On the practical convergence of coda-based correlations: a window optimization approach

J. Chaput, V. Clerc, M. Campillo, P. Roux, H. Knox

On the practical convergence of coda-based correlations: a window optimization approach

J. Chaput,[1] V. Clerc,[1] M. Campillo,[1] P. Roux[1] and H. Knox[2]

[1]*ISTERRE, Université Joseph Fourier, Grenoble, F-38000, France. E-mail: jchaput82@gmail.com*
[2]*Geophysics and Atmospheric Sciences, Sandia National Laboratories, Albuquerque, NM, USA*

SUMMARY

We present a novel optimization approach to improve the convergence of interstation coda correlation functions towards the medium's empirical Green's function. For two stations recording a series of impulsive events in a multiply scattering medium, we explore the impact of coda window selection through a Markov Chain Monte Carlo scheme, with the aim of generating a gather of correlation functions that is the most coherent and symmetric over events, thus recovering intuitive elements of the interstation Green's function without any nonlinear post-processing techniques. This approach is tested here for a 2-D acoustic finite difference model, where a much improved correlation function is obtained, as well as for a database of small impulsive icequakes recorded on Erebus Volcano, Antarctica, where similar robust results are shown. The average coda solutions, as deduced from the posterior probability distributions of the optimization, are further representative of the scattering strength of the medium, with stronger scattering resulting in a slightly delayed overall coda sampling. The recovery of singly scattered arrivals in the coda of correlation functions are also shown to be possible through this approach, and surface wave reflections from outer craters on Erebus volcano were mapped in this fashion. We also note that, due to the improvement of correlation functions over subsequent events, this approach can further be used to improve the resolution of passive temporal monitoring.

Key words: Inverse theory; Interferometry; Volcano seismology; Antarctica.

1 INTRODUCTION

Seismic interferometry is a relatively new methodology that allows for new seismic responses to be created by cross-correlating seismic signals at different locations in a medium. Though the principle is simple, a great deal of effort has been put forward to establish theoretical and numerical scenarios through which the cross-correlation of a given wavefield recorded between two receivers A and B does in fact approach the Earth's impulse response, or Green's function (GF). The wavefield must satisfy certain theoretical requirements, namely that the medium be sampled isotropically in terms of incident amplitude. For the 3-D case, the derivative of the cross-correlation function with respect to time is proportional to the GF from A to B minus the time reversed GF from B to A (e.g. Sabra *et al.* 2005), or

$$\frac{\partial (C_{AB})}{\partial t} \approx G(A, B; t) - G(B, A; -t). \tag{1}$$

Here, C_{AB} denotes the cross-correlation between receivers A and B. The correlation function, prior to time derivation, is therefore expected to be symmetric in its causal and acausal portions (referred to here as 'causal–acausal symmetry') for this relation to hold true.

In practice, the wavefield requirements are satisfied by one of two particular scenarios, or a combination thereof.

The first, for which substantial medium complexity is not a prerequisite, relies on a complete source distribution surrounding the medium to make use of stationary phase components of the source averaged integrand (Snieder 2004; Roux *et al.* 2005; Wapenaar & Fokkema 2006). In practice, this is essentially unattainable given the source distribution requirements, though partial reconstructions are often possible when the source is a distribution of ambient noise and is not extremely skewed (Campillo & Paul 2003; Shapiro and Campillo 2004; Paul *et al.* 2005; Sabra *et al.* 2005; Tonegawa *et al.* 2009; Tonegawa & Nishida 2010; Poli *et al.* 2012a). However, while direct phases are easily reconstructed from such correlations in real media, the scattered portion of the GF has proven far more elusive, particularly in light of the general asymmetry of the recovered correlation functions for a single station pair.

The second case relies on scattering from complexity inherent to the medium to progressively randomize the source wavefield and homogenize the apparent source distribution, thus indirectly satisfying the requirements of the first case (Shapiro *et al.* 2000; Lobkis & Weaver 2001; Weaver & Lobkis 2004; Sanchez-Sesma & Campillo 2006; Sato 2009; Colombi *et al.* 2014). For a given

local impulsive event, this homogenization of the wavefield occurs in the coda portion of the seismogram if multiple scattering effects can lead to a state of diffusion before absorption effects become important (Paul *et al.* 2005). The subsequent cross-correlation of a given window in this multiply scattered coda ideally converges towards the impulse response between the two receivers. Stehly *et al.* (2008) has further noted that correlations of the coda of noise correlations (C3) also converge towards the GF, indicating that ambient noise correlations do contain a certain level of scattered phases. In the same stroke, it was noted that correlations of coda converge towards the GF faster than correlations of ambient noise. C3 studies notwithstanding, the difficulty here lies in the general paucity of relevant events, and finding suitably scattering media to enable sufficient wavefield randomization.

Focusing on a complex scattering medium where a large amount of small impulsive icequakes were recorded on a dense network of stations deployed on Erebus Volcano, Antarctica, Chaput *et al.* (2015) noted, through interstation cross-correlations of the coda of these icequakes, that (1) the direct Rayleigh waves in the cross-correlations were stable over most chosen coda windows and (2) the coda of the cross-correlation function was subject to fluctuations in phase and amplitude for these same time windows. Furthermore, the correlation functions converged towards maximal causal–acausal symmetry after 300–500 stacked correlations, and any further addition of events to the source average (up to 3318) resulted in minimal or inexistent gains to this metric, a phenomenon previously documented by Larose *et al.* (2007). Though excellent causa–acausal symmetries in the correlation functions were noted in some cases, the fluctuations in the coda portion of the correlation function imposed limits on how well the reconstructions could be related to the interstation GF being estimated. Thus, even in situations where discrete events are abundant and the medium is highly heterogeneous, a given choice of event coda windows may result in an incorrectly reconstructed correlation function despite yielding very well resolved direct waves.

In essence, this means that scattering-based methodologies where the coda of the correlation function plays a dominant role (i.e. efforts aiming to separate the single scattered from multiple scattered wavefield arrivals (Shahjahan *et al.* 2013), or direct reflection studies using correlation gathers of ambient noise or coda (Draganov *et al.* 2009; Chaput *et al.* 2012), will contain an unknown degree of error unless a better estimate of the true GF can be made.

In this study, we begin by exploring the following question: Given that coda-driven correlation functions for highly scattering media are built from a heuristic choice of coda windows applied uniformly to all events, could we improve the reconstruction by selecting better, or ideal, coda windows for each event? Given that coda in real media undergo absorption, there should be a time window in the coda generated by a given discrete event where the wavefield is relatively diffuse and still contains adequate energy. The cross-correlation of this window at two receivers should maximize the constructive contribution to the GF of that particular event. Furthermore, in the cases explored here, the media of interest are either invariable (numerical models) or very stable (Erebus upper edifice) at the time scales explored. We expect that the correlation of a repeatedly excited diffuse wavefield will yield not only causal–acausal symmetry for each correlation function, but the same symmetry (and not simply a different reconstruction of symmetry) for every computed correlation function (referred to in this paper as 'event-time', or 'event-time coherence'). We have therefore identified an intuitive set of criteria that allows the optimization of the coda windows of individual events, with a goal of recovering a matrix of correla-

tion functions that is both symmetric with respect to the causal and acausal portions, and coherent over events.

Here, we make use of numerical experiments as well as a database of small icequake signals recorded on Erebus volcano (Chaput *et al.* 2015; Knox 2012) to explore the window selection problem. In essence, we wish to test the effect of event-dependent coda time windows on the resulting correlation gathers, while weighing the intuitively derived criteria described above. It should be noted here that correlation functions are typically computed using spectral whitening and temporal normalization processing of the raw signal. Here, we simply bandpass filter the data (between 1.5 and 4 Hz for Erebus icequakes), and cross-correlate tapered segments of coda, thus circumventing the somewhat unclear question of the impact of these nonlinear methods on the correlation function.

2 CODA WINDOW OPTIMIZATION

The optimization problem described in the introduction can be solved in a variety of ways. Here, the problem is explored through a Markov Chain Monte Carlo (MCMC) scheme, with a goal of iteratively readjusting the coda window selected for each event to create an ideally converging correlation gather. Though computationally expensive, this approach allows us to obtain a distribution of coda time window solutions (in the form of the posterior probability distribution, PPD; see Appendix A) and not only a single best fit.

As stated above, if we assume that the coda of the events used are diffuse and approaching equipartition, then the recovered correlation functions, stacked over enough events to suppress random noise, should converge towards symmetry. If the events are further recorded within a relatively small time frame under conditions of minimal structural change, then the GFs should be coherent over event-time.

Fig. 1 presents a general conceptual flowchart of the optimization process. Note that this description is not specific to MCMC implementations, but simply describes the basic idea.

