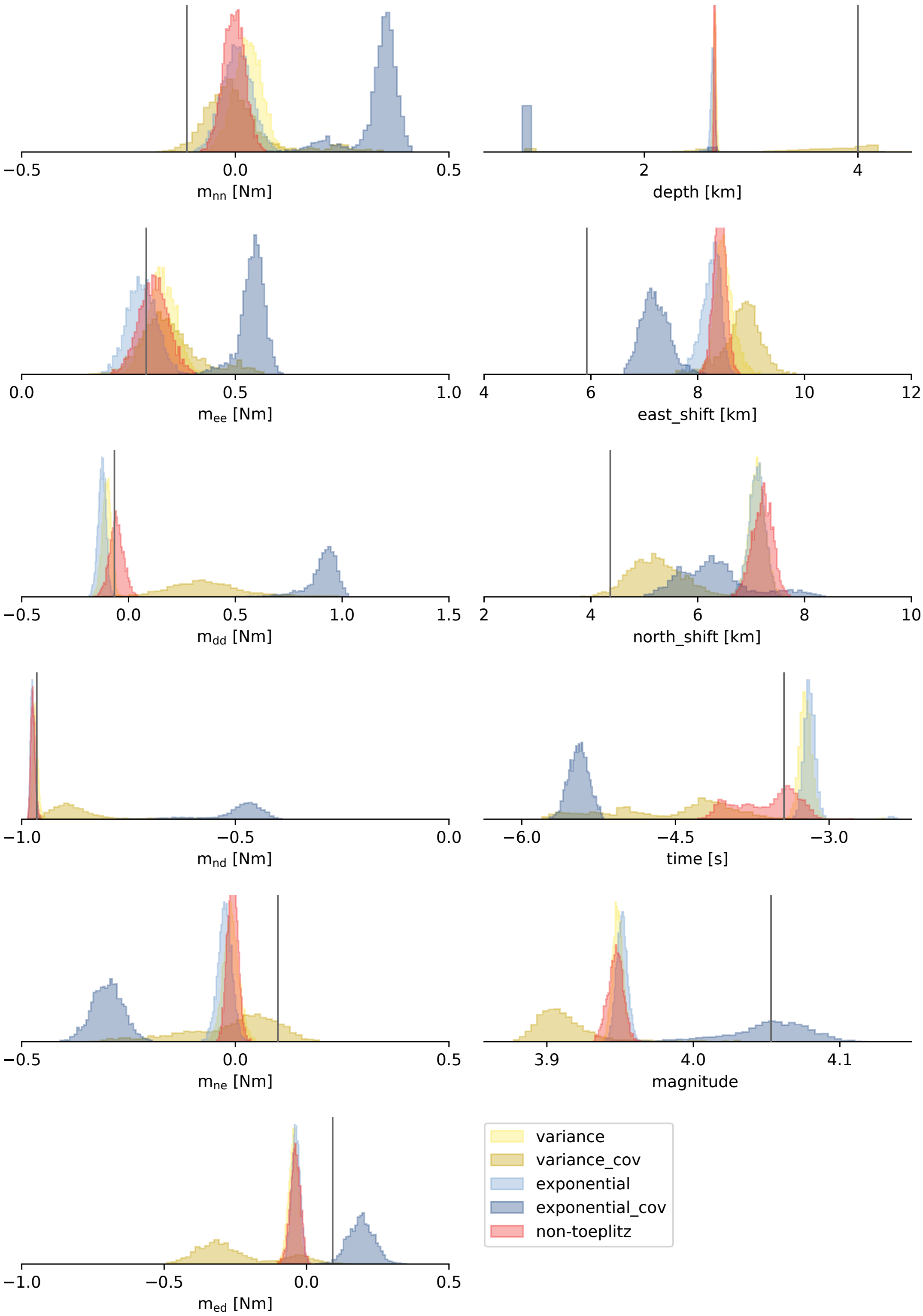

Figure 10. Posterior marginals for the 2015 Fox Creek event inverted assuming a regional velocity model. The location estimates are relative to the reference location (NEIC, 54.102° N and 116.95°W). Colors same as Fig. 4. The solution of Wang et al. (2016) are also shown (gray lines). Here, overestimation of theory errors leads to parameter biases for the modelling approaches.

