Common conversion point stacking of receiver functions versus passive-source reverse time migration and wavefield regularization

Xuefeng Shang,[1],[*] Maarten V. de Hoop[2] and Robert D. van der Hilst[1]

[1]*Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology,* 77 *Massachusetts Avenue, Cambridge, MA* 02139, *USA.*
E-mail: xuefeng.shang@gmail.com
[2]*Department of Computational and Applied Mathematics, Rice University,* 2036 *Duncan Hall, Houston, TX* 77005, *USA*

SUMMARY

We demonstrate with synthetic and field data that with sufficiently dense sampling wave-equation-based methods such as reverse time migration (RTM), implicitly forming array receiver functions (ARFs), perform better resolution wise than migration of common conversion point (CCP) stacks of traditional receiver functions. However, even with modern array deployments the sampling requirement is typically not met for teleseismic (earthquake) data. To enable RTM imaging with sparsely (and irregularly) sampled wavefields at the surface, we use an intermediate reconstruction based on sparsity promoting optimization using a curvelet (or wave packet) representation of the data, as an important and necessary pre-processing step. To suppress artefacts, the curvelet coefficients are constrained to represent the range of known directions present in the data. We show that our proposed pre-processing procedure (which may be viewed as generating 'missing' traces) can produce artefact-free data for RTM even if only 20 per cent of necessary data are available in the original data set. With synthetic data, we also demonstrate that if the sampling criteria is not met, CCP can produce results that are superior over wave-equation methods such as RTM. As a proof-of-concept with field data, we image the structure of the crust beneath the Himalayas with passive-source RTM of teleseismic data from Hi-CLIMB project. For Hi-CLIMB data, the CCP and RTM results are similar because sampling is still too sparse for RTM and the structure is simple enough for successful CCP. Both results are improved by wavefield regularization and reveal that the Moho is continuous beneath most of the array, and not fragmented as suggested by some earlier studies.

Key words: Crustal imaging; Wave scattering and diffraction; Continental tectonics: compressional.
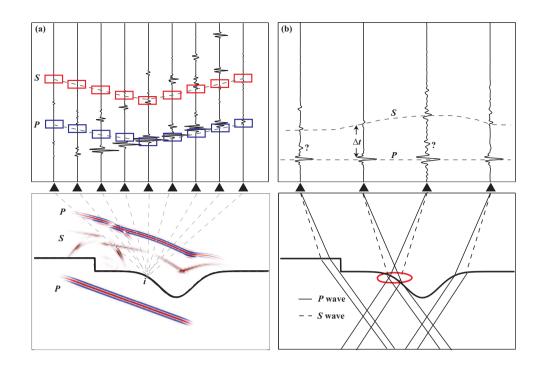
1 INTRODUCTION

In past decades, increasingly dense seismographic arrays—such as Hi-CLIMB in Tibet (Nábělek *et al.* 2005)—have been deployed with the aim of exploring finer structures in the crust and upper mantle. At some point, however, the underlying assumptions about the medium (and associated simplifications of wave propagation or inversion theories) begin to limit the length scales that can be resolved with traditional methods. For example, common conversion point (CCP) stacking is applied routinely in receiver functions studies (e.g. Dueker & Sheehan 1997; Gilbert *et al.* 2003), but the assumption that (at least locally) the interface is horizontal (Fig. 1b) degrades the imaging of geologically complex structures,

such as steep faults and laterally discontinuous interfaces. While reasonable for single station applications or (relatively) sparse sensor distributions such as in USArray (Meltzer *et al.* 1999), such simplifications will eventually diminish the scientific return of the investment in dense array deployments. Indeed, modern trends in data acquisition increasingly call for imaging methods, such as seismic migration/inversion, that exploit the full complexity of recorded wavefields and rely less on prior information about the Earth's structures.

As in reflection seismology, different types of migration methods can be applied to teleseismic studies, such as ray-based Kirchhoff migration (e.g. Ryberg & Weber 2000; Rondenay *et al.* 2001; Poppeliers & Pavlis 2003) and wave-based migration (e.g. Chen *et al.* 2005; Shang *et al.* 2012). In this paper, we apply the method of (wave-based) passive-source reverse time migration (RTM) developed by Shang *et al.* (2012) to teleseismic data of the Hi-CLIMB

* Now at: Shell Oil Company, Houston TX 77079, USA.

Figure 1. Schematic illustration of (a) passive-source reverse time migration and (b) common conversion point (CCP) stacking of traditional receiver functions. (a) Bottom: *P* wave impinging on an interface (solid black line) produces direct *P* and converted *S* waves, which are recorded by an array of seismographic stations at the surface (black triangles). Top: *P* and *S* energy due to scattering at, say, point *i* arrives at different times (blue and red boxes, respectively). In inverse sense, the location point *i* can be reconstructed by optimization of the correlation between the backprojected *P* and *S* wavefields. (b) Bottom: in traditional receiver functions, the *P*–*SV* conversion is assumed to occur at an interface that is (locally) horizontal. Data redundancy is obtained by stacking over CCP, shown as a red oval. Top: the receiver functions aligned with direct *P* wave. The traveltime difference Δt between transmitted *P* and converted *S* is a measure of interface depth, which is ambiguous at complex regions shown as question marks.

project in Tibet (Nábělek *et al.* 2005). It is worth noting that passive-source RTM requires only one elastic backward propagation to form an image rather than two (one forward from source side and one backward from data side) as in conventional RTM. Therefore source related uncertainties, such as in source location and origin time, are eliminated in teleseismic imaging of the structures beneath stations.

Teleseismic *P* (or *S*) waves can convert to other modes (*P*-to-*S* or *S*-to-*P*) when impinging from below on an interface or a scatterer, and the transmitted and converted waves then propagate to receivers on the Earth's surface (Fig. 1). In traditional receiver function analysis, the time difference between direct and converted phases is measured on records from individual stations and used to estimate the depth to a causative interface assuming that (at least locally) the latter is horizontal. In RTM, the relevant parts of the transmitted (e.g. direct *P*) and converted wave (e.g. *Ps*) fronts recorded at a seismograph array are backprojected (to be precise, reverse time continued) until they correlate at the (time and space) point of mode conversion. The conversion time (and location, for a given a reference velocity model) can be determined by applying a cross-correlation like imaging condition. Shang *et al.* (2012) developed passive-source RTM (which implicitly forms array receiver functions, ARFs) and demonstrated with synthetic data that RTM is superior over traditional receiver function analysis in the presence of strong topography on the interfaces being imaged.