(1) For a matrix of M events recorded on a given station pair (denoted as A and B), the initial step consists of computing a first guess cross-correlation gather MCF, consisting of M cross-correlations, by selecting an identical arbitrary window of length T and start time t_i in the coda of all events in the raw data matrices, and cross-correlating these coda windows between stations. A single event cross-correlation will typically present very low coherence with any other cross-correlation regardless of the window chosen. Therefore, the rows of matrix MCF must be averaged by an arbitrary number n_{bin} to produce a binned MCF gather of N rows, where $N = M/n_{bin}$. n_{bin} is dependent on the nature of the data set, and is chosen experimentally to ensure adequate correlation function convergence in for each row of the binned MCF matrix. Two metrics are then computed: (i) a matrix COH of size $(N \times N)$, consisting of the coherence of each row MCF_i with each other row MCF_j, and (ii) a vector SYM of binned MCF row symmetries of size $(N \times 1)$, consisting of the coherence of the acausal and causal portion of every row of MCF (i.e. coherence of the causal portion of MCF_i with the time reversed acausal portion of MCF_i).

(2) The subsequent optimization can be formulated as such: Select at random an event in the raw data matrix, change the starting time of its coda window, and compute the corresponding cross-correlation function. Update the matrix MCF by inserting the new cross-correlation (replacing the old version), and recalculate COH and SYM.
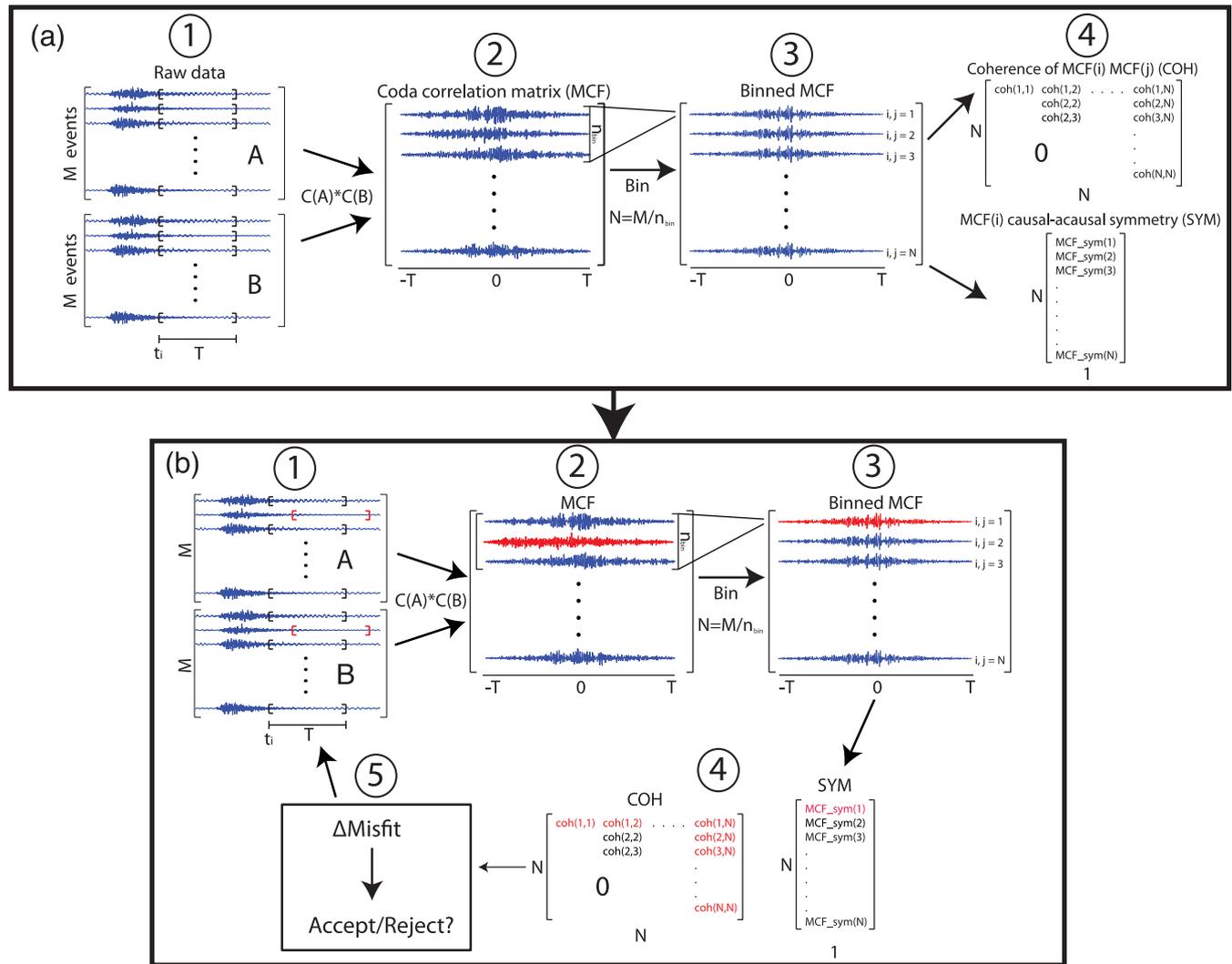
Figure 1. Description of the optimization procedure. The initial step (a) consists of the standard procedure for coda based cross-correlation functions, and is used as a starting point for the optimization (b). (A1) Begin with two gathers of M raw-picked events recorded at stations A and B, and select an arbitrary coda window of length T starting at t_i for all events. (A2) Assemble a matrix MCF of M coda cross-correlations. (A3) Average MCF rows by a number n_{bin} to obtain a binned MCF gather of N binned cross-correlations, where $N = M/n_{\text{bin}}$. (A4) Compute metrics COH (calculated by taking the coherence of each binned MCF row with each other row) and SYM (defined as the coherence of the causal and acausal portions of each MCF row). (b) Optimization process. Red traces indicate proposed changes for a given iteration. (B1) Choose an event in the data matrix and shift the coda window at random. (B2) Recalculate MCF. (B3) Recalculate the binned MCF. (B4) Recalculate both COH and SYM. (B5) Accept or reject the proposed shift based on the misfit defined by eq. (2). Repeat B1–B5 until convergence is achieved.

(3) For this new perturbed set of metrics COH and SYM, accept or reject the change based on a misfit function defined by eq. (2). Repeat steps 2 and 3 until the change in misfit falls consistently below a preset value, or for a set number of iterations (in the MCMC case).

The goal of the optimization problem is to maximize both COH and SYM, and the misfit can thus be defined simply as:

$$\text{MSF} = A \left(1 - \frac{2}{(N^2 - N)} \sum_{i>j} \text{COH}(i, j) \right)$$

$$+ B \left(1 - \frac{1}{N} \sum_{i} \text{SYM}(i) \right) \tag{2}$$

where i, j define the ith row and jth column of the coherence matrix COH, or the ith element of the symmetry vector SYM. Note that

it is more computationally efficient to only calculate the upper half of matrix COH (as it is symmetric). Further rejecting the coherence of MCF_i with MCF_i (which is always equal to 1) leads to the normalization factor $1/(N^2 - N)/2$. A and B are weights that cause the solutions to converge preferentially towards one trait or another. For our experiments, we weight COH and SYM evenly. To do so, we allow 1000 iterations prior to burn-in and test the sensitivity of the misfit function to random variations t_i. The goal here is to choose values of A and B that result in equal contributions by the first and second terms of the right-hand side of eq. (2) to a change in *MSF*.

Given that we apply a MCMC algorithm here, it may be useful to define some of the terminology used in subsequent sections. The term 'model' refers to the set of coda window start times t_i used in the construction of the current cross-correlation function gather at a given iteration in the optimization. Naturally, t_i varies on an event to event basis. The model changes with every iteration (if the proposed model is accepted), and the final model is typically
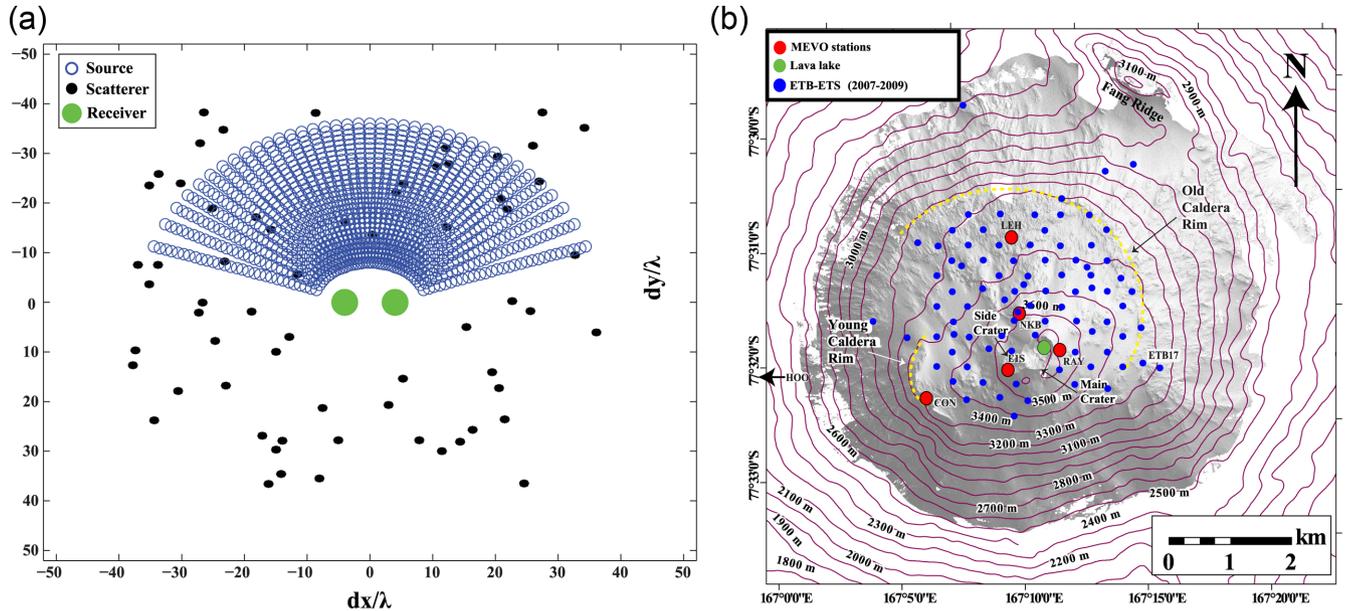
Figure 2. Experimental geometries. (a) Example of a medium used in numerical studies for 64 scatterers, 2 receivers, and 1130 sources. Such random media were also generated for 128 scatterers. (b) Map of Erebus volcano and seismic stations considered in this study.

