Answer to reviewer 1

This paper presents a novel seismic approach aiming to temporally monitor the formation of sea ice through the deployment of a dense nodal seismic network. Thorough processing of ambient noise through beamformed cross-correlations resulted in the recovery of waveguided modes that were subsequently used as the basis for a Bayesian inverse scheme.

The paper is well written and constructed, as are most coming out of this group, and is suitable for publication upon very minor revisions. My few questions/comments are as follow.

We thank the reviewer for the constructive feedback. We have answered the comments below, and carefully reviewed our manuscript. In accounting for the comments of the reviewers, we have added a figure in section 2.2.2 to describe the workflow for recovering the noise correlation function. We also added two figures in Appendix A to show the characteristics of the recorded seismic noise, and a figure in Appendix B to show the covariance of sea ice parameters.

1) On paragraph 165: consider describing a bit better in math the procedure related to the SVD decomposition of the FK transform.

We have added a more detailed description of the processing. However, it is complicated to add more mathematical details without going into the full description of the method. Those details can be found in Minonzio et al. (2010).

This processing consists in the following steps:

1. The matrix of transmit-receive signals has three-dimensions: sources (M = 2 or 4), receivers (N), and time. The first step is the application of the Fourier transform to the temporal dimension of this matrix.

2. At each frequency, the resulting Fourier-domain matrix is sliced into 2D transmit-receive matrices. These matrices are then decomposed into singular values. The singular vectors define an orthonormal basis of the space dimensions along the transmitters (left-singular vector) and receivers (right-singular vector). The underlying idea behind this processing step is that the different levels of modal energy are distributed onto the singular vectors, the energy information being contained in the singular values. This allows a heuristic separation of the noise and signal subspaces, in a classical way for singular value-based filters.

3. The last step consists of defining test vectors that are representative of the wave propagation problem. In the present case, we use plane waves of the form $e^{-ik_{test}x_n}$, k_{test} is the wavenumber to be tested, and x_n (n = 1,2,...,N) is the coordinate of receiver n along the propagation. Finally, the test vectors are projected onto the singular vectors of the receivers' basis. This leads to a scalar product that is maximized when the wavenumber in the test vector matches that of the waves in the measured wavefield. In practice, this projection step is equivalent to calculating the discrete spatial Fourier transform of each singular vector.
Step 3 is performed at each frequency resulting from step 1.

Once steps 1-3 are performed, the resulting frequency-wavenumber spectrum significantly enhances the identification of the dispersion curves, for two reasons: (i) it is possible to separate signal from noise by applying a threshold to the singular values; and (ii) modal amplitude stands out at all frequencies and for all modes with the same spectral intensity (Fig. 4), despite their different relative amplitude in the wavefield, because singular vectors have a unit norm. The dispersion curves can therefore be identified on a larger bandwidth and with less SNR-related uncertainties than with conventional beamforming techniques (Moreau, Boué et al, 2020).

2) Could additional dispersion information have been retrieved by simply picking the maximum of the beamformer at every frequency?

This is a good question, because beamforming is known to be very robust to SNR. However, it is limited by spatial sampling. To apply beamforming, the full 2D array should be used. In the 2D array, spatial sampling is 4 meters, which allows to beamform the data without aliasing only up to 16 Hz (Moreau, Boué et al, 2020). Moreover, the SVD processing accounts for the multiplicity of sources with much more sensitivity (see figure B1 in Moreau, Boué et al, 2020). Hence, no additional information could be retrieved from beamforming, it is actually the opposite.

We have added the following sentence in section 2.2.2.

"Applying beamforming at frequencies higher than 16 Hz would cause aliasing problems, which prevents dispersion information to be extracted beyond 16 Hz."

We have also added a statement after the more detailed description of the SVD-based dispersion curves extraction:

"The dispersion curves can therefore be identified on a larger bandwidth and with less SNR-related uncertainties than with conventional beamforming techniques (Moreau, Boué et al, 2020)."

3) Why pose as an MCMC instead of a full grid search? In your case, your forward model is purely analytical unless I am mistaken, which means that a quick parallel implementation of a grid search should be quite feasible.

We understand from this question that the reviewer is asking about the efficiency of parallelized computation versus that of MCMC, which is sequential and thus cannot be parallelized. There are three main reasons why we have chosen to use MCMC:

1- When using a standard grid search, accuracy depends on the grid size. When using MCMC, the steps in the Markov Chain depend on the likelihood of the position in the parameters space. Typically, positions with high likelihood will be scanned with a much finer resolution than positions with low likelihood. For example, to achieve the same resolution with a grid search, one should use a grid size of 0.01 m for the thickness, 0.02 GPa for the Young's modulus, 0.02 for the Poisson's ratio and 40 kg for the density. This would require about 240 000 combinations to search through a parameters space of the same dimension as the one in the manuscript. With MCMC we require half this amount. Of course, a parallelized implementation of the grid search would still be much faster, but the carbon footprint of the inversion would be at least twice.
2- The main interest for Bayesian inference instead of a grid search is to obtain statistical information about the inverted parameters. The MCMC output is the joint probability density function of the parameters.
3- The simulated annealing part of our MCMC implementation allows the variance of the measurement errors to be estimated, which is a useful information when trying to identify the origin of potential discrepancies in the results. For example, in the present case the pattern shown in figure 5 tends to indicate that ice properties might vary over time more quickly in the NS direction than they do in the EW direction.