Before providing a proof-of-concept application of wave-based RTM to data from the Hi-CLIMB project in Tibet

(Nábělek *et al.* 2005), we use synthetic tests to explore the influence of spatial aliasing and low signal-to-noise ratio (SNR), determine the parameters that control trace interpolation and subsequent inversion, and verify the viability of applying RTM to teleseismic data. We show that accurate crustal imaging with RTM requires a station spacing of the order of 2 km—an actual station spacing of 5–10 km (as in the Hi-CLIMB array) thus implies that some 50–80 per cent of the data needed for wave-based imaging must be estimated from the recorded wavefield. To prepare the Hi-CLIMB data for crustal imaging with wave-based RTM we must, therefore, resample the recorded wavefield on a sufficiently fine regular grid and interpolate it where the original data set is too sparse.

Hi-CLIMB comprises an 800-km-long, densely spaced seismic array that was deployed to investigate the structure of the lithosphere beneath the Himalayas and the southern Tibetan Plateau. A major aim of the project was to image the crust–mantle interface (hereinafter referred to as Moho) but studies with Hi-CLIMB data have yielded conflicting results. From receiver function analysis, Nábělek *et al.* (2009) argue that the Moho is continuous across the Himalayas into at least the Qiangtang Block in southern Tibet, whereas Nowack *et al.* (2010) present evidence from Gaussian beam migration for segmentation by Moho cutting faults. These structural differences have important implications for our understanding of continental deformation on a lithospheric scale across the most prominent zone of active continental collision. While station spacing is still sparser than ideal for wave-based methods, RTM can provide additional constraints on crustal structure—and help distinguish between the types of Moho structure previously proposed.
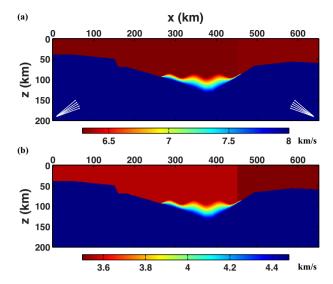
2 WAVEFIELD REGULARIZATION

Wavefield reconstruction algorithms usually exploit transforms to compress data in another domain, exploiting redundancy, such as the Fourier transform (e.g. Spitz 1991; Sacchi *et al.* 1998; Zwartjes & Sacchi 2007), the Radon transform (e.g. Kabir & Verschuur 1995; Trad *et al.* 2002) and the curvelet transform (e.g. Herrmann & Hennenfent 2008; Naghizadeh & Sacchi 2010a). Curvelets appear in solution constructions of the wave equation, revealing a degree of concentration depending on the smoothness of the wave speeds (Smith 1998; Candès & Demanet 2005; Andersson *et al.* 2008). For detailed implementation and synthetic data experiments, we refer to Appendix and references therein.

In seismic data reconstruction, we can consider the observed data **d** as a subset of data recorded continuously and ubiquitously at Earth's surface, or, more practically, as a subset of desired interpolated data **m**, so that $\mathbf{d} = \mathbf{Gm} + \mathbf{n}$, with **G** a sampling operator and **n** additive noise. From **m** one can then reconstruct any subset of data for imaging, including traces that were not available in—or 'missing' from—the original data set. This motivates an inversion problem aimed at estimating **m** from observations **d**—see Appendix for a more detailed formulation. As in any underdetermined inversion, uneven sampling and the presence of noise require regularization (or damping) and cause trade-offs between data misfit and model roughness. We note that **m** relates to curvelet coefficients **x** according to $\mathbf{m} = \mathbf{C}^T\mathbf{x}$, with $\mathbf{C}^T$ the adjoint of curvelet transform operator **C** (Candès & Donoho 2004), which can be found by minimizing, with some vector norm, the difference between **d** and **Ax**, with $\mathbf{A} = \mathbf{GC}^T$. For details, see Appendix.

We recover 'missing' traces in the curvelet domain (Figs A1 and A2) by solving a series of Lasso problems (Appendix). Upon transformation, the set of curvelet coefficient **x** is controlled by the range of the single scattering operator (i.e. it is restricted to represent directions present in the data), and in the inversion a damping parameter is used to control the trade-off between data misfit and the solution sparsity (Appendix). By exploring the *L*-curve explicitly (Fig. A3), as in classic Tikhonov regularization analysis, the optimal damping parameter is obtained for a given sampling geometry and data noise level. We show (Fig. A4) that for our purposes RTM can be performed successfully even if the observed data **d** contains only 15 per cent of total traces in **m** (with **m**, through the above relationship, informed by **d**). Wavefield regularization and trace interpolation is robust, see data example below, but the ability to recover fine-scale structure in the data depends (of course) on sampling density and noise level. The main artefacts are randomly oriented fine-scale curvelets within acquisition gaps, but these can be suppressed by stacking over different sources.

We tested the performance of RTM without and with wavefield regularization with synthetic data produced by 2-D finite-difference wave simulation. The RTM results are compared to results of traditional CCP stacking. In a first series of tests, RTM and CCP are applied without wavefield regularization. In a second series of tests, RTM is performed on data after curvelet regularization.

The test model (Fig. 2) was designed with application to Hi-CLIMB data in mind and is used to tune the parameters needed to recover essential features. The model comprises two main layers, which could represent crust and mantle, separated from each other by an interface, say the 'Moho'. In the middle part of this section the synthetic Moho is disrupted and the velocity increases gradually from shallow to deep layer. Furthermore, the shallow layer consists of two blocks (with boundary at 450 km horizontal distance) in order

Figure 2. Synthetic models for (a) *P* and (b) *S* wave speeds. There are generally two layers in the models, crust (red) and mantle (blue). Two blocks are presented in the crust, separated at 450 km horizontal distance. In the middle part (from 250 to 450 km), a transitional zone is added between the crust and mantle, in which the velocity increases gradually with depth. Ten events with plane wave incidence are simulated, and the incident angle (with respect to the vertical axis) is from 20° to 40° with an increment of 5°. Five events are injected from the left-hand side, and others are from the right-hand side (white arrows in (a)).

to represent different geological blocks in Tibet, and a depth offset (of 20 km) of the synthetic Moho is introduced at 150 km horizontal distance.

The synthetic data are generated by a finite-difference scheme. Ten localized plane *P* waves are injected in the mantle as incident waves, five of which are from the left-hand side and others from the right-hand side (Fig. 2a). The incident angle with vertical axis is from 20° to 40° with 5° increment. The corresponding epicentral distance varies from 30° to 90°. A Ricker wavelet is used as the source time function and the central frequency is 0.5 Hz (the maximum frequency ~1 Hz). For simplicity but without loss of generality, absorbing boundary condition is applied in the forward simulation, and there are only direct *P* wave and converted *S* waves in the data without surface related multiples. In order to investigate the effect of receiver spacing, the recorded wavefield is randomly sampled along the spatial coordinate with an average interval of 2, 6, 10 and 20 km.