the mean of the posterior distribution of these accepted models (see Appendix A for Bayesian description). The step involving the n_{bin}-averaged MCF can be justified by noting that a correlation function converges to stability after many individual correlation functions are summed, and indeed, in our experiments, attempts to optimize windows without this binning step failed due to low coherence between individual correlation functions. The number n_{bin}, which defines the number of correlations in MCF used to create each row of the binned MCF (see Fig. 1), should reflect the rate at which the correlation function converges to stability, and will vary as a function of the data set. This number should furthermore be large enough to allow adequate correlation stability within every row of the binned MCF, but small enough to allow a suitably large number of these rows. The consideration here is the trade-off between computational cost (i.e. a very small n_{bin} means more coherences need to be calculated at every iteration) and a more full description of temporal stability of the resulting correlation functions. In our case, this somewhat arbitrary number is set at 400 correlations/bin for real data (for icequakes, (Chaput *et al.* 2015), and 100 correlations/bin for numerical models. COH and SYM are therefore calculated from this binned version of MCF, and a 'correlation bin' refers to a row of this binned MCF. The impact of the window time length T is not evaluated here, and is arbitrarily set at 45 periods for synthetic media, and 15 s for subsequent real data.

We will apply this optimization to several examples of synthetic media, where the particularities of the solutions may be thoroughly analysed, and will then demonstrate the applicability of such an approach to real data. The advantage here is a natural convergence towards a solution that presents traits inherent to the GF, essentially retrieving persistently symmetric elements in the cross-correlation functions.

3 NUMERICAL SIMULATIONS AND REAL DATA GEOMETRY

On the tail of rapid advances in computation power, numerical schemes have recently been used to explore the character of multiply

scattered diffuse wavefields, both from a radiative transfer perspective (Margerin *et al.* 2000; Margerin & Nolet 2003) and from direct 2-D simulations (De Rosny & Roux 2001; Catheline *et al.* 2011; Davy *et al.* 2013). We adapt a 2-D finite difference (Liu & Tao 1997) acoustic simulation based on a linearization of Euler equations with conservation of mass and momentum. The simulation consists of a second order staggered leapfrog scheme, with perfectly match layer (PML) boundaries, so as to absorb a maximum of resonant waves in the medium and permit adequate field randomization over long run times. Discretization parameters in time and space for this numerical simulation are dt and dx and are defined such that $f * dt = 12.5$ and $\lambda/dx = 10$, where f and λ are the central frequency and central wavelength of the acoustic signal. Our setup consists of a medium surrounded by absorbing boundaries and embedded with 64 or 128 dense impenetrable circular scatterers with a radius of 0.8λ. Two receivers (at $[-4\lambda\ 0]$ and $[4\lambda\ 0]$) are located in the centre of the scattering zone, and we use an array of 1160 sources distributed in a symmetric pattern in the upper half of the medium around the receivers. The total synthetic medium size is $104\lambda * 104\lambda$ with a scattering zone of dimensions $80\lambda * 80\lambda$. The coda window length is arbitrarily chosen to be 45 periods for numerical tests, and the impact of this value on final window solutions is not explored here.

Fig. 2(a) depicts an example of the distribution of sources, scatterers, and receivers used in this study (64 scatterers shown). We are ultimately interested in the differences in reconstruction between a heuristically chosen stack of correlation windows (as described in the initial step of Fig. 1) and the optimized window solutions (optimization step of Fig. 1). Furthermore, the final distribution of coda window start times t_i for each event as provided by the PPD output of the MCMC optimization could also contain information about the nature of the wavefield, should the resulting GF estimate be systematically superior. For synthetic media, we allow 60 000 MCMC iterations in total. For a value of n_{bin} of 100 and 1160, the run time for one model on a single core system is roughly 15 min, though this can likely be improved upon.

Additionally, a database of small impulsive icequakes recorded on Erebus volcano is used here to compute coda correlations for a one month period during the 2008–2009 Antarctica summer season

(a description of this data set and of Erebus volcano characteristics can be found in (Aster *et al.* 2004; Knox 2012; Chaput *et al.* 2015). A total of 3318 small impulsive events were recorded on a network of 94 broadband and short period stations (ETS-ETB temporary array) deployed over the summit region (Fig. 2b), with 30 broad-band stations further recording over 21 000 icequakes over the year. Correlation gathers were computed for every station pair, and we also included data from permanent Mount Erebus Volcanic Observatory (MEVO) broad-band stations that overlapped with this time period. Here we re-compute cross-correlation functions via MCMC for a large subset of 2850 interstation paths (for the database of 3318 icequakes) to evaluate the gains in symmetry with respect to the original stacks as described in Chaput *et al.* (2015). We then compare MCMC cross-correlations to standard cross-correlations for a series of arbitrarily chosen lines within the array to qualitatively evaluate gathers. Finally, we perform basic surface wave elliptic back-propagation for both standard and MCMC cross-correlations to test whether the recovered scattered information corresponds to any known topography for a single-scatter assumption.

4 NUMERICAL SIMULATION RESULTS

Fig. 3 presents examples of numerical simulations for synthetic media featuring 64 scatterers. Fig. 3(a) presents the mean posterior model (mean of the PPD) compared to a single window standard cross-correlation and the true GF. Fig. 3(b) shows the convergence of the average window solution (i.e. average of figure 3 c over all sources, for every iteration, including burn-in) for four runs of the same medium with different prior MCF computed by a single window starting at periods 30, 70, 120 and 200. The average converged coda time-window solutions are essentially identical despite the varying choice of prior windows, pointing to the existence of an 'ideal window range'.

When computing coda correlations conventionally, a heuristic approach is generally assumed given the lack of prior information about the true GF (i.e. all coda windows, barring analysis of modal stability (Hennino *et al.* 2001; Chaput *et al.* 2015) or knowledge of the mean free path, are equal). Should the coda of recorded events be sufficiently long, it is common to stack the cross-correlated contributions of several sequential coda windows.

In the interest therefore of computing the best possible standard correlation for comparison (that is, beyond the single window approach described in the top window of Fig. 1, here denoted as the 'single window standard' correlation), we use the following approach for our numerical simulations. We select a first correlation

window of length $T = 45\lambda$ starting early in the coda at an arbitrary time t_i (estimated), allow a 75 per cent overlap while shifting t_i slightly later in the coda, and compute as many subsequent windows as possible until the end of the coda. We then average over this set of correlations. This approach is subsequently referred to as the 'multi-window standard' correlation. Though there is absorption at the medium boundaries, the recorded signals show very weak decay on account of the strong scattering and the otherwise lossless wave propagation. Thus, the standard correlations in our setup feature the upper bounds of what should be recoverable through the standard approach given that there is a large amount of multiply scattered coda to stack.

Given the relatively tight bounds of the average solution (Fig. 3b), we can infer that the algorithm converges to a general balance between the extent of the scattering in the wavefield, the gradual wavefield decay due to absorption at the boundaries, and the chosen parameter weights. The solutions for each prior converge towards the same general average, though there are fluctuations that do not fully stabilize given the limited number of iterations and the nature of MCMC sampling. Fig. 3(c) shows an example of the misfit function for an MCMC run. Note that we only retain the models after 30 000 iterations.