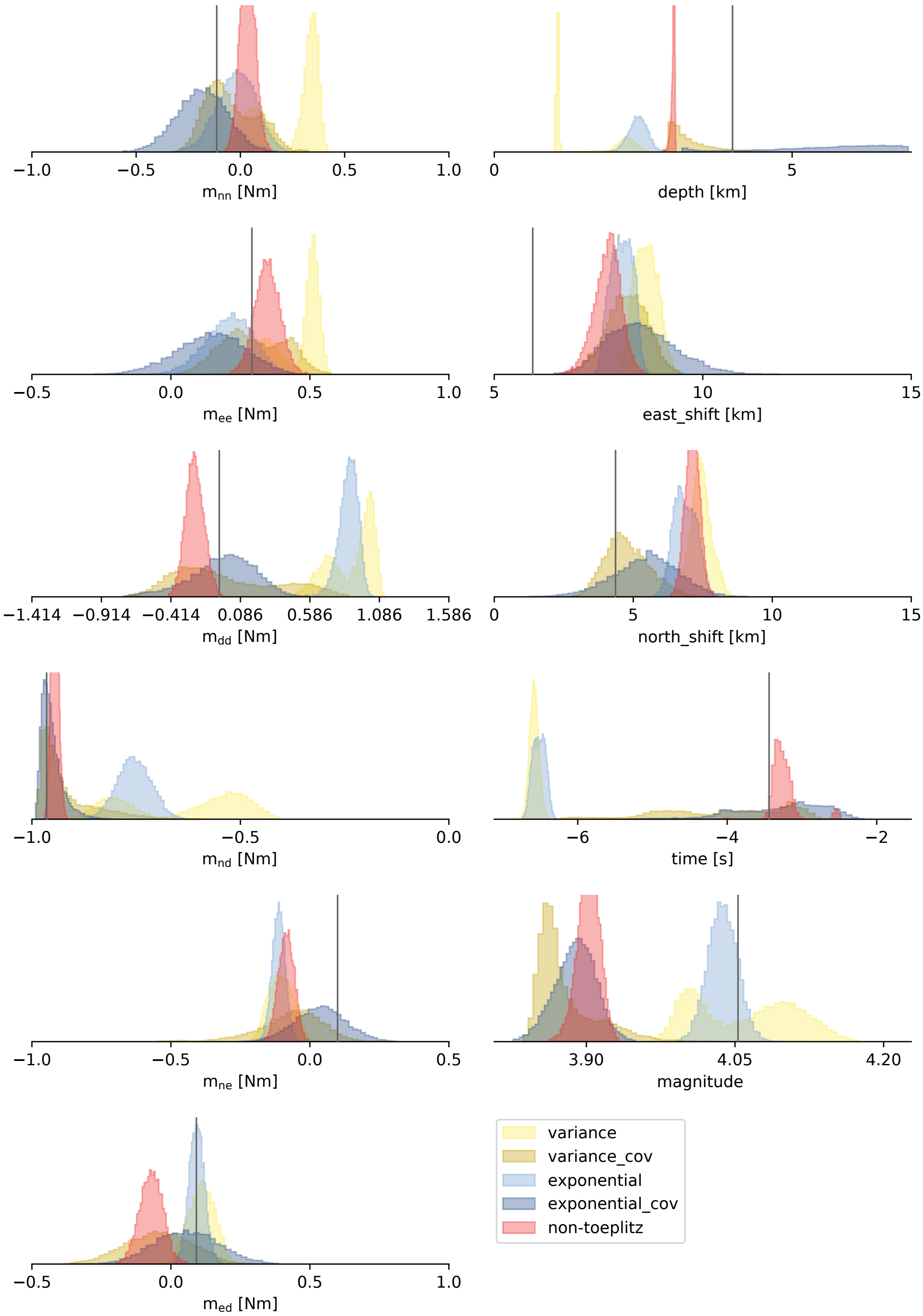

Figure 11. Same as Fig. 10 but GFs were computed for global velocity model. Results show the advantage of including theory errors when the velocity structure is poorly known *a priori*.