CCP stacks of data from 10 events with different station spacing (from 2 to 20 km) are shown in Fig. 3. Average 1-D wave speed models are used as the background models for ray tracing. The relatively flat part of the synthetic Moho (i.e. at 0–100 km and 500–600 km horizontally) is well imaged even with 20 km spacing samplings (Fig. 3d). The image quality of the dipping Moho (100–300 and 400–500 km) gradually improves as the receiver spacing decreases; see, for instance, the diminishing staircases from Figs 3(d) to (c) to (b). The improvement becomes trivial at certain point, and Figs 3(a) and (b) are almost identical. Even after such convergence, however, neither the steeply dipping structures nor the details of the disrupted Moho are well imaged.

We then conduct passive-source RTM (without wavefield regularization) on the same data sets. For a fair comparison, the background velocity models are the same as above. The images from differently
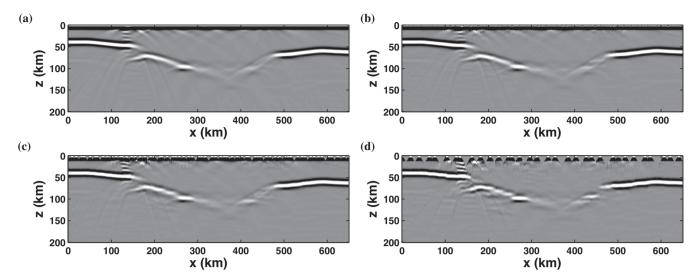
Figure 3. CCP stacking results with various station intervals. The stations are randomly distributed on the surface with an average spacing of (a) 2 km, (b) 6 km, (c) 10 km and (d) 20 km. For the ray tracing, 1-D linear velocity models are used.
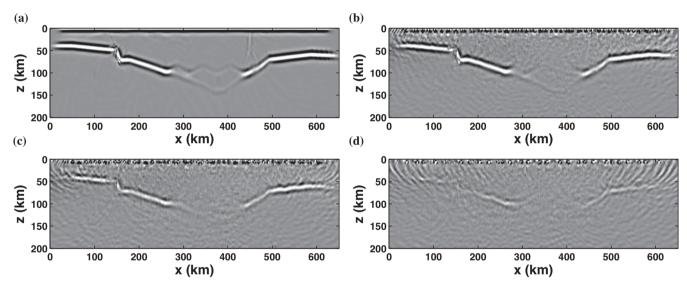

Figure 4. Passive-source RTM results without wavefield interpolation. The input data and background velocity models are the same as in CCP stacking. The average station spacing is (a) 2 km, (b) 6 km, (c) 10 km and (d) 20 km.

spaced data are shown in Fig. 4. For coarse sampling, for example, 20 km in Fig. 4(d), there are many circle-like artefacts due to spatial aliasing and the Moho is barely seen in the image. With increasing number of stations, such artefacts are gradually suppressed. The edge at 150 km horizontal distance starts to show up when the station interval decreases below 10 km (Fig. 4c). The topography of the gradual Moho in the middle is observable in the case of 6 km (Fig. 4b). With a 2 km spacing array (at least two sample points per wavelength; Fig. 4a), the Moho is delineated remarkably well with the absence of migration artefacts. The Moho kink, weak disrupted Moho and even the vertical suture (450 km at horizontal axis) are imaged with high clarity.

In a second set of experiments, the wavefield is enhanced by curvelet regularization. The randomly sampled data with different station intervals are interpolated to regular grids with 2 km spacing. After wavefield interpolation, the passive-source RTM images are shown in Fig. 5, from top to bottom, for the original data with 6, 10 and 20 km spacing, respectively. The spatial alias is remarkably

suppressed compared with the images based on the unprocessed data (Figs 4b–d). Using the data interpolated from 6 km grids, the RTM image (Fig. 5a) is almost identical to the one with 2 km interval (Fig. 4a). The disrupted Moho is clearly imaged in the case of 10 km (Fig. 5b) as well, and the Moho kink is observable in the extreme case in Fig. 5(c) (20 km interval). The artefacts introduced in the curvelet interpolation are not apparent in the stacked RTM images, since they are localized and randomly oriented features.

These tests demonstrate that: (1) as expected, the ability to use the full power of RTM used here (Shang *et al.* 2012) depends on data coverage and the success of regularization and interpolation; (2) with sufficient sampling RTM is superior over traditional migration of receiver function stacks; (3) traditional methods (such as CCP migration) can yield better results than wave-based methods if sampling is not sufficient for accurate wavefield reconstruction and (4) wavefield reconstruction as a pre-processing tool improves RTM as well as CCP migration.
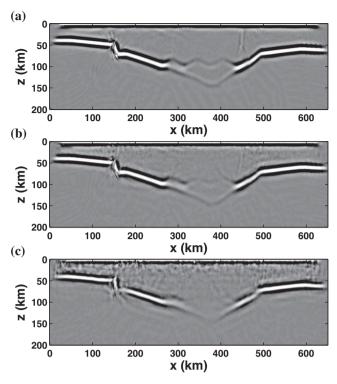
Figure 5. Passive-source RTM results with curvelet interpolations. The input data are regularly sampled data with 2 km spacing interpolated from randomly distributed receivers with an average interval of (a) 6 km, (b) 10 km and (c) 20 km.
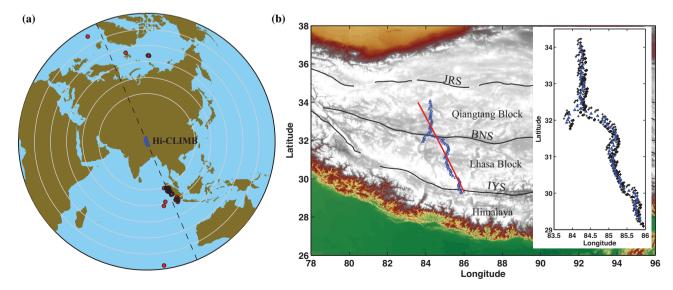
3 APPLICATION TO Hi-CLIMB ARRAY DATA

In this section, we apply 2-D passive-source RTM to Hi-CLIMB data (northern segment only) to image the lithosphere beneath Tibet. For this purpose, we project the data to the great circle (estimated from a least-squares fit of 71 station locations) shown as dashed line in Fig. 6(a) (red line in Fig. 6b). The horizontal co-ordinate is set along the linear profile, starting from point (29°N, 86°E). Note the linear profile (*x*-axis hereafter) extends northwards to the centre of the Qiangtang block, across Indus–Yarlung suture and Bangong–Nujiang suture (BNS), which are the main collision frontiers between different tectonic blocks in Tibet (Yin & Harrison 2000).