In the interest of seeing if the average coda window start time t_i changes as a function of scattering strength and optimization parameters, we compute MCMC correlations for 90 realizations of random media. Fig. 4(a) shows an example of the PPD for one such medium, where each row represents a colour-mapped histogram depicting the probability of t_i for a given source. For each medium realization, we take the average of Fig. 4(a) over all sources (rows) to recover an estimate of the total range of t_i explored by the optimization for various parameter weights and scattering densities (noted as the 'Total Average PPD'). In particular, we are interested in seeing if solutions converge to similar windows for 64 and 128 scatterer media, and if the solutions t_i are mainly governed by event-time coherence or by causal–acausal symmetry. Fig. 4(b) shows several examples of these total average PPDs. We can make two general observations here: 1) The total average PPD solutions are slightly later for increased scatterer density, and 2) The optimization, at least for this numerical setup, is largely driven by symmetry and not coherence (i.e. correlation symmetry leads to coherence over events, but the reciprocal is not necessarily true). This may in part be explained by a higher level of trapped scattering for denser scatterer distributions, and more gradual propagation of the diffusive energy halo. The differences in the total averaged PPDs between media with 64 and 128 scatterers are however relatively small, which can in part
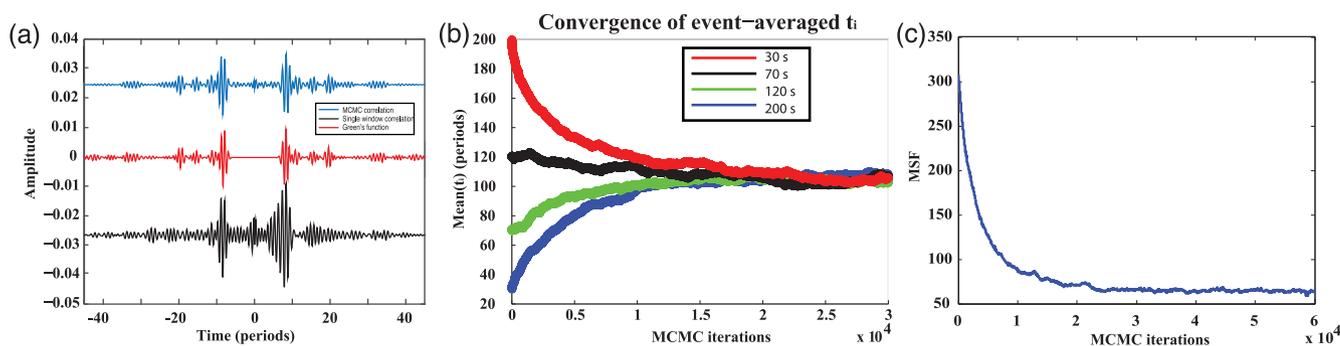

Figure 3. Numerical experiments, 64 scatterers. (a) Optimized results versus single window standard correlation and the true GF for one version of the scattering medium. (b) Convergence of the average window solution for a variety of average prior windows. The window is 45 periods long, and the total length of the receiver recording is 250 periods. The colours are indicative of runs with different window start positions. (c) Example misfit function for a random medium featuring 64 scatterers. Models after 30 000 iterations are stored.
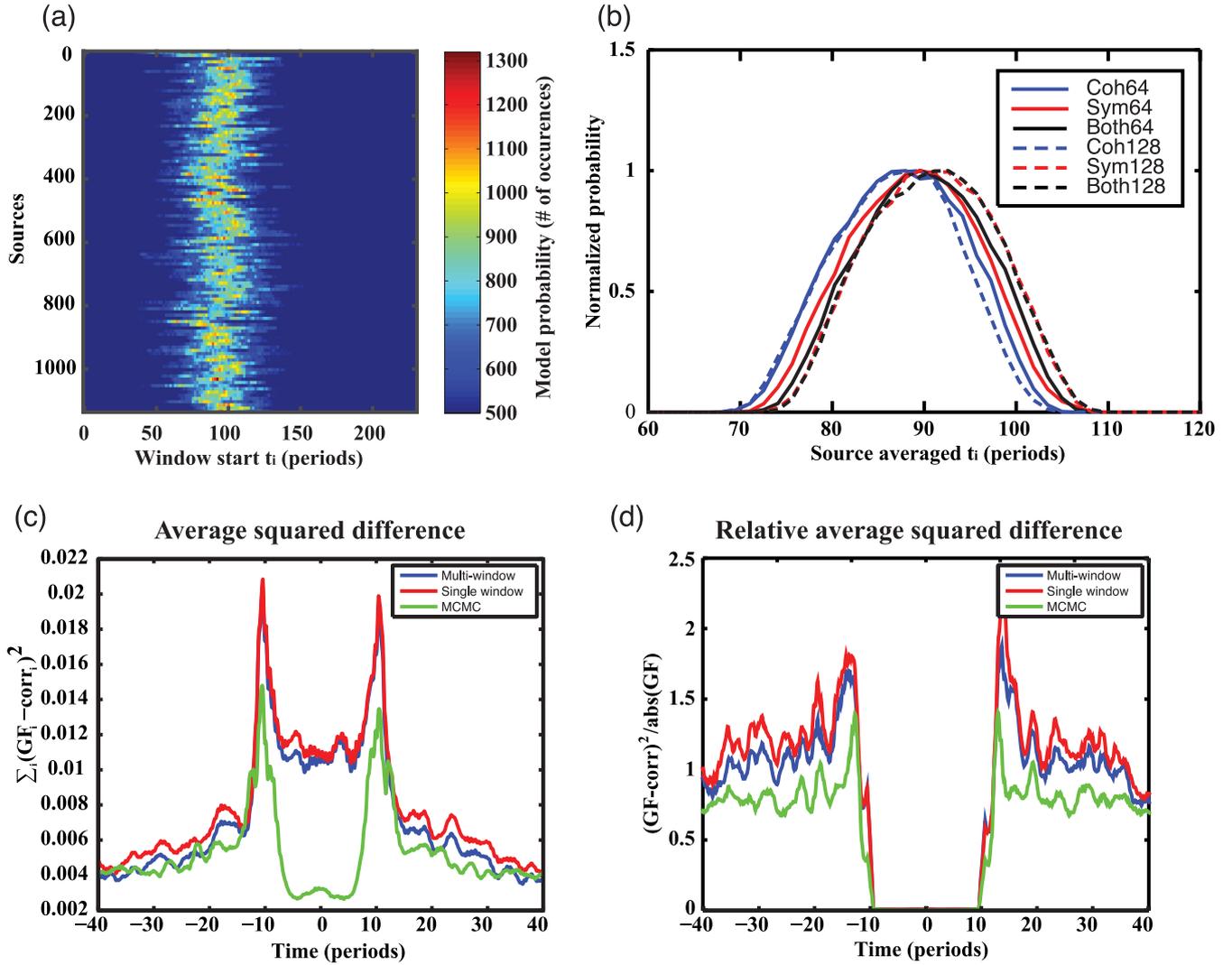
Figure 4. (a) Colour-mapped histogram output of the MCMC algorithm for 1 medium realization. (b) Fig. 4(a) averaged over all sources and 90 media. Curves labelled 'Coh' indicate a 20-1 weighting in favour of event-time coherence over causal–acausal symmetry. 'Sym' labels indicate a 20-1 weighting in favour of causal–acausal symmetry, and 'Both' indicates equal weighting. (c) Squared difference of the MCMC correlations and standard correlations (with single and multiple window methods) with respect to the GF averaged over 90 media realizations with 128 scatterers. The matching of differences between the methods after ±35 periods is an effect of the window length chosen in the correlations (i.e. the amplitude resolution falls off). (d) Relative squared differences. Same as panel (c), but normalized by the amplitude of the GF for every medium realization. Given that the amplitude in the true GF before the arrival of the direct wave is zero, relative squared differences do not show the noise reduction provided by the optimization process in this range.

be explained by considering the theoretical evolution of correlation symmetry in a multiply scattered wavefield setting. Malcolm *et al.* (2004) derived a simple form for the temporal evolution of causal–acausal symmetry, which holds true if the wavefield can be approximated by the diffusion equation. In this case:

$$\Gamma(R, t) = \frac{1 + 3R/2vt}{1 - 3R/2vt}. \tag{3}$$

Here, Γ is the ratio between causal and acausal portions of the correlation, R is the distance between the sources and the midpoint between receivers and v is the wave velocity (which in practice is variable depending on modes and frequency). Importantly, we note here that the evolution of correlation symmetry does not depend on the strength of the scattering once the field is diffuse. This expression is however highly simplified and assumes an already diffuse field. In practice, the slight delay for stronger scattering models can be explained by considering the general evolution of the wavefield

before this state is globally attained. A discrete source in a scattering medium will generate direct and single scattered waves followed by a diffusive multiply scattered halo. The propagation velocity of this halo decreases as a function of increasing scattering, thus resulting in a delayed full spatial sampling of the medium by this halo. As such, the slight delay of total averaged PPDs for stronger scattering models is well accounted for. It should be noted that these trends would not necessarily persist in real world data, given that attenuation is typically related to scattering, and full medium sampling is naturally limited.

We are further interested in how well, on average, the optimized cross-correlations fare with respect to standard cross-correlations when compared to the true GF. For a randomly generated medium, we calculate the squared differences between the derivative of the cross-correlations (MCMC optimized, multi-window standard, single window standard) and the GF (as in eq. (1), defined by $G(A, B, t) - G(B, A, -t)$). These differences are then averaged over 90

medium realizations. Figs 4(c) and (d) show the squared differences between the cross-correlations (single window, multiwindow, and optimized) and the GFs averaged over 90 random media realizations featuring 128 scatterers. It is immediately evident that the window optimization yields solutions that are much closer to the empirical GF. For real events in strongly scattering media, such as the icequakes described below, the amount of coda available is limited to 25 s at most, meaning that the standard approach will at best have 2–3 windows to stack per event. The best standard correlations for icequake data are therefore likely closer to the single window example in Figs 4(c) and (d) than the multiple window stack. It should therefore not be surprising that the standard correlation estimates are far from symmetric for real data, and show substantial differences with respect to the optimized solutions.