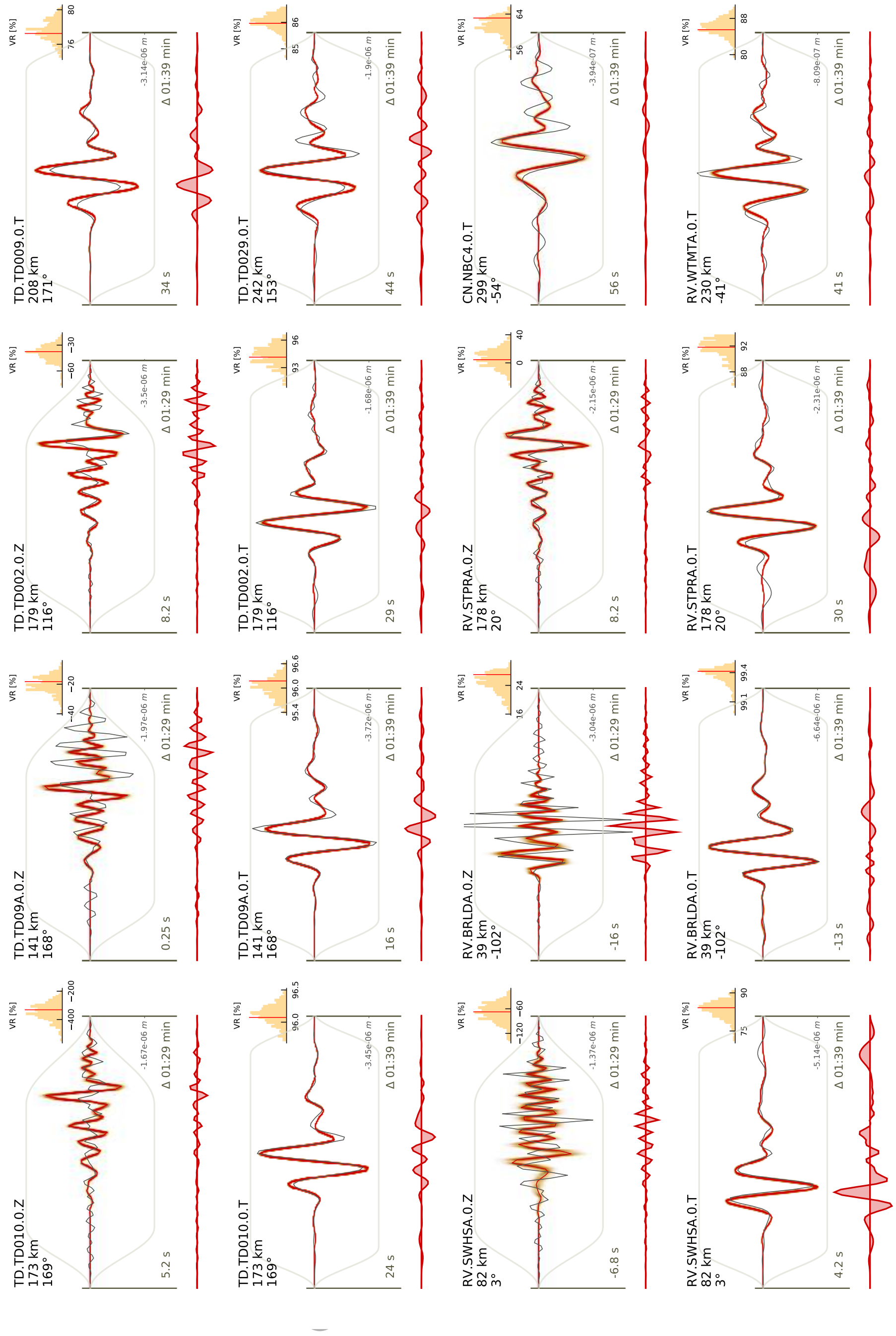

Figure 12. Waveform fits for the full moment tensor solution with *variance* noise parameterisation using the regional subsurface structure. The filtered displacement waveform data (dark grey solid line) for body (vertical Z-component 0.08-0.3Hz) or surface wave arrivals (transverse T-component 0.04-0.1Hz) and appropriate predictions (red solid line) are shown. The brown shading is for 100 randomly selected waveforms from the posterrior predictive distribution. The residual waveforms are shown below each waveform as red lines with filled polygons. Waveform are normalized with respect to the component (Z and T). Traces are annotated with station name, component, epicentral distance, and azimuth obtained for the maximum *a-posteriori* centroid. The arrival time with respect to the centroid time, and the duration of each window are shown in the lower left and right, respectively. The orange histogram in the top right of each panel shows the weighted variance reduction (VR) for the posterior predictive distribution.