In order to minimize 3-D effects, only earthquakes ($m_b > 5.0$) approximately aligned with the *x*-axis are selected. The event location is confined to within a 30° cone with respect to the *x*-axis, and the epi-distance range is restricted from 30° to 90° (Fig. 6a). After careful scrutiny of all available data, 75 earthquakes with high SNR are used in the RTM imaging, of which 70 events are from southeast and only five are from northwest due to low seismicity (Fig. 6a). The *Ps* conversion points at Moho are plotted in Fig. 6(b), assuming a Moho depth of 70 km. The half-width of the swath is less than 15 km. This is comparable to the scale of the Fresnel zone (~14–23 km), which suggests that small-scale heterogeneities away from the *x*-axis form at most a limited contribution to the final 2-D image (Nowack *et al.* 2010).

For each earthquake, the recorded station locations are projected along the epi-distance contours onto the *x*-axis. Two horizontal components (E–W and N–S components) of the seismic data are rotated to the *x*-axis and its perpendicular direction. In 2-D RTM, only *x*- and *z*- (vertical) components are used for wavefield time reverse continuation. After removal of bad data, there are a total of about 4000 traces for each component. The gap between adjacent traces varies from 1 to 40 km with an average of 10 km. The *x*-axis is discretized into 2 km grids, and the trace locations are rounded up to the nearest grids with an average round-up error 0.5 km.

For each event, curvelet regularization is then performed component by component due to varying noise level among components. As an example shown in Fig. 7(a), the vertical component of one earthquake recorded by Hi-CLIMB array is aligned with the first *P*-wave arrival. The frequency band is from 0.05 to 1 Hz. There are 58 stations with various spacing from 2 to 26 km. After the wavefield regularization, the trace data are distributed to a regular grid with 2 km spacing, which implies that 'missing' traces amount to approximately 80 per cent of the total data used.

Figure 6. (a) Source distribution (red circles) used for the RTM imaging. Blue triangles represent stations in the north segment of Hi-CLIMB. Black dashed line is the great circle from a least-squares fit of a total of 71 station locations, which is 23.1° counter-clockwise from the north. (b) A close-up of north part of Hi-CLIMB stations (blue triangles) and projected linear profile (red line, the same as the great circle in (a)). IYS stands for Indus–Yarlung suture; BNS is for Bangong–Nujiang suture and JRS is Jinsha River suture. Inset: *Ps* conversion point (black dots) distribution assuming the Moho depth is 70 km.
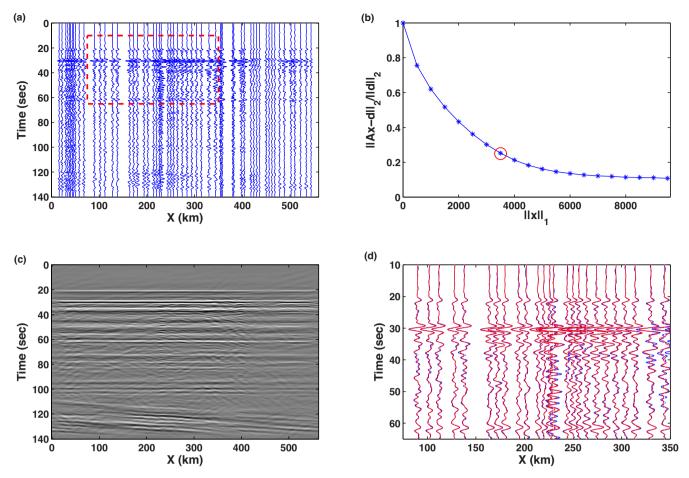
Figure 7. A field data example of wavefield interpolation. (a) The recorded data aligned with the first *P*-wave arrival. (b) The *L*-curve of the curvelet interpolation with a mask function, and the apparent velocity is chosen as 6 km s^{-1}. (c) Interpolated data with the parameter τ marked as a red circle in panel (b). (d) Zoomed-in comparison (clipped from panel (a), depicted as a red dashed box) between the interpolated and recorded traces. The recorded traces are shown in blue wiggles, and the interpolated data are plotted in red.

As shown above, we explore the *L*-curve by solving a series of Lasso problems to find the best regularization parameter τ, depicted in Fig. 7(b). The apparent velocity for the mask function in the interpolation is chosen as 6 km s^{-1} (see Appendix for detailed explanation of the mask function). It is slightly higher than the crust *S*-wave velocity since the first *P*-wave moveout is already corrected in Fig. 7(a). The best parameter τ is selected as a red circle along the *L*-curve (Fig. 7b). The corresponding interpolation result is demonstrated in Fig. 7(c). The amplitudes along the horizontal and slant (e.g. around 120 s) events vary smoothly. There are no significant artefacts after the interpolation. A zoomed-in comparison with the raw traces is plotted in Fig. 7(d). Major events are honored very well, whereas some subtle wiggles in the raw data are considered as uncorrelated noise and dimmed out after the recovery.

After trace interpolation, the source time function is estimated from the vertical component by principle component analysis (e.g. Rondenay *et al.* 2005) and then deconvolved from both components by Wiener deconvolution (e.g. Chen *et al.* 2010a). The data are filtered between 0.05 and 0.8 Hz. Passive-source RTM is then applied to the interpolated data sets. The background velocity models used in RTM is constructed by smoothing the reference models in Nowack *et al.* (2010) and superposing *P*- and *S*-wave tomographic models in Hung *et al.* (2010), shown in Fig. 8.

The final RTM image is shown in Fig. 9(d). For comparison, the CCP stacking results of receiver functions before (Fig. 9b) and after