In theory, given that any correlation window in the diffuse coda should converge directly towards the GF, it is only possible to offer intuitive arguments pertaining to the convergence towards a specific range of time windows defined by t_i. Our optimization process converges to solutions that are both symmetric and coherent over MCF rows, which are characteristics of the true GF if changes in the medium are minimal. These characteristics are typically destroyed by low signal to noise ratios or an insufficiently scattered wavefield, and it is not surprising that numerical experiments yield the same solution regardless of the chosen initial guess, since a window range

that trades off multiple scattering and decay is ideally sought out. It can be noted that simply choosing the maximum of each PPD for each source (as pictured in Fig. 4a) as a single window correlation does not yield as good a correlation gather as the stack of the post burn-in solutions. Though the PPD yields an estimate of where most solutions fall in the coda of each event, it would seem that the most coherent of correlation gathers rely on subtle interdependencies between event solutions, and not just arbitrary windows chosen in an ideal range.

Interestingly, the width of the PPD averaged over all sources and 90 media is very nearly identical for scenarios with 64 and 128 scatterers (Fig. 4b). This would indicate that overall, the strength of the scattering in these simulations does not affect the time window sampling of accepted models beyond the small shift to later times for stronger scattering. The widths of the PPDs are therefore largely dependent on optimization parameters.

5 ICEQUAKES ON EREBUS VOLCANO

Fig. 5 shows examples of MCMC processing for stations on Erebus volcano, where we used either a month-long database of 3318 icequakes recorded on 94 broadband and short period stations, or a year-long database of 21 000+ icequakes recorded on
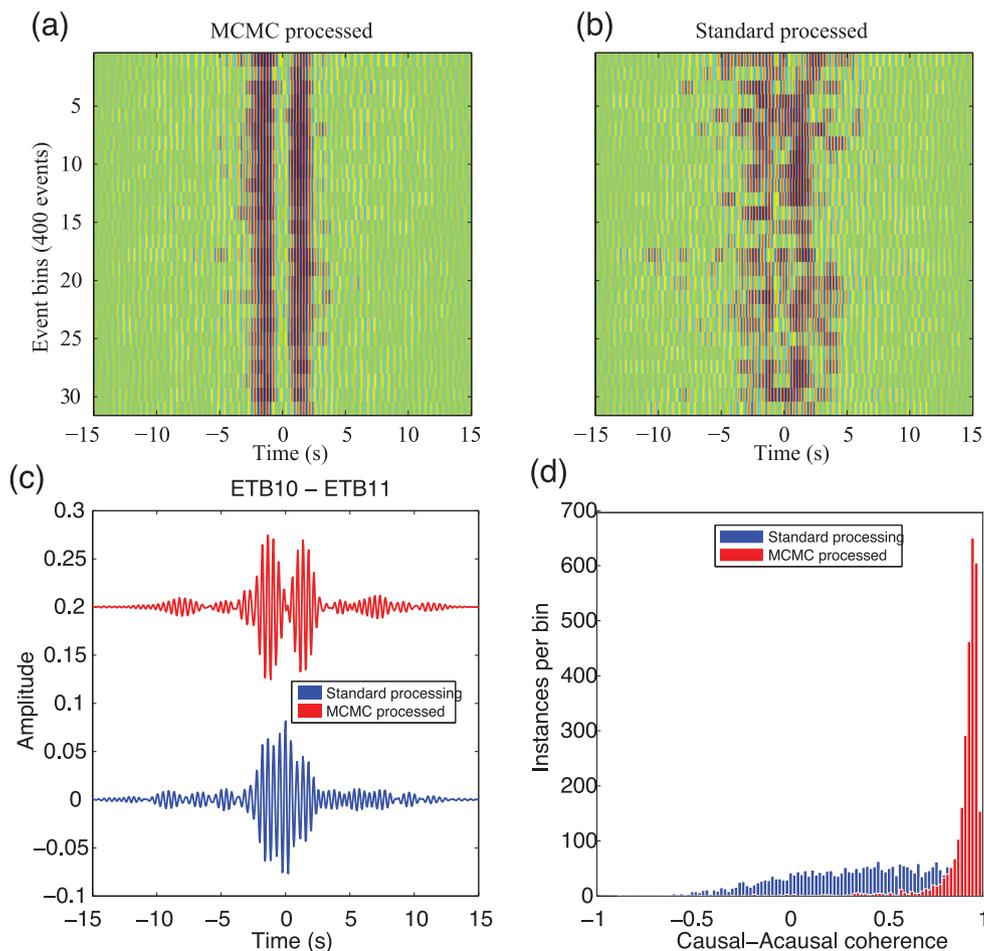

Figure 5. Erebus icequake correlations. (a) Example post burn-in binned correlation matrix for a 1 yr database with 21 000+ events, stations ETB10-ETB11. (b) Matrix MCF binned by 400 correlations built from an initial arbitrary window selection. (c) Example correlation stacks for MCMC and standard multi-window stack using a database of 21 000+ icequakes recorded over 1 yr. (d) Overall comparison in causal–acausal coherence featuring 2850 pairs of stations for MCMC and standard processing using a small database of 3318 icequakes.

30 broad-band stations in the same manner as the numerical examples described above. For both cases, we allow for 60 000 iterations, though the burn-in periods for these numbers of events are typically under 10 000 and 25 000 iterations, respectively. Given the relatively small number of sources, the varying nature of the source signatures and locations, and the limited amount of coda, differences between the standard correlations and the MCMC stacks are quite substantial in most cases, though some standard correlations yield good initial results as well. In all cases tested, there is an improvement of the coherence over event-time and an increase in causal–acausal symmetry. Fig. 5(a) displays an example of a post burn-in MCF matrix binned for sets of 400 cross-correlations for two ETB (21 000+ events) stations. Fig. 5(b) shows the associated initial MCF matrix also binned by 400 events, computed by taking a best guess single window correlation. Note the remarkable improvement in coherence over sequential bins, as well as the net improvement in causal–acausal symmetry of this matrix. Fig. 5(c) shows the mean of the PPD (mean of Fig. 5 a for many models sampled from the PPD) compared to the corresponding multi-window correlation computed for this station pair. Fig. 5(d) shows the histogram of causal–acausal coherences for standard correlations (Chaput *et al.* 2015) versus MCMC results for 2850 interstation pairs using 3318 icequakes recorded over one month.

For icequake correlations, we have no *a priori* information on the GF. We therefore choose two somewhat qualitative approaches to evaluate the GF estimated through both MCMC and standard processes. We first select a series of lines from the array of 94 stations and seek out the presence of coherent moveout signatures. We perform a rough moveout test by binning the correlation functions for similar interstation distances, effectively stacking correlation functions to enhance contributions from potential linear reflectors along the strike of the line (such as large surface crater features, or flat layers underlying the line). Naturally, on the volcanic edifice, dominant reflectors are likely to be of arbitrary shape and size, but in the very least, it allows us to evaluate the coherent moveout of the direct waves, and observe potential coherent reflection patterns.

Fig. 6(b) shows comparative examples of line gathers depicted by Fig. 6(a). It is immediately apparent that both the Rayleigh wave and scattered portions of the correlation function show increased symmetry in the MCMC correlations, and there are emerging coherent moveout patterns for several of the lines. Naturally, this simple assumption results in incorrect stacking for reflections not along strike, but it is sufficient from a qualitative standpoint.

The highly variable topography of Erebus volcano also allows us to perform a further test. Considering the amplitude and timing of many of the scattered arrivals displayed in Fig. 6(b), we hypothesize that at least part of this energy corresponds to surface waves reflected from pronounced topography on the volcanic edifice. Given the total decay of the Rayleigh wave for interstation distances exceeding 3 kms (Chaput *et al.* 2015) and the relatively short correlation windows available, resulting in limited correlation timings that could contribute to scattered surface wave modes, we use all station pairs with interstation distances shorter than 2.5 km, and repropagate the absolute envelope of their correlation functions assuming a purely Rayleigh wave composition. We first sum the causal and acausal portions of the correlation functions, remove the entirety of the direct interstation Rayleigh wave energy for a given pair (as estimated from the lowest observed velocities) plus an additional 0.5 s to avoid the mapping of direct waves as scattering events, and then compensate for cylindrical geometric spreading. Note that the same decay factor is removed from both the standard and MCMC correlation functions. Using a bulk Rayleigh wave velocity of 1 km/s (Chaput *et al.* 2015), we then generate a travel distance ellipse for every time sample of a given station pair's correlation function weighted by the amplitude at that sample. Finally, we create an averaged interpolated map of surface wave scattering strength for the volcanic upper edifice as depicted by Fig. 7.