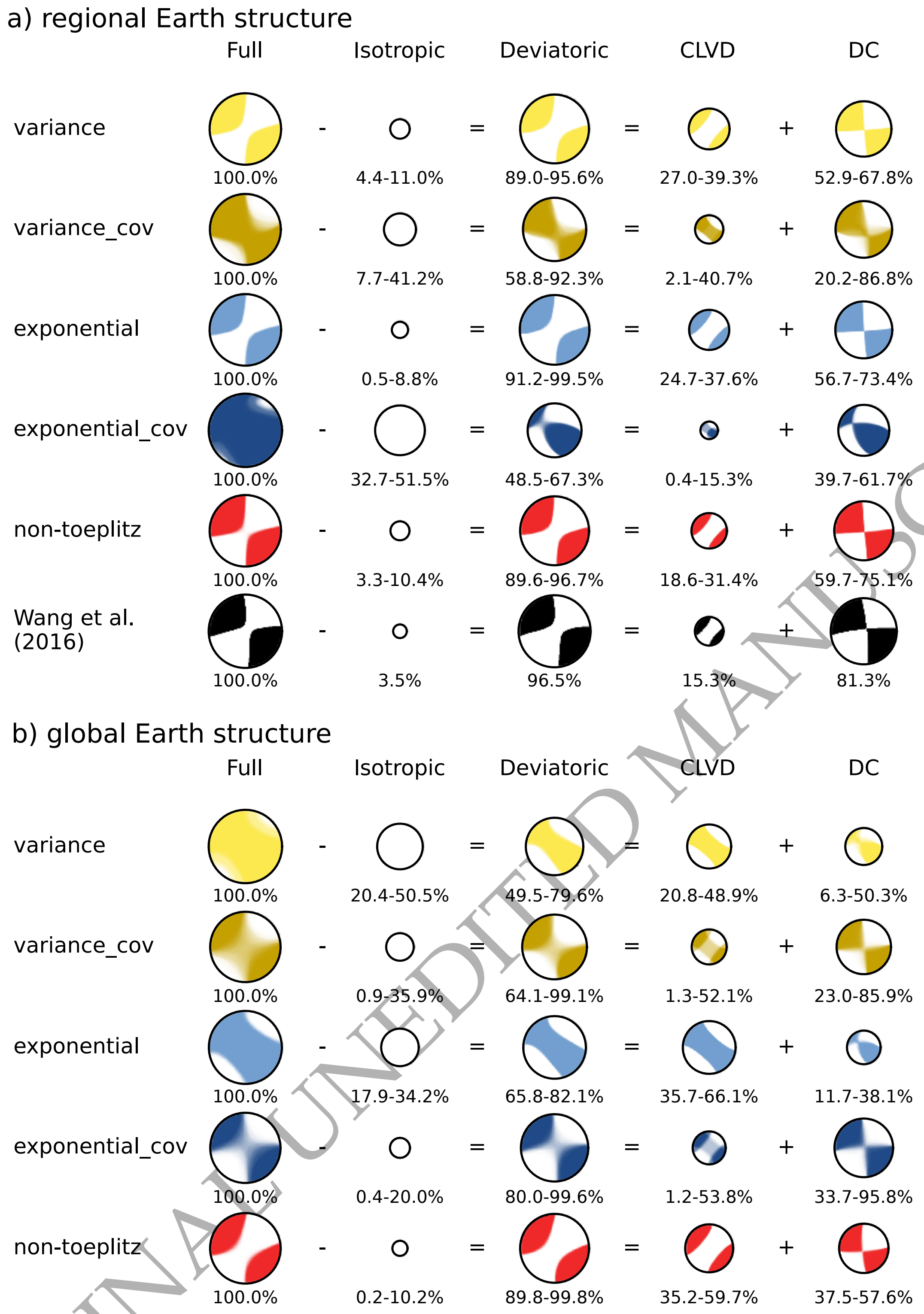

Figure 13. Same as Fig. 7 but for the 2015 Fox Creek event: Moment tensor decompositions for results from the various noise parameterisations for a) regional Earth structure and b) global Earth structure.

APPENDIX A: SOURCE PARAMETERISATIONS

The moment tensor parameterisations in this manuscript include unknown centroid locations (east-shift, north-shift, depth), centroid time, source duration, and the MT specific parameterisations described here. The MT components are used as weights for the Green's Functions (GF) to compute synthetic seismograms (see eq. (8) in Heimann et al. (2019)).

A1 Full Moment Tensor

A seismic source can be represented by a point source if its seismic moment is sufficiently small so that spatial extent of the source is small compared to the distance where it has been recorded. Such representations are symmetric 3 × 3 tensors **MT**, with components $m_{xx}, m_{yy}, m_{zz}, m_{xy}, m_{xz}, m_{yz}$ (e.g., Aki & Richards 2002). Here, x, y, z are coordinates with various conventions possible. The scalar moment of the full MT is (Silver & Jordan 1982; Stähler & Sigloch 2014)