(Fig. 9c) wavefield regularization are presented as well. Comparison of Figs 9(c) and (b) demonstrates that the amplitude anomaly between 32° and 32.5° latitude, caused by stations close to BNS suture zone, is considerably attenuated after wavefield regularization, without the loss of lateral resolution. Both CCP stacking and RTM show a relatively flat strong interface (Moho) at depth ∼70 km. The lateral continuity of the Moho is consistent with previous studies with receiver functions (Nábělek *et al.* 2009; Xu *et al.* 2015) and virtual deep seismic sounding (Tseng *et al.* 2009) but does not reveal major disruption of the Moho in the vicinity of the BNS, as suggested by Nowack *et al.* (2010) from Gaussian beam imaging. Changes in depth of the Moho discontinuity are apparent near 31° and 32.5°. The latter may be related to the BNS but seems to occur slight further north than inferred from (active source) deep seismic reflection profiling (Gao *et al.* 2013). Underneath the Lhasa block, strong negative interfaces are observed in middle-to-lower crust, as well as positive doublets at 60–70 km depth (e.g. between 29.7° and 30.5°). The flat Moho beneath Qiangtang block is somewhat fuzzier in RTM image (Fig. 9d) than in the CCP stacking image (Fig. 9c). One reason is that most of earthquakes are clustered in south, so the illumination underneath Qiangtang is much weaker than Lhasa block. Moreover, the station is sparser in Qiangtang (the spacing is more than 10 km), and the above synthetic analysis suggests that in that case CCP stacking is more robust than RTM.
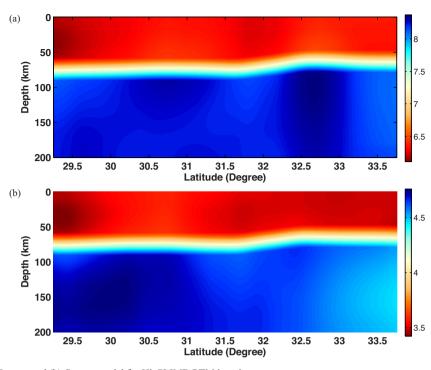
Figure 8. Background (a) *P*-wave and (b) *S*-wave model for Hi-CLIMB RTM imaging.
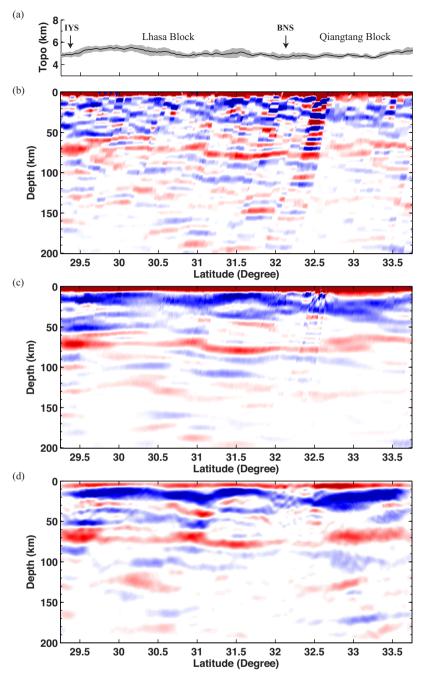
4 DISCUSSION

The workflow of the passive-source RTM presented here includes several steps that could potentially introduce uncertainties (or artefacts) to the final image. In the wavefield regularization, the station locations are rounded to the nearest regular grids, so the curvelet coefficients calculated by equispaced curvelet transform are an approximation. The round-up error can introduce jittering move-out, especially for dipping events in the data. In the Hi-CLIMB application, the crustal structures appear relatively flat, so major events are mostly horizontal after alignment with the first *P* arrival, see Fig. 7(a). So the jittering effect on the curvelet coefficient estimation is negligible in our application, especially for coarse-scale curvelets. If needed, this artefact can be suppressed by using a non-equispaced curvelet transform (e.g. Duchkov *et al.* 2010; Hennenfent *et al.* 2010).

As all other imaging methods the accuracy of RTM images depends, to some extent, on the background model. Small fluctuations of wave speed in the model will affect the amplitude of RTM image, but the locations of prominent scatterers are not very sensitive to the details of background models. Here, we assume that the background models are isotropic, so that *P* and *S* waves can be separated by polarization decomposition (Shang *et al.* 2012). In the presence of anisotropy q*P* and q*S* waves can be separated by solving the Christoffel equation (e.g. Dencker 1982; Dellinger & Etgen 1990; Stolk & De Hoop 2002). Along the northern segment of the Hi-CLIMB array, significant shear wave splitting (∼0.8 s and mainly east–west as the fast direction) suggests strong anisotropy in the crust and upper mantle (Chen *et al.* 2010b). Because of the nearly north–south linear geometry of sources and receivers, RTM image is not severely affected by the azimuthal anisotropy. In the future, out-of-plane events can be included in RTM using a 2.5-D or 3-D wave propagator (e.g. Roecker *et al.* 2010; Tong *et al.* 2014) and a heterogeneous and anisotropic background model (e.g. Foss *et al.* 2005).

In the north Hi-CLIMB application, the final images from CCP stacking and RTM are fairly comparable (Figs 9c and d), which is what we expect since the crustal structure is fairly simple and the station spacing is on the edge of where RTM would begin to produce superior results. Nevertheless, there is subtle difference between two images. For instance, the location of double interfaces (at 60–70 km depth) beneath Lhasa block is slightly different. Such a feature is observed in Nábělek *et al.* (2009) as well, using CCP stacking (across piecewise horizontal interfaces) of both *Ps* and surface multiples *PpPs* and *PpSs,* and interpreting the deeper one as eclogitic Indian Moho which might subduct northward up to 31°N. To reduce the uncertainty and improve spatial resolution, surface multiples can be regarded as virtual source reflection data (Yu *et al.* 2012) and incorporated in the RTM framework (e.g. Burdick *et al.* 2014), without the need to assume that the interfaces where conversions occur are horizontal.

5 CONCLUSIONS

In this paper, we demonstrate that passive-source RTM is superior to CCP stacking as long as stringent sampling requirements can be met. If the latter is not the case, however, traditional methods (such as CCP stacks) may yield better results than wave-based methods. We have presented a pragmatic method for wavefield regularization in the curvelet domain, which can be used to prepare array data for more accurate passive-source RTM or CCP migration. With trace interpolation as a pre-processing step, about 6 km station spacing is enough for typical teleseismic imaging of the crust and lithosphere (e.g. the dominant frequency is around 1 Hz), but where the effective spacing is larger CCP is likely to perform better (on the same, regularized data). The field data application on Hi-CLIMB data in Tibet clearly reveals a continuous Moho near 70 km depth, which is consistent with most previous

Figure 9. (a) Surface topography along the *x*-axis. The solid line is the mean elevation and the grey shading denotes the standard deviation across a swath of about 50 km in width. (b) CCP stacking image of receiver functions from 75 events. The receiver function frequency range is from 0.05 to 0.8 Hz. The elevation is not corrected here. (c) CCP stacking result after curvelet trace interpolation. (d) Passive-source RTM image after curvelet interpolation. The input data are the same as in (c).

studies. The station spacing in most currently available (passive source) arrays is on the edge of where RTM makes a difference over traditional methods, but if the trend in data acquisition continues toward further reduction of sensor spacing and, thus, denser wavefield sampling, the improvement of images produced by RTM like methods will become more significant and important. We note that incorporating surface multiples and iterative least-squares migration (e.g. Burdick *et al.* 2014) is expected to improve sampling and, thus, image quality compared to the application presented here.

REFERENCES

Andersson, F., De Hoop, M.V., Smith, H.F. & Uhlmann, G., 2008. A multiscale approach to hyperbolic evolution equations with limited smoothness, *Commun. Partial Differ. Equ.*, **33,** 988–1017.

Burdick, S., de Hoop, M.V., Wang, S. & van der Hilst, R.D., 2014. Reverse-time migration-based reflection tomography using teleseismic free surface multiples, *Geophys. J. Int.,* **196**(2), 996–1017.