The most striking feature of the MCMC correlation map is the emergence of coherent scattering from the nearest older upper caldera rim (crater features denoted by black dotted line), only slightly represented by standard correlations. This poor reconstruction could be due to imperfect symmetry in the standard correlations, where the averaging over the causal and acausal portions results in largely diminished amplitudes (not the case for MCMC correlations). The large coherent scattering feature visible near the top right of scattering maps, visible in both processing methods, approximately corresponds to a large surface expression of an even older caldera rim (outer black dotted line), named Fang Ridge (see Fig. 2). Other bright spots enhanced by the MCMC approach also map fairly well to particularly large surface expressions of topographic gradient, as is the case near station CON, where there is a pronounced ridge followed by a steep slope. This particular feature is also somewhat resolved by standard correlations, pointing to the strength of this particular scatterer. The innermost crater (near station RAY in Fig. 7) hosting the present day lava lake is also somewhat enhanced, though much of the close range scattering is obliterated during the removal of the direct Rayleigh wave. The southern face of the volcanic edifice, featuring the most prominent scattering in the MCMC resolved map, and to a lesser extent, the standard map, presents the highest topographic gradient of the upper edifice, and is clearly well represented here. Note that there are also a plethora of potentially mismapped arrivals, including later bands of strong scattering between older crater features, and scattering to the right of the present-day crater. Many scattered arrivals in the correlation functions could very well be due to reflected body waves from strong scatterers at depth, and such shallow structures have already been mapped out in previous studies (Chaput *et al.* 2012; Zandomeneghi *et al.* 2013). Furthermore, due to the natural decay of the correlation function (given relatively short time windows considered here), it is perhaps not surprising that the outermost crater is only somewhat represented (in the general area of Fang Ridge). Nonetheless, we present this as practical evidence of improved reconstruction of the scattered portion of the GF over traditional processing of correlation functions.

6 DISCUSSION

The type of processing shown here presents a variety of implications. The GF resulting from optimized windows is clearly superior overall in simulations, which is useful for passive imaging techniques, particularly where stations are sparse. The recovered PPDs further offer additional information about the nature of the wavefield, and potentially the nature of the medium's scattering strength. The slight delay on PPDs estimated in more scattering media (128 versus 64 scatterers) very likely results from a counterintuitive trade-off between how scattered the wavefield is, and the pervasiveness of this wavefield within the medium. Though theoretical predictions requiring an already diffuse wavefield suggest that a higher degree of scattering results in a better GF reconstruction, discrete sources in practice generate a combination of direct waves followed by a diffuse energy halo that progressively samples the medium. The propagation of this halo is slowed by increased scattering strength, resulting in a delayed sampling of the medium as a whole, and a
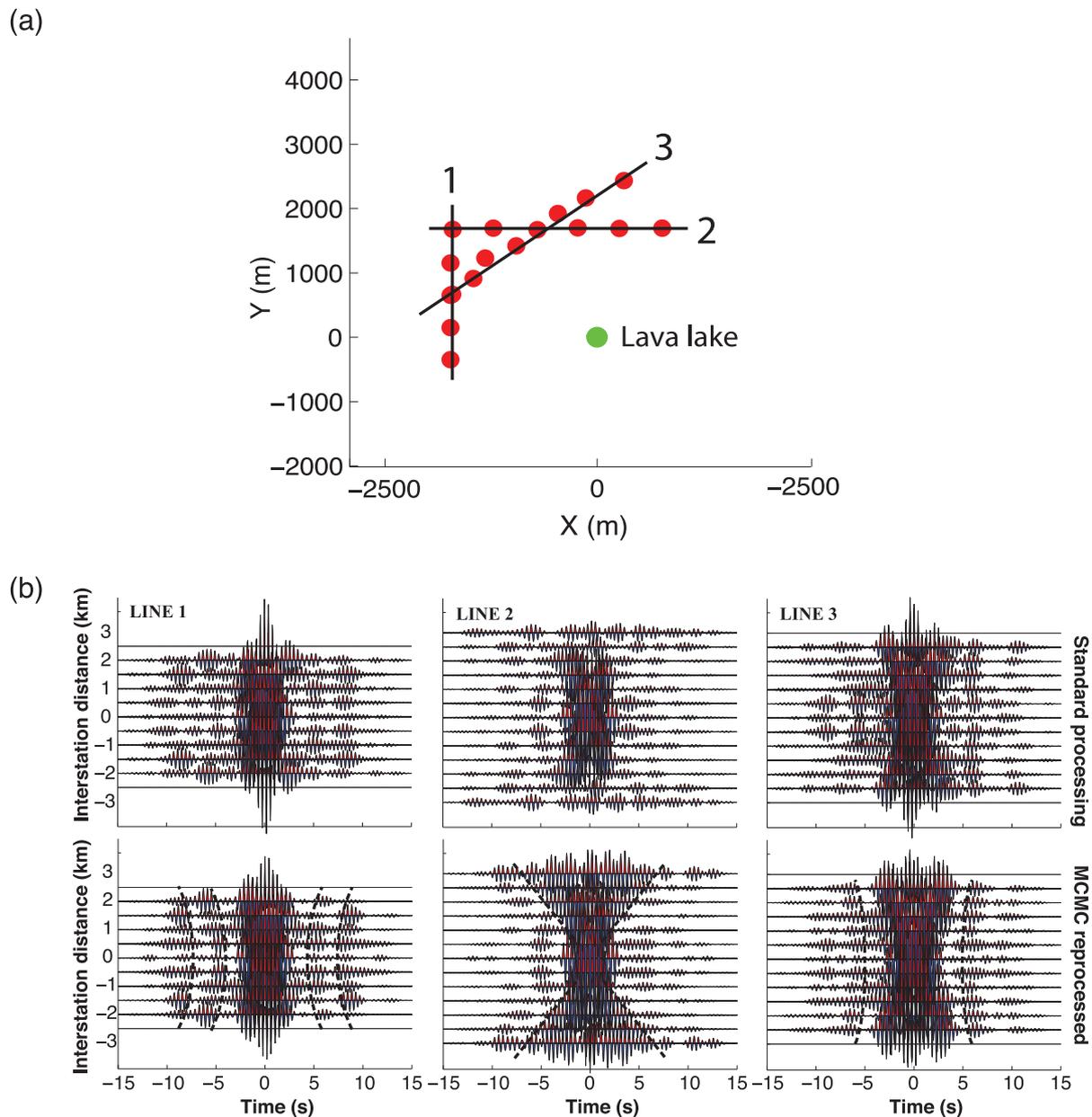
Figure 6. Demonstration of scattering reconstruction for icequakes on Erebus volcano. (a) Example station lines chosen from the dense array of 94 stations. (b) Corresponding correlation gathers. Top panels correspond to standard processed correlations, and bottom panels correspond to MCMC processed correlations. Cross-correlations are binned by interstation distances, effectively enhancing contributions from potential flat layering under the line, or large linear surface reflectors along strike of the line. Reconstructed symmetric arrivals with prominent moveouts are indicated with dotted lines.

consequent delay in the convergence to symmetry in the correlation function of this wavefield. Furthermore, the optimization parameters chosen here suggest that correlation symmetry has a stronger influence on the PPD than the coherence over event-time, thus further explaining this delay. An aspect of MCMC solutions not explored here is the spatial dependence of the individual source PPD maxima (i.e. the mapping of individual PPD maxima to their source locations). Such mapping could be representative of the distribution of scatterers, though more research is needed.

For icequake data, the exact time window solutions are less informative, given the variations in source amplitude and phase signatures, imprecise first arrival picks, and non-uniform distributions. The net result of the solution stack however, as demonstrated clearly

by Fig. 5 and numerical results, yields a much improved estimate of a given receiver pair's GF. The simple surface wave reflection assumption, explored in Figs 6 and 7, also clearly indicates that at least part of the recovered coda energy consists of singly scattered surface waves from older crater features. The circular shape of these craters around the station array likely results in strong focusing, which would help explain why these particular features dominate the early coda in the MCMC reconstructed correlations. It should be noted however that small scale scattering is of particular importance in the mixing of wavefield directionality, and the direct mapping of the corresponding scatterers remains a subject of much interest.

An interesting side effect of the optimization approach, even where the final stacked solution is relatively unchanged, is the vast
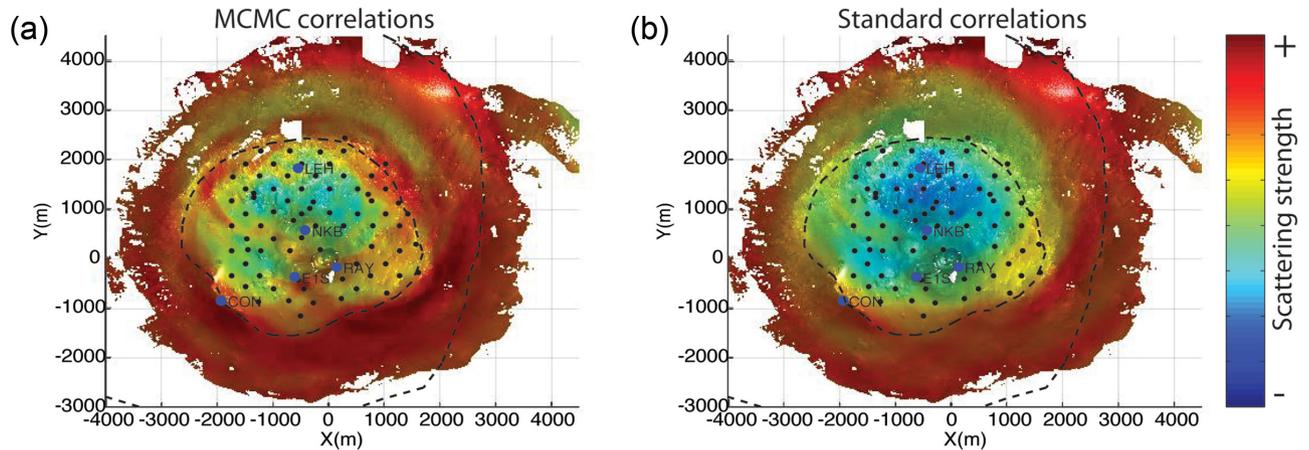
Figure 7. Elliptically repropagated correlation functions for both MCMC (a) and standard (b) processed correlations. Warmer colours correspond to coherently stacked energy over all station pairs. Dotted lines correspond to known older crater features, and are largely representative of rapid increases in slope gradient.