$$M_0 = \frac{1}{\sqrt{2}}\sqrt{m_{xx}^2 + m_{yy}^2 + m_{zz}^2 + 2(m_{xy}^2 + m_{xz}^2 + m_{yz}^2)}. \tag{A.1}$$

To sample numerically stable values for scalar moment, we sample moment magnitude $M = 1.5 * \log 10(M_0 * 10^7) - 10.7$ (Hanks & Kanamori 1979). Note, that we use SI units i.e., Nm and the formulation is valid for $dyne - cm$, thus the conversion factor of 10^7 is needed. Here, we adopt the coordinate system of North, East and down ($\{n, e, d\}$) (Aki & Richards 2002).

$$\begin{gathered} M_0 = \frac{1}{\sqrt{2}}\sqrt{m_{nn}^2 + m_{ee}^2 + m_{dd}^2 + 2(m_{ne}^2 + m_{nd}^2 + m_{ed}^2)} \\ M_0 = 10.0^{1.5*(M+10.7)} * 1.0^{-7} \\ m_{nn}, m_{ee}, m_{dd} \sim U(\sqrt{-2.}, \sqrt{2.}) \\ m_{ne}, m_{nd}, m_{ed} \sim U(-1., 1.), \end{gathered} \tag{A.2}$$

where m_{ij}, with $i, j \in \{n, e, d\}$, are the moment tensor components with uniform prior probabilities.

A2 Deviatoric Moment Tensor

For the parameterisation of a deviatoric moment tensor $\mathbf{M}_{dev}$ we sample the solution space according to the formulation in Sec. A1, but we subtract the isotropic part of the moment tensor $\mathbf{M}_{iso}$ prior to forward modeling (Jost & Herrmann 1989).