Candès, E.J. & Demanet, L., 2005. The curvelet representation of wave propagators is optimally sparse, *Commun. Pure appl. Math.,* **58**, 1472–1528.

Candès, E.J. & Donoho, D.L., 2004. New tight frames of curvelets and optimal representations of objects with piecewise C2 singularities, *Commun. Pure appl. Math.,* **57**, 219–266.

Chen, C.W., Miller, D., Djikpesse, H., Haldorsen, J. & Rondenay, S., 2010a. Array-conditioned deconvolution of multiple-component teleseismic recordings, *Geophys. J. Int.,* **182**, 967–976.

Chen, L., Wen, L. & Zheng, T., 2005. A wave equation migration method for receiver function imaging: 1. Theory, *J. geophys. Res.,* **110**, B11309, doi:10.1029/2005JB003665.

Chen, W.-P., Martin, M., Tseng, T.-L., Nowack, R.L., Hung, S.-H. & Huang, B.-S., 2010b. Shear-wave birefringence and current configuration of converging lithosphere under Tibet, *Earth planet. Sci. Lett.,* **295**, 297–304.

Dellinger, J. & Etgen, J., 1990. Wave-field separation in two-dimensional anisotropic media, *Geophysics,* **55**, 914–919.

Dencker, N., 1982. On the propagation of polarization sets for systems of real principal type, *J. Funct. Anal.,* **46**, 351–372.

Duchkov, A.A., Andersson, F. & De Hoop, M.V., 2010. Discrete almost-symmetric wave packets and multiscale geometrical representation of (seismic) waves, *IEEE Trans. Geosci. Remote Sens.,* **48**, 3408–3423.

Dueker, K.G. & Sheehan, A.F., 1997. Mantle discontinuity structure from midpoint stacks of converted P to S waves across the Yellowstone hotspot track, *J. geophys. Res.,* **102**, 8313–8327.

Figueiredo, M.A.T., Nowak, R.D. & Wright, S.J., 2007. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, *IEEE J. Sel. Top. Signal Process.,* **1**, 586–597.

Foss, S.-K, De Hoop, M.V. & Ursin, B., 2005. Linearized 2.5-dimensional parameter imaging-inversion in anisotropic elastic media, *Geophys. J. Int.,* **161**, 722–738.

Gao, R., Chen, C., Lu, Z., Brown, L.D., Xiong, X., Li, W. & Deng, G., 2013. New constraints on crustal structure and Moho topography in Central Tibet revealed by SinoProbe deep seismic reflection profiling, *Tectonophysics,* **606**, 160–170.

Gilbert, H.J., Sheehan, A.F., Dueker, K.G. & Molnar, P., 2003. Receiver functions in the western United States, with implications for upper mantle structure and dynamics, *J. geophys. Res.,* **108**, 2229, doi:10.1029/2001JB001194.

Hennenfent, G., Fenelon, L. & Herrmann, F.J., 2010. Nonequispaced curvelet transform for seismic data reconstruction: a sparsity-promoting approach, *Geophysics,* **75**, Wb203–Wb210.

Herrmann, F.J. & Hennenfent, G., 2008. Non-parametric seismic data recovery with curvelet frames, *Geophys. J. Int.,* **173**, 233–248.

Hung, S.-H., Chen, W.-P., Chiao, L.-Y. & Tseng, T.-L., 2010. First multiscale, finite-frequency tomography illuminates 3-D anatomy of the Tibetan plateau, *Geophys. Res. Lett.,* **37**, L06304, doi:10.1029/2009GL041875.

Kabir, M.N. & Verschuur, D., 1995. Restoration of missing offsets by parabolic Radon transform, *Geophys. Prospect.,* **43**, 347–368.

Meltzer, A. *et al.*, 1999. The USArray initiative, *GSA Today,* **9**, 8–10.

Nábělek, J. *et al.*, 2009. Underplating in the Himalaya-Tibet collision zone revealed by the Hi-CLIMB experiment, *Science,* **325**, 1371–1374.

Nábělek, J.L., Vergne, J. & Hetényi, G., 2005. Project Hi-CLIMB: a synoptic view of the Himalayan collision zone and Southern Tibet, in *EoS, Trans. AGU,* **86**(52), Fall Meet. Suppl., Abstract T52A-02.

Naghizadeh, M. & Sacchi, M.D., 2010a. Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data, *Geophysics,* **75**, WB189–WB202.

Naghizadeh, M. & Sacchi, M.D., 2010b. On sampling functions and Fourier reconstruction methods, *Geophysics,* **75**, Wb137–Wb151.

Nowack, R.L., Chen, W.P. & Tseng, T.L., 2010. Application of Gaussian-beam migration to multiscale imaging of the lithosphere beneath the Hi-CLIMB array in Tibet, *Bull. seism. Soc. Am.,* **100**, 1743–1754.

Poppeliers, C. & Pavlis, G.L., 2003. Three-dimensional, prestack, plane wave migration of teleseismic P-to-S converted phases: 1. Theory, *J. geophys. Res.,* **108**, 2112, doi:10.1029/2001JB000216.

Roecker, S., Baker, B. & McLaughlin, J., 2010. A finite-difference algorithm for full waveform teleseismic tomography, *Geophys. J. Int.,* **181**(2), 1017–1040.

Rondenay, S., Bostock, M.G. & Fischer, K.M., 2005. Multichannel inversion of scattered teleseismic body waves: practical considerations and applicability, in *Seismic Earth: Array Analysis of Broadband Seismograms,* Geophys. Monogr. Ser., Vol. 157, pp. 187–203, eds Levander, A. & Nolet, G., American Geophysical Union.

Rondenay, S., Bostock, M.G. & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves 3. Application to the Cascadia 1993 data set, *J. geophys. Res.,* **106**, 30 795–30 807.

Ryberg, T. & Weber, M., 2000. Receiver function arrays: a reflection seismic approach, *Geophys. J. Int.,* **141**, 1–11.

Sacchi, M.D., Ulrych, T.J. & Walker, C.J., 1998. Interpolation and extrapolation using a high-resolution discrete Fourier transform, *IEEE Trans. Signal Process.,* **46**, 31–38.

Shang, X., de Hoop, M.V. & van der Hilst, R.D., 2012. Beyond receiver functions: passive source reverse time migration and inverse scattering of converted waves, *Geophys. Res. Lett.,* **39**, L15308, doi:10.1029/2012GL052289.

Smith, H.F., 1998. A parametrix construction for wave equations with $C^{1,1}$ coefficients, in *Annales de l'institut Fourier*, pp. 797–835, Grenoble.

Spitz, S., 1991. Seismic trace interpolation in the F-X domain, *Geophysics,* **56**, 785–794.

Stolk, C.C. & De Hoop, M.V., 2002. Microlocal analysis of seismic inverse scattering in anisotropic elastic media, *Commun. Pure appl. Math.,* **55**, 261–301.