increase in the event to event coherence (Fig. 5a), yielding much smoother correlation gathers without necessitating post-processing techniques (F-K smoothing, wavelet/curvelet transform). Whether this particular increase in coherence can be used to reduce noise but preserve real effects of temporal variability is the subject of ongoing research. However, tests featuring particularly asymmetric heuristic correlations (not shown) often do not converge toward amplitude symmetry when subjected to MCMC processing. That the algorithm doesn't ensure perfect symmetry is not troubling, given the finite length of the coda used and the extreme case of scattering geometry chosen here. It is simply a reflection of the best possible coda window range given the nature of the recorded wavefield, and a demonstration of the algorithm's inability to artificially create data from coherent stacking of random fluctuations. In a rapidly changing medium, it may further be possible to implement a sequential MCMC scheme that would allow for updating the correlation gather with any new information in a computationally efficient manner (Dettmer *et al.* 2011) while ensuring that a maximum of the real temporal variations are preserved.

It should be noted that for the specific application described here, certain prerequisites are desirable. First, the correlation function theoretically converges towards the GF in media capable to hosting of near equipartitioned wavefields. Alternately, the equipartition criterion is interchangeable with a homogeneous source distribution around the medium of interest (particularly rare in real media). The criteria in the objective function (i.e. symmetry and coherence over events, see Fig. 1) would however not apply to this second case, given that we would only expect the correlation function to converge after the stack of all events. Furthermore, it is desirable to have as much coda as possible to use for any event, and thus a medium in which the scattering mean free path is much shorter than the absorption mean free path should be considered optimal. This is certainly the case for numerical model, and is also the case for Erebus volcano (Chaput *et al.* 2015).

The implications of these improved GF estimates are diverse. First, next generation methods aiming to separate the singly scattered from the multiply scattered wavefield for imaging purposes (Shahjahan *et al.* 2013), rely necessarily on very accurate impulse responses to extract subtle coherence signatures along receiver lines. Such an approach could become commonplace in highly scattering media appropriately sampled by local events, if recorded coda are sufficiently diffuse to allow for window tuning. Second, works by Stehly *et al.* (2008) have shown that the scattered component of the

GF is at least partially recovered in the correlation of the coda of ambient noise correlations (C3). The optimization approach described here, somewhat less intuitive for raw ambient noise correlations, could however be applied to the computation of these C3 functions from ambient noise correlations. Resulting improved GF recovery from virtual source coda (the basis of C3) could permit the use of scattering based methods in conjunction with ambient noise signals (i.e. robust reconstruction of scattered phases instead of merely the direct waves), thus far only accomplished for bulk earth phases and fault trapped signals (Boue *et al.* 2014; Hillers *et al.* 2014; Poli *et al.* 2012b). Furthermore, the advent of ambient noise based methodologies for temporal monitoring has led to the development of coda sensitivity kernels aiming to constrain the general region in which change is occurring (Obermann *et al.* 2013a,b; Planes *et al.* 2015; Pacheco & Snieder 2005). Ideally, it would be much more useful to map these changes to known reflection-based substructure, particularly in the case of active faults or volcanic media. The great improvement of event-time coherence implies that correlations can converge much more rapidly towards the true GF, or in the very least towards stability, than with the standard correlation approach. The event convergence limits of this method have not been explored here, but MCMC processing necessarily results in better coherence over time regardless of the event-binning chosen, and thus the temporal resolution for monitoring purposes could be greatly increased.

7 CONCLUSIONS

We have presented a window optimization scheme by which we are able to greatly improve interstation GF estimates from multiply scattered coda wavefields. Though this optimization can be solved by various means, we approach the problem with a MCMC algorithm that provides us with additional information concerning the individual event distributions of solutions. Numerical experiments have demonstrated an overall better convergence towards the empirical GF when compared with several standard coda correlation approaches (single- or multiwindow stacking). The model space is evidently well explored by the MCMC approach, given that various prior estimates converge to the same general solution range. Coda time window solutions were averaged over many random synthetic media, and it is noted that the density of scatterers has a small effect on the average window solution in the posterior distribution, though

these effects may be specific to the nature of the synthetic media used here. The stack of windows solutions after the burn-in period also results in a much better GF estimate than recovered via standard means. We further applied this approach to icequake data from Erebus volcano, where great increases in event-time coherence and correlation symmetry were observed. These gains were persistent for data sets spanning one month through to 1 yr. As suggested by the numerical experiments, the scattered portion of the correlation function appears to be more representative of the real GF. This was demonstrated through effectively mapping out known surface structures by back-propagating the correlation functions while assuming single surface wave scatter modes. The increase in event-time coherence also decreases the number of events necessary for correlation convergence, implying that resolution gains in temporal monitoring experiments could be made this way.

ACKNOWLEDGEMENTS

REFERENCES

Aster, R. *et al.*, 2004. Real-time data received from Mount Erebus volcano, Antarctica, *EOS, Trans. Am. geophys. Un.,* **85,** 97, 100–101.

Boue, P., Poli, P. & Campillo, M., 2014. Reverberations, coda waves and ambient noise: correlations at the global scale and retrieval of the deep earth phases, *Earth planet. Sci. Lett.,* **391,** 137–145, doi:10.1016/j.epsl.2014.01.047.

Campillo, M. & Paul, A., 2003. Long-range correlations in the diffuse seismic coda, *Science,* **299,** 547–549.

Catheline, S., Gallot, T., Roux, P., Ribay Guillemette, R. & de Rosny, J., 2011. Coherent backscattering enhancement in cavities: The simple-shape cavity revisited, *Wave Motion,* **48**(3), 214–222.

Chaput, J., Zandomeneghi, D., Aster, R., Knox, H. & Kyle, P.R., 2012. Imaging of Erebus volcano using body wave seismic interferometry of Strombolian eruption coda, *Geophys. Res. Lett.,* **39**(7), doi:10.1029/2012GL050956.

Chaput, J., Campillo, M., Aster, R.C., Roux, P., Kyle, P.R., Knox, H. & Czoski, P., 2015. Multiple scattering from icequakes at Erebus volcano, Antarctica: Implications for imaging at glaciated volcanoes, *J. geophys. Res.,* **120**(2), 1129–1141.

Colombi, A., Boschi, L., Roux, P. & Campillo, M., 2014. Green's function retrieval through cross-correlations in a two-dimensional complex reverberating medium, *J. acoust. Soc. Am.,* **135**(3), 1034, doi:10.1121/1.4864485.

Davy, M., Fink, M. & de Rosny, J., 2013. Green's function retrieval and passive imaging from correlations of wideband thermal radiation, *Phys. Rev. Lett.,* **110,** 203901, doi:10.1103/PhysRevLett.110.203901.

De Rosny, J. & Roux, P., 2001. Multiple scattering in a reflecting cavity: Application to fish counting in a tank, *J. acoust. Soc. Am.,* **109**(6), 2587–2597.

Dettmer, J., Dosso, S. & Holland, C.V., 2011. Sequential trans-dimensional Monte Carlo for range-dependent geoacoustical inversion, *J. acoust. Soc. Am.,* **129**(4), 1794–1806.

Draganov, D., Campman, X., Thorbecke, J., Verdel, A. & Wapenaar, K., 2009. Reflection images from ambient seismic noise, *Geophysics,* **74,** A63–A67.

Hennino, R., Tregoures., N., Shapiro., N., Margerin., L., Campillo., M., van Tiggelen, V. & Weaver, R., 2001. Observation of equipartition of seismic waves, *Phys. Rev. Lett.,* **86**(15), 3447–3450.

Hillers, G., Campillo, M., Ben-Zion, Y. & Roux, P., 2014. Seismic fault zone trapped noise, *J. geophs. Res.,* **119,** doi:10.1002/2014JB011217.

Knox, H., 2012. Eruptive characteristics and glacial earthquake investigation on Erebus volcano, Antarctica, *PhD dissertation,* New Mexico Institute of Mining and Technology.

Larose, E., Roux, P. & Campillo, M., 2007. Reconstruction of Rayleigh-Lamb dispersion spectrum based on noise obtained from an air-jet forcing, *J. acoust. Soc. Am.,* **122**(6), 3437, doi:10.1121/1.2799913.

Liu, Q. & Tao, J., 1997. Perfectly matched layers for acoustic waves in viscous media: Application to underwater acoustics, *J. acoust. Soc. Am.,* **101**(5), 3182–3183.