4) I'm curious as to why things seem to be generally insensitive to density, since this is an important stated objective of your study. I would recommend you explaining what you mean by "The inversions give a value that is very stable, at around 910±82 kg.m−3 for the EW direction, and 908 ± 80 kg.m−3 for the NS direction" on 320, and a second (perhaps unintentionally repeated) time: "Nevertheless, the actual uncertainty is probably much smaller, as all of the inversions give a very stable value" on 325, and again . The posterior density is more or less flat, and thus the mean value here, regardless of its consistency across lines, should probably not be treated as a constrained parameter.

Perhaps an added statement as to why you think the inversion fails to robustly constrain density might be helpful, or in what ways it could be accounted for.

We agree that density is not well-constrained. We believe that the larger standard deviation reflects the limits of our forward model more than it is an indicator of the limits of the methodology. The fact that the estimated density remains constant through days and directions is an indicator that this assumption is very likely.

The model is only sensitive to the effective properties of the {ice+snow} system, because it cannot account for the snow layer (about 40 cm thick, on average), which modifies the effective properties. Intuitively, it appears that the weight of the snow layer modifies the density of the {ice+snow} system more than it does its rigidity (Young's modulus) and expansion/contraction (Poisson's ratio). Presumably, a forward model able to account for snow would be a significant improvement, which should constrain the density in a better way.

We have added the following statement at the end of section 2.4.2.

Interestingly, the covariance of the parameters (see figure B1 in Appendix B) indicates that Poisson's ratio, despite being well-constrained, seems rather uncorrelated from the other parameters. On the other hand, Young's modulus, density and thickness appear to be strongly correlated, despite the density being not very well-constrained.

The observations seem to indicate that density having a flatter PDF reflects the limits of our forward model more than it is an indicator of the limits of the methodology. The model is only sensitive to the effective properties of the {ice+snow} system, because it cannot account for the snow layer (about 40 cm thick, on average), which modifies the effective properties. The weight of the snow layer modifies the density of the {ice+snow} system more than it does its rigidity (Young's modulus) and expansion/contraction (Poisson's ratio). Presumably, a forward model able to account for snow would be a significant improvement, which should constrain the density in a better way. The development of such a forward model is therefore an important follow up of this work.

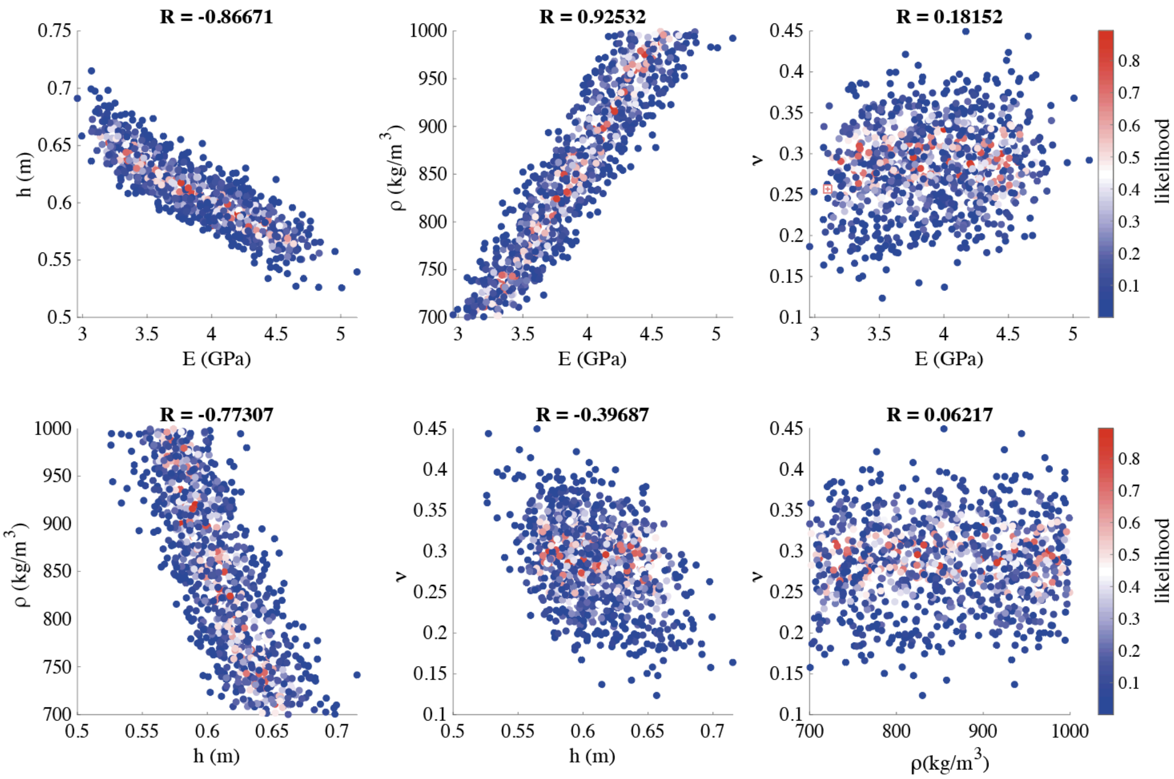

Figure B1. Covariance between Young's modulus, thickness, density and Poisson's ratio, with the associated correlation coefficient