Tong, P., Chen, C.W., Komatitsch, D., Basini, P. & Liu, Q., 2014. High-resolution seismic array imaging based on an SEM-FK hybrid method, *Geophys. J. Int.,* **197**(1), 369–395.

Trad, D.O., Ulrych, T.J. & Sacchi, M.D., 2002. Accurate interpolation with high-resolution time-variant Radon transforms, *Geophysics,* **67**, 644–656.

Tseng, T.-L., Chen, W.-P. & Nowack, R.L., 2009. Northward thinning of Tibetan crust revealed by virtual seismic profiles, *Geophys. Res. Lett.,* **36**, L24304, doi:10.1029/2009GL040457.

van den Berg, E. & Friedlander, M.P., 2008. Probing the Pareto frontier for basis pursuit solutions, *SIAM J. Sci. Comput.,* **31**, 890–912.

Xu, Q., Zhao, J., Yuan, X., Liu, H. & Pei, S., 2015. Mapping crustal structure beneath southern Tibet: seismic evidence for continental crustal underthrusting, *Gondwana Res.,* **27**, 1487–1493.

Yin, A. & Harrison, T.M., 2000. Geologic evolution of the Himalayan-Tibetan orogen, *Annu. Rev. Earth planet. Sci.,* **28**, 211–280.

Yu, C.-Q., Chen, W.-P., Ning, J.-Y., Tao, K., Tseng, T.-L., Fang, X.-D., Chen, Y.J. & van der Hilst, R.D., 2012. Thick crust beneath the Ordos plateau: implications for instability of the North China craton, *Earth planet. Sci. Lett.,* **357**, 366–375.

Zwartjes, P.M. & Sacchi, M.D., 2007. Fourier reconstruction of nonuniformly sampled, aliased seismic data, *Geophysics,* **72**, V21–V32.

APPENDIX: WAVEFIELD RECONSTRUCTION WITH SPARSITY PROMOTION

In seismic data reconstruction, we consider the observed data **d** as a subset of the desired interpolated data **m**. This relationship can be represented as

$$\mathbf{d} = \mathbf{Gm} + \mathbf{n}, \tag{A1}$$

where **G** is a sampling operator and **n** is additive noise (Naghizadeh & Sacchi 2010b). The desired data **m** can be synthesized by curvelet frames **x** as:

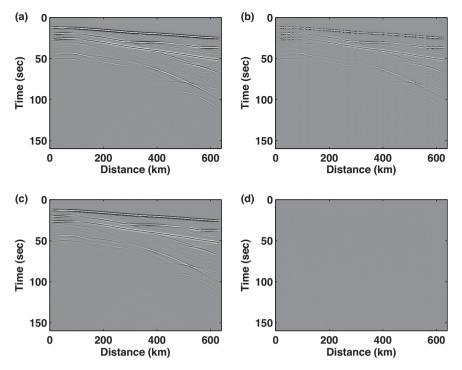$$\mathbf{m} = \mathbf{C}^{\mathrm{T}} \mathbf{x}, \tag{A2}$$

Figure A1. A synthetic example of curvelet interpolation. (a) The original data typically in teleseismic study. (b) Data after removing 50 per cent of the original traces randomly. (c) Recovered data by curvelet interpolation with sparsity promotion. (d) The difference between the (c) interpolated data and the (a) original data. All figures are displayed on the same colour scale, clipped by ± 5 per cent of the maximum value of the original data.

where $\mathbf{C}^{\mathrm{T}}$ is the adjoint (and inverse) of curvelet transform operator $\mathbf{C}$ (Candès & Donoho 2004). Substituting eq. (A2) into eq. (A1), the curvelet coefficients $\mathbf{x}$ with sparsity promotion can be found by minimizing the cost function

$$J = \frac{1}{2} ||\mathbf{d} - \mathbf{Ax}||_2 + \lambda ||\mathbf{x}||_1, \tag{A3}$$

where $\mathbf{A} = \mathbf{GC}^{\mathrm{T}}$ and the positive parameter λ is the Lagrange multiplier indicating the trade-off between L_2 norm of the data misfit and L_1 norm of the solution. The desired data $\mathbf{m}$ can then be estimated by eq. (A2).

Two other optimization approaches, basis pursuit denoise (BPDN) and Lasso problem, are closely related to eq. (A3). BPDN problem is described as

$$\min_{\mathbf{x}} ||\mathbf{x}||_1, \text{ subject to } ||\mathbf{Ax} - \mathbf{d}||_2 \leq \sigma \tag{A4}$$

where σ is a character of data noise level. The Lasso problem is

$$\min_{\mathbf{x}} ||\mathbf{Ax} - \mathbf{d}||_2, \text{ subject to } ||\mathbf{x}||_1 \leq \tau \tag{A5}$$

where τ is the upper bound of curvelet L_1 norm. These three problems are equivalent in some sense, and for appropriate parameters of λ, τ and σ, the solutions of eqs (A3)–(A5) coincide (van den Berg & Friedlander 2008).

Herrmann & Hennenfent (2008) solved eq. (A3) for noise-free data by iterative soft thresholding with cooling (ISTc) method, and the damping parameter λ gradually decreases towards zero during iterations. In practice, in order to mitigate the undersampling and noise issues, we solve a series of Lasso problems (or BPDN) to sample the L-curve explicitly, and then find the optimal constraint parameter given a teleseismic data set. We employ spectral projected-gradient ℓ_1 solver (SPG ℓ_1) for the Lasso problem (and BPDN), which converges faster than ISTc algorithm (Figueiredo *et al*. 2007; van den Berg & Friedlander 2008).

Since curvelets are direction selective, one can choose a weight function to penalize fine-scale nearly vertical curvelets (horizontally propagated high-frequency wave packets) which are more likely noise in the t–x domain data gather. A mask function $\mathbf{M}$ can be introduced in curvelet domain, and the operator $\mathbf{A}$ in eqs (A3)–(A5) changes as $\mathbf{A} = \mathbf{GC}^{\mathrm{T}}\mathbf{M}$. Assuming no evanescent waves, the mask function $\mathbf{M}$ can be designed according to the maximum slope in f–k domain, which is

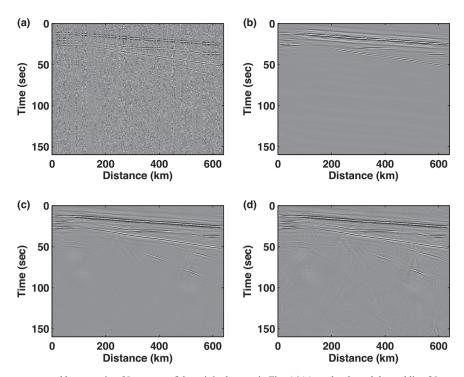$$s_{\max} = \frac{2\pi}{v_{\min}} \tag{A6}$$

where $v_{\min}$ is the minimum medium velocity or apparent velocity.