Lobkis, O. & Weaver, R., 2001. On the emergence of the Green's function in the correlations of a diffuse field, *J. acoust. Soc. Am.,* **110,** 3011–3017.

Malcolm, A.E., Scales, J. & van Tiggelen, B., 2004. Extracting the Green function from diffuse, equipartitioned waves, *Phys. Rev. E,* **70,** 015601, doi:10.1103/PhysRevE.70.015601.

Margerin, L. & Nolet, G., 2003. Multiple scattering of high-frequency seismic waves in the deep earth: modeling and numerical examples, *J. geophys. Res.,* **108**(B5), 2234–2249.

Margerin, L., Campillo., M. & Von Tiggelen, B., 2000. Monte Carlo simulation of multiple scattering of elastic waves, *J. geophys. Res.,* **105,** 7873–7892.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100**(B7), 12 431–12 447.

Obermann, A., Planes, T.A., Larose, E., Sens-Schonfelder, C. & Campillo, M., 2013a. Depth sensitivity of seismic coda waves to velocity perturbations in an elastic heterogernous medium, *Geophys. J. Int.,* doi:10.1093/gji/ggt043.

Obermann, A., Planes, T.A., Larose, E. & Campillo, M., 2013b. Imaging preeruptive and coeruptive structural and mechanical changes of a volcano with ambient noise, *J. Geophs. Res.,* **118,** 6285–6294.

Pacheco, C. & Snieder. R. 2005. Time-lapse travel time change of multiply scattered acoustic waves, *J. acoust. Soc. Am.,* **118,** 1300–1310.

Paul, A., Campillo, M., Margerin, L. & Larose, E., 2005. Empirical synthesis of time-asymmetrical Green functions from the correlation of coda waves, *J. geophys. Res.,* **110,** B08302, doi:10.1029/2004JB003521.

Planes, T., Larose, E., Rossetto, V. & Margerin, L., 2015. Imaging multiple local changes in heterogeneous media with diffuse waves, *J. acoust. Soc. Am.,* **137,** 660, doi:10.1121/1.4906824.

Poli, P., Pedersen, A. & Campillo. M. 2012a. Emergence of body waves from cross-correlation of short-period seismic noise, *Geophys. J. Int.,* **188,** 549–558.

Poli, P., Campillo, M. & Pedersen, H.A., 2012b. Body wave imaging of Earth's mantle discontinuities from ambient seismic noise, *Science,* **338**(6110), 1063–1065.

Roux, P., Sabra, K.G., Kuperman, W.A. & Roux, A., 2005. Ambient noise cross correlation in free space: theoretical approach, *J. acoust. Soc. Am.,* **117**(1), 79–84.

Sabra, K.G., Gerstoft, P., Roux, P. & Kuperman, W.A., 2005. Extracting time-domain Green's function estimates from ambient seismic noise, *Geophys. Res. Lett.,* **32,** L03310, doi:10.1029/2004GL021862.

Sanchez-Sesma, F. & Campillo, M., 2006. Retrieval of the Green's function from cross correlation: The canonical elastic problem, *Bull. seism. Soc. Am.,* **96**(3), 1182–1191.

Sato, H., 2009. Green's function retrieval from the CCF of coda waves in a scattering medium, *Geophys. J. Int.,* **179,** 1580–1583.

Shahjahan, S., Aubry, A., Rupin, B., Chassignole, A. & Derode, A., 2013. A randrom matrix approach to detect defects in a strongly scattering polycrystal; how the memory effect can help overcome multiple scattering, *Appl. Phys. Lett.,* **104,** 234105, doi:10.1063/1.4882421.

Shapiro, N.M., Campillo, M., Margerin, L., Singh, S.K., Kostoglodov, V. & Pacheco J. 2000. The energy partitioning and the diffusive character of the seismic coda, *Bull. seism. Soc. Am.,* **90,** 655–665.

Shapiro, N. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geop. Res. Lett.,* **31,** L07614, doi:10.1029/2004GL019491.

Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase, *Phys. Rev. E,* **69,** 046610, doi:10.1103/PhysRevE.69.046610.

Stehly, L., Campillo, M., roment, B. & Weaver, R.L., 2008. Reconstructing Green's function by correlation of the coda of the correlation (C3) of ambient seismic noise, *J. geophys. Res.,* **113,** B11306, doi:10.1029/2008JB005693.

Tonegawa, T. & Nishida, K., 2010. Inter-source body wave propagations derived from seismic interferometry, *Geophys. J. Int.,* **183,** 861–868.

Tonegawa, T., Nishida., K., Watanabe., T. & Shiomi, K., 2009. Seismic interferometry of teleseismic S-wave coda for retrieval of body waves: an application to the Philippine Sea slab underneath the Japanese Islands, *Geophys. J. Int.,* **178,** 1574–1586.

Wapenaar, K. & Fokkema, J., 2006. Green's function representations for seismic interferometry, *Geophysics,* **71,** 33–46.

Weaver, R. & Lobkis, O., 2004. Diffuse fields in open systems and the emergence of the Green's function, *J. acoust. Soc. Am.,* **116,** 2731–2734.

Zandomeneghi, D. *et al.*, 2013. Internal structure of Erebus volcano, Antarctica imaged by high-resolution active-source seismic tomography and coda interferometry, *J. geophys. Res.,* **118**(3), 1067–1078.

APPENDIX A

Bayes theorem in its most general form states that

$$P\left(m|d\right) = \frac{P\left(m\right)P\left(d|m\right)}{P\left(d\right)} = \frac{P\left(m\right)P\left(d|m\right)}{\int_M P\left(m_i\right)P\left(d|m_i\right)d_i'}. \qquad \text{(A1)}$$

The left-hand side constitutes the PPD, representing the probability of the model given the data information. The term $P(d|m)$ is interpreted as the likelihood L for fixed data d, and $P(m)$ is the prior information known about the model. $P(d)$ is often referred to as the Bayesian evidence, and is the probability of the data independent of m, and is furthermore equivalent to an integral over the model space M. Metropolis–Hastings sampling for an MCMC approach is typically used to sample the PPD, with the general acceptance term for such sampling being defined as

$$A\left(m'|m\right) = min\left[1, \frac{Q\left(m|m'\right)}{Q\left(m'|m\right)}, \frac{P\left(m'\right)}{P\left(m\right)}, \frac{L\left(m'\right)}{L\left(m\right)}\right]. \qquad \text{(A2)}$$

Here, $Q(m'|m)$ is defined as the proposal distribution from which perturbation parameters are drawn to shift from the current model m to the new model m'. The new model is evaluated by drawing a random number from a uniform distribution on [0,1], and accepting the perturbed model if this number is smaller than $A(m'|m)$. For our particular experiment, we wish to find combinations of time windows in the coda of a number of events that yield the most coherent and symmetric matrix of cross-correlation gathers possible. Given that we have prior knowledge concerning the average coda range (for icequakes) in which the wavefield is modally partitioned (Chaput *et al.* 2015), we define priors P as uniform distributions around a heuristic estimate of the ideal coda window. For icequakes, this estimate is roughly 11 s after the event onset, and we set 11 ± 9 s as the uniform bounds for each event to ensure that the correlation remains representative of icequake coda and not background noise or ballistic waves. Proposal distributions Q are typically set as Gaussian distributions around current model values with standard deviations that can be tuned to allow for more efficient convergence towards the PPD. Though we do not provide a full analysis of convergence parameters here, a standard deviation of 1/30 times the width of the prior bounds allowed for adequate run times. For synthetic tests, we simply set the prior as an arbitrary window in the coda, and allow the full range of the time series to be used given the absence of signal other than the scattered wavefield. Here, given the fixed number of parameters (i.e. number of total events used), the uniformly bounded prior distributions and the symmetric proposal distributions, the Metropolis-Hastings acceptance defined in eq. (2) simplifies to

$$A\left(m'|m\right) = min\left[1, \frac{L\left(m'\right)}{L\left(m\right)}\right]. \qquad \text{(A3)}$$

As such, though the prior and proposal distributions play a role in fine tuning the efficiency of the algorithm, they play no part in the acceptance probability for each individual iteration. Various forms of the likelihood L can be used. Note that the acceptance function depends solely on the ratio of likelihoods between model steps, and is here defined as

$$P_{\text{accept}} = \exp\left(-\left(\frac{\delta_{\text{misf}}}{F^2}\right)\right). \qquad \text{(A4)}$$

Here, P_{accept} represents a true or false clause, and δ_{misf} is defined as $\text{MSF}_k - \text{MSF}_{k-1}$, where k is the iteration number. If a positive number chosen from a uniform distribution centred around zero is larger than the current value of P_{accept}, then the new model is rejected. Otherwise, it is kept as the updated model. F, which roughly represents the standard deviation of the data error, is adjusted until the acceptance ratio falls between 30–50 per cent, shown to be the ideal range in efficiency for this type of algorithm (Mosegaard & Tarantola 1995). A 'burn in' period is allocated for each run of the algorithm, where the misfit rapidly decreases until a state of relative stability is attained. Models generated during this burn in period are rejected in the evaluation of the posterior distribution of the model space.