Referee 2 (Erdinc Saygin):

Lines 30-, I am not sure if DSurfTomo will qualify for ambient noise data processing tool.
We agree with the reviewer that DSurfTomo is a tool for tomography and thus does not fit into the suite of packages mentioned here. Originally, the manuscript contained also references to packages for ambient noise tomography, which we have since removed for brevity and clarity. Thus, we removed the mention of DSurfTomo to be more consistent.

Please introduce "Syngine" as "The IRIS Synthetics Engine" at the beginning.
We've adapted the text accordingly and added the DOI of the IRIS repository as follows:
"This is achieved, for example, by the IRIS Synthetics engine (Syngine) repository (IRIS, 2015; Krischer et al., 2017) and by tools for the extraction and management of Green's function databases (van Driel et al., 2015a; Heimann et al., 2019)."

For equations 1 and 2, convolution and correlation terms are used in the text. However, the equations itself are in the frequency domain and are in multiplicative form.
This could indeed be misunderstood. We have added explicit mention of the change to frequency domain to both the relevant sentences. They now read:
"One component of ground motion u i observed at a seismic receiver at location x can be modeled as the convolution of the noise source time series with the impulse response of the Earth or Green's function G. In frequency domain, this relation is expressed as

$$u_i(\boldsymbol{x},\omega) = \int\limits_{\partial\oplus} G_{in}(\boldsymbol{x},\boldsymbol{\xi},\omega) N_n(\boldsymbol{\xi},\omega)\,\mathrm{d}\boldsymbol{\xi}^2,$$

(Aki and Richards, 2002), where summation over repeated indices is implied. The correlation of two such signals, averaged over an observation period, can be expressed by multiplication in the frequency domain, i.e.

$$\begin{aligned} C_{ij}(\boldsymbol{x}_1,\boldsymbol{x}_2,\omega) &= \langle u_i^*(\boldsymbol{x}_1,\omega) u_j(\boldsymbol{x}_2,\omega)\rangle \\ &= \left\langle \iint\limits_{\delta\oplus} G_{in}^*(\boldsymbol{x}_1,\boldsymbol{\xi}_1,\omega) N_n^*(\boldsymbol{\xi}_1,\omega) G_{jm}(\boldsymbol{x}_2,\boldsymbol{\xi}_2,\omega) N_m(\boldsymbol{\xi}_2,\omega)\,\mathrm{d}\boldsymbol{\xi}_1 \mathrm{d}\boldsymbol{\xi}_2 \right\rangle, \end{aligned}$$

etc.

Equation 3 has a minor typo.
Thank you, the typo was corrected.

Page 9, Line 240 is vague, and itself does not make much sense.
The sentence in question stated: "The simulation in AxiSEM3D is run for a global mesh, but dropping the number of Fourier expansion coefficients to 0 at a depth of 800 km and at a distance above 90 degree." We understand that more detail is required to make the sentence less vague. We suggest to replace the sentence by the following short paragraph:

"In AxiSEM3D, a method that couples a spectral-element discretisation with a pseudospectral expansion along the azimuth (Leng et al., 2019) we simulate the full desired 3D resolution inside the domain of interest. Rather than using absorbing boundaries as in the simulation with SPECFEM3D_GLOBE, we avoid spurious reflections in AxiSEM3D by using a global computational domain. The azimuthal Fourier expansion is tapered to a minimum of two Fourier coefficients outside of our domain of interest, which strongly reduces the additional compute time accrued due to the global simulation."

Note that we used this particular approach because absorbing boundaries didn't ship with the published version of AxiSEM3D at the time the simulations were conducted. Their development was ongoing at the time and is soon to be published by Haindl et al.

Figure 3 caption has autocorrelation, but all of the waveforms are from cross-correlations.
This has been corrected.

Section 5.2: Are the sources in Figure 6 simultaneously or randomly occurring. Is there any delay? And what is the frequency content?
The sources are assumed to have converged to the expected value of a random, uncorrelated field of sources as expressed in equation 3. Equation 3 yields power spectral density estimate Snm(xi, omega) as source term, i.e. phase information is no longer contained in the source term. This assumption may be questioned, but it is pervasively used in ambient noise studies, both of the Green's function retrieval and the cross-correlation modeling kind.

We propose to extend the sentence in lines 325-326 as follows:
"The frequency content of the starting model is homogeneous for all sources (background and blobs), with Gaussian power spectral density S(xi, omega) of equation 3 having a mean frequency 0.05 Hz and standard deviation of 0.02 Hz."