We first create a synthetic teleseismic section, shown in Fig. A1. The first arrival is approximately a plane wave, and later arrivals are converted waves and multiples. Note that polarity changes, conflicting dips and caustics can be observed in the later events. This original data set $\mathbf{m}_0$ is used as a ground truth solution to evaluate the quality of curvelet interpolation. The reconstruction quality is evaluated in decibels (dB) by the measure
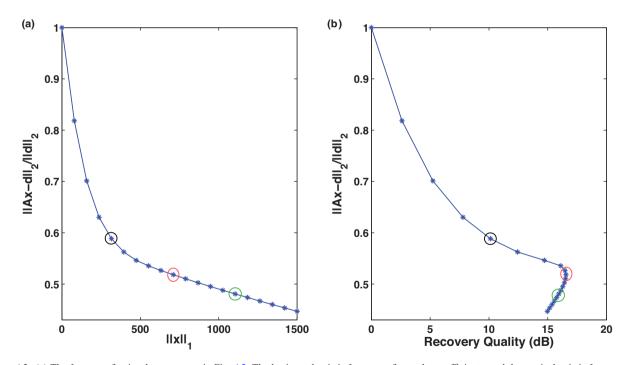
$$Q = -20 \log_{10} \left(\frac{||\mathbf{m}_0 - \hat{\mathbf{m}}||_2}{||\mathbf{m}_0||_2} \right), \tag{A7}$$

where $\hat{\mathbf{m}}$ is the reconstructed data.

In the first experiment, 50 per cent of original traces are randomly removed, as in Fig. A1(b) (the missing traces are replaced by zero traces). Since it is noise free, we solve the BPDN problem with $\sigma \to 0$ (here $\sigma = 0.001||d||_2$ is used) in eq. (A4), and the interpolated data $\mathbf{m}$ are then obtained by applying the adjoint curvelet transform in eq. (A2). The recovered data are shown in Fig. A1(c). The difference between the recovered data and the ground truth solution is illustrated in Fig. A1(d). The recovery quality Q measured by eq. (A7) is 34.60 dB. It is clear that wavefield interpolation works excellently and the missing traces are recovered almost perfectly.

Figure A2. (a) Noisy data constructed by removing 50 per cent of the original traces in Fig. A1(a) randomly and then adding 30 per cent white Gaussian noise. (b)–(d) Interpolated data for different parameter τ along the *L*-curve indicated as black, red and green circles in Fig. A3, respectively.

Figure A3. (a) The *L*-curve of noise data recovery in Fig. A2. The horizontal axis is L_1 norm of curvelet coefficients, and the vertical axis is L_2 norm of the data residual after interpolation (normalized by the data norm). The star symbols are those numerical sampling points along the *L*-curve by solving a series of Lasso problems. Three points depicted as colour circles are selected to investigate the influence of parameter τ on the curvelet interpolation. The corresponding interpolated results are shown in Figs A2(b)–(d). (b) The recovery quality (measured in dB) varies with parameter τ sampled along the *L*-curve in the left-hand panel.

In the presence of noise, the *L*-curve is explored to find the best trade-off parameters λ, σ and τ in eqs (A3)–(A5). Here, we sample the *L*-curve by a series of Lasso problems in eq. (A5). The upper bound τ is estimated in the following. For the recorded data in *t*–*x* domain, in each row (fixed time) we replace each missing

sample with the mean value of its horizontal neighbours. Such a process is conducted iteratively to fill the gaps from both sides. This simple method gives a low-quality interpolation, but provides a good reference for the upper bound of τ (by transforming it to curvelet domain). We provide two scenarios with 50 and
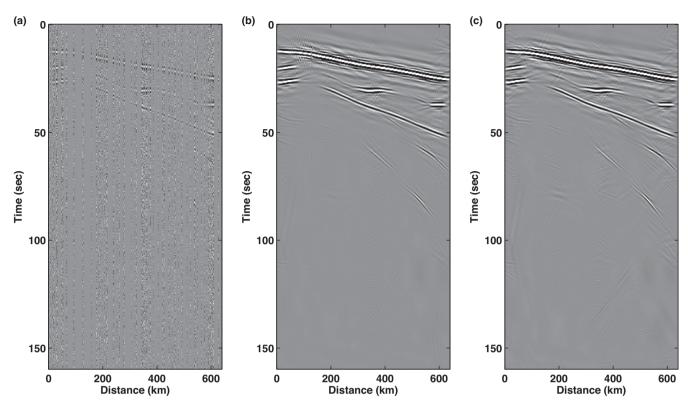
Figure A4. (a) Noisy data constructed by randomly removing 85 per cent of the original traces in Fig. A1(a), and then adding 30 per cent white Gaussian noise. (b) Recovered data after sparsity-promoted interpolation. (c) The same as Fig. A4(b) but with a mask function in the interpolation. The mask function is designed with apparent velocity 4 km s^{-1}.

85 per cent missing data, respectively. In each case, 30 per cent white Gaussian noise is added to the recorded data. We first remove 50 per cent traces and add random noise, depicted in Fig. A2(a). A series of parameter τ are used to sample the *L*-curve, plotted in Fig. A3(a). The recovery quality varies with respect to the parameter τ as well (Fig. A3b). To understand better the effects of parameter τ, three points on the *L*-curve are selected (colour circles in Fig. A3), and the corresponding interpolated results are illustrated in Figs A2(b)–(d).

It is noted that with a small τ, only the first arrival and some of strong converted phases are recovered (Fig. A2b). There are no evident smearing and oscillatory artefacts which are common in the Fourier domain interpolation methods. Subtle features such as multiples and caustics can be observed as τ increases (Fig. A2c). Above certain point, however, fine scale and nearly vertical curvelets

are introduced to overfit the noisy data (Fig. A2d). The recovery quality even decreases though the residual of data fitting decreases (Fig. A3b). The best recovery quality is 16.55 dB in this example (the red circle in Fig. A3).

Next, we remove 85 per cent of traces and add 30 per cent noise to the data, shown in Fig. A4(a). The best parameter τ yields 9.09 dB recovery quality in the absence of a mask function, and it increases to 12.54 dB with a mask function **M** (the apparent velocity is chosen 4 km s^{-1} here). The interpolated data are shown in Figs A4(b) and (c), respectively. Compared with the original data (Fig. A1a), the direct wave and primary converted waves are recovered well, though weak signals such as multiples around 50 s are not observable after interpolation. After applying a simple mask in the curvelet domain, some fine-scale artefacts are notably suppressed in the interpolated result (Fig. A4c).