$$M_{trace} = \frac{m_{nn} + m_{ee} + m_{dd}}{3}$$

$$\mathbf{M}_{iso} = \begin{bmatrix} M_{trace} & 0 & 0 \\ 0 & M_{trace} & 0 \\ 0 & 0 & M_{trace} \end{bmatrix} \tag{A.3}$$

$$\mathbf{M}_{dev} = M_0 \cdot (\mathbf{MT} - \mathbf{M}_{iso}).$$

A3 Double Couple Source

A double couple (DC) source can be described by the dip, strike and rake angles of the fault. For slip on a buried horizontal plane in only the x direction, the moment tensor is (Aki & Richards 2002):

$$\hat{\mathbf{M}}_{dc} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}. \tag{A.4}$$

This plane can be rotated by the dip (α), strike (β) and rake (γ) angles around the x, y and z axes, respectively. We use the Euler angle rotation formulation (Goldstein et al. 2001) to calculate a rotation matrix $\mathbf{R}$. The moment tensor expression for a pure DC source is obtained by

$$\mathbf{M}_{dc} = M_0\, \mathbf{R}(\alpha, \beta, -\gamma)^T\, \hat{\mathbf{M}}_{dc}\, \mathbf{R}(\alpha, \beta, -\gamma). \tag{A.5}$$

APPENDIX B: SAMPLING ALGORITHM

Using a Monte Carlo method allows drawing samples from a posterior PDF (eq. 1); once enough samples are drawn the resulting distribution is a valid approximation of the posterior probability density (PPD). To sample the PPD we use a Sequential Monte Carlo (SMC) sampler (Moral et al. 2006; Ching & Chen 2007), similar to Minson et al. (2013). Here, we outline the main features of the algorithm, however, for more details we refer the reader to the original references. Obtaining samples from a posterior probability density function (PDF) that has a complex topology (high-dimensional, multimodal, flat, ...) is difficult and inefficient. Therefore, sampling is done starting from the prior PDF via several intermediate PDFs that change following a self adjusting cooling parameter starting at zero (similar to Simulated Annealing (Sambridge & Mosegaard 2002)) (Moral et al. 2006; Minson et al. 2013):

$$f(\mathbf{m}|\mathbf{d}^{obs}, \beta_l) \propto p(\mathbf{d}^{obs}|\mathbf{m})^{\beta_l} p(\mathbf{m})$$
$$l = 0, 1, ..., L \qquad \text{(B.1)}$$
$$0 = \beta_0 < \beta_1 < ... < \beta_L = 1$$

Each intermediate PDF $f(\mathbf{m}|\mathbf{d}^{obs}, \beta_l)$ is sampled in parallel by a pre-defined number of Monte Carlo (MC) chains. Each chain samples the solution space with a predefined number of steps, where step size and directions are determined according to a proposal distribution. When sampling of all chains for the intermediate PDF is completed the algorithm enters a transitional stage:

(i) The likelihood of each Markov chain end-point is used to form an intermediate likelihood distribution.

(ii) This likelihood distribution (at β_l) is compared to the previous intermediate likelihood distribution (at β_{l-1}) by evaluating the coefficient of variation (COV). If they differ significantly (COV $>$ 1) the cooling parameter β_l is incremented only by a small amount. On the other hand, if the distributions are similar (COV $<$ 1) the tempering parameter β_l is increasing faster.

(iii) The proposal distribution is updated based on the distribution of model parameters in the MC chain end-points.

(iv) Optional: update $\mathbf{C}_k$ in each transitional stage using the mean of each model parameter distribution (Dettmer et al. 2007; Minson et al. 2013; Duputel et al. 2014) (see eq. 3).

(v) The ensemble of Markov chain end-points at β_{l-1} is resampled according to the intermediate likelihoods. Hence, the next stage of Markov chains at β_l are seeded on the end-points of the previous chains, which had the highest likelihoods; unlikely chains are discarded.

Finally, if the cooling parameter satisfies $\beta_l \geq 1$, the posterior distribution is reached

$f(\mathbf{m}|\mathbf{d}^{obs}, \beta_L = 1) \propto p(\mathbf{m}|\mathbf{d}^{obs})$ and one last sampling of all MC chains with the defined number of steps is executed; then the algorithm stops. For the proposal distribution we use a multivariate Gaussian distribution similarly to Minson et al. (2013).