We did not modify the figure caption itself, as it is already a little bit lengthy.

Between 345 and 350 "wave forms" should be "waveforms".
Thank you, this has been corrected.

Figure 6: Rather than giving coordinates, please mark/number the selected stations and use them in the waveform plots.
This is an excellent suggestion, and we have added a marker for the station pairs in question.

I found the use of logarithmic signal ratio and asymmetry inversion confusing. Can you please describe it further or rephrase the part in Lines 310 onwards.
Since Referee 1 also commented on this part, we have attempted to revise the paragraph describing the respective sensitivity kernels in lines 310 ff thoroughly, and we have included the equations describing both misfit functions (full waveform and asymmetry). We hope that this clarifies the usage of the asymmetry in the inversion. The paragraph now reads:

“Sensitivity kernels computed with noisi can be used to run gradient-based inversion for the distribution of ambient seismic sources from a data set of observed ambient noise cross-correlations. To demonstrate the effectiveness of this approach, we conduct two synthetic inversions using two different functions to measure the misfit between observations and model. The sensitivity kernel of any misfit function can be expressed as

$$K_{zz}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{\xi}) = \int_{\omega=0}^{\omega_{Nyq}} G^*_{zz}(\boldsymbol{x}_1,\boldsymbol{\xi},\omega) G_{zz}(\boldsymbol{x}_2,\boldsymbol{\xi},\omega) f_{zz}(\boldsymbol{x}_1,\boldsymbol{x}_2,\omega)\, d\omega,$$

where merely the function f zz is determined by the chosen misfit function and corresponds to the derivative of the misfit function with respect to the modelled cross-correlation.
As first misfit function, we use the L2-norm of the synthetic (C syn) and observed (C obs) correlation waveforms, i.e.

$$\chi_{\text{fwi}} = \frac{1}{2}\left[\mathcal{C}^{syn} - \mathcal{C}^{obs}\right]^2.$$

in time domain, yielding

$$f(\boldsymbol{x}_1,\boldsymbol{x}_2,\omega) = \mathcal{F}\left[\mathcal{C}^{\text{syn}} - \mathcal{C}^{\text{obs}}\right],$$

where we denote the Fourier
transform by F. An exemplary waveform sensitivity kernel for the z-components of both receivers, and vertical sources, is shown in the left panel of Figure 5. It reveals how various locations of the source distribution affect the measurement. One can clearly recognize the pattern of stationary phase regions behind the stations and the oscillating sensitivity in between the stations (e.g. Snieder, 2004; Xu et al., 2019).
In contrast, the right panel of Figure 5 shows sensitivity K zz of another misfit function,

$$\chi_{\text{A}} = \frac{1}{2}\left[A(\mathcal{C}^{syn}) - A(\mathcal{C}^{obs})\right]^2,$$

where

$$A(x(\tau)) = \ln\left(\frac{\int [w_+(\tau)x(\tau)]^2 d\tau}{\int [w_-(\tau)x(\tau)]^2 d\tau}\right),$$

and w + , w − denote causal and a-causal window of the cross-correlation, respectively, and f becomes (where the dependency on the lag τ is omitted):

$$f(\boldsymbol{x}_1,\boldsymbol{x}_2,\omega) = \mathcal{F}\left[\left[A^{\text{syn}} - A^{\text{obs}}\right] \cdot \left[\frac{w_+^2 \mathcal{C}^{\text{syn}}}{\int [w_+ \mathcal{C}^{\text{syn}}]^2 d\tau} - \frac{w_-^2 \mathcal{C}^{\text{syn}}}{\int [w_- \mathcal{C}^{\text{syn}}]^2 d\tau}\right]\right].$$

For simplicity, we will refer to this second measurement as asymmetry in the following. This second sensitivity kernel (Figure 5, right panel) is smoother than the full-waveform one: The oscillating sensitivity between the stations is removed due to the windowing by w_-, w_+, and the stationary phase regions have opposite signs of sensitivity due to the ratio

$$\frac{\int [w_+(\tau)C(\tau)]^2\, d\tau}{\int [w_-(\tau)C(\tau)]^2\, d\tau}.$$

A body wave is caught in the measurement window, adding a faint ring of sensitivity near the stations probably due to body-wave surface-wave interaction (Sager et al., 2018a). The term $f_{zz}(x_1, x_2, \omega)$ encompasses the differences between both sensitivity kernels of Figure 5, by taking the form of equations 10 and 13 for waveforms and asymmetry measurement, respectively.”