Probabilistic multiparameter Backus–Gilbert method—Application to density inversion

Noami Kaplunov [●],[1] Coen Hofstede,[2] Dimitri Zigone [●],[3] Olaf Eisen,[2,3,4] Brian L. N. Kennett [●][5] and Andreas Fichtner [●][1]

[1]*Department of Earth Sciences, Swiss Federal Institute of Technology (ETH),* 8092 *Zurich, Switzerland. E-mail:* nkaplunov@student.ethz.ch
[2]*Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research,* 27570 *Bremerhaven, Germany*
[3]*Université de Strasbourg/CNRS, Institut Terre et Environnement de Strasbourg,* 67084 *Strasbourg, France*
[4]*Department of Geosciences, University of Bremen,* 28359 *Bremen, Germany*
[5]*Research School of Earth Sciences, The Australian National University,* 0200 *Canberra, Australia*

SUMMARY

We present an adaptation of the Backus–Gilbert method that enables (i) the incorporation of arbitrary prior knowledge and (ii) the solution of multiparameter inverse problems, providing a tunable balance between spatial resolution, inference errors and interparameter trade-offs. This yields a powerful approach for solving a class of inverse problems where the forward relation is linear or weakly nonlinear. The method rests on a probabilistic reformulation of Backus–Gilbert inversion and the solution of an optimization problem that maximizes deltaness while minimizing interparameter trade-offs. Applying the theory to multimode surface wave dispersion data collected by distributed acoustic sensing on the Northeast Greenland Ice Stream, we show that density in the firn layer may be constrained directly and without the need for scaling relations to depths of around ten metres, provided that dispersion data up to at least the third overtone of Rayleigh waves are available in the ∼10–50 Hz frequency band. The limiting factor that prevents the resolution of density at greater depth is data quality. Hence, progress on the direct inference of density could be made by repeated experiments or higher signal-to-noise ratios that would require better coupling and shielding of fibre-optic cables from wind and temperature fluctuations.

Key words: Inverse theory; Distributed acoustic sensing; Wave propagation.

1 INTRODUCTION

George Backus' and Freeman Gilbert's early work on inferences of Earth structure based on noisy data is frequently considered the hour of birth of modern geophysical inverse theory (Backus & Gilbert 1967, 1968, 1970). In addition to providing profound insight into the nature of inverse problems and circumventing the need to parametrize the model space, they posed and answered the question of how to negotiate a compromise between two antagonists: spatial resolution and certainty.

Both mathematically and philosophically, the Backus–Gilbert method differs from alternative approaches, including, at the opposite ends of the methodological spectrum, Bayesian inference (e.g. Tarantola 2005; Fichtner 2021) and misfit minimization by numerical optimization (e.g. Nocedal & Wright 1999). The major advantage of Backus–Gilbert theory is that spatial discretization, which is unavoidably subjective and affects inversion outcomes, is not required. Generalizations of the original Backus–Gilbert method include the SOLA method and efficient extensions to higher dimensions (e.g. Yanovskaya 2000; Zaroli *et al.* 2017), as well as variants for mildly nonlinear problems (Snieder 1991). Applications of Backus–Gilbert inversion can mostly be found in seismic tomography (e.g. Zaroli 2016, 2019; Amiri *et al.* 2023) and the imaging of the solar interior (e.g. Pijpers & Thompson 1992; Pijpers 1995; Kosovichev 1999).

Notwithstanding its elegance and numerous successful applications, the Backus–Gilbert method has limitations that prevent its more widespread use. By design, Backus–Gilbert inversion does not infer model parameters but their spatial averages, which have the advantageous property of being unique. While uniqueness greatly facilitates interpretation, averages may not be used directly in simulations that aim to predict new data, because most forward problem solvers require the actual model parameters as input. Therefore, Backus–Gilbert inversion primarily serves the purpose of quantitative interpretation instead of prediction.

A deficiency of the original Backus–Gilbert method that weighs more heavily is the restriction to single-parameter problems. From the perspective of an interpreter, knowledge of multiple parameters is often required to produce derived inferences. For instance, seismic wave speeds, attenuation and density are required to distinguish thermal from chemical heterogeneities in the Earth (e.g. Trampert *et al.* 2004;

Koelemeijer *et al.* 2017; Cobden *et al.* 2018). Furthermore, optimizing the inference of just one parameter may increase trade-offs with other parameters, thereby producing potentially useless results. For these reasons, multiparameter inversion has long been and continues to be considered one of the main challenges in seismic inversion (Kennett *et al.* 1988; Kennett & Sambridge 1998; Fichtner *et al.* 2024).

In addition to multiparameter problems, the incorporation of prior knowledge, frequently needed to obtain meaningful inferences, is not considered in Backus' and Gilbert's original work. While prior knowledge in the form of a probability density is an integral part of Bayesian inference, the inherently deterministic Backus–Gilbert approach does not naturally include this option.

Several attempts to extend the basic Backus–Gilbert method have been made during the past decades. Masters (1979) and Masters & Gubbins (2003) proposed a multiparameter variant in the context of global-scale density inversion using normal-mode data but could not include essential prior information, for example, on the total mass or moment of inertia of the Earth. Backus (1970a, b) developed the special case where the norm of inferred model properties is *a priori* bounded but did not allow for arbitrary prior knowledge.

The primary objectives of this work are twofold. First, we derive a Backus–Gilbert method for multiparameter problems where arbitrary prior knowledge can be incorporated. This approach is intended to answer the question of how inferences of one parameter may be optimized while (i) simultaneously reducing trade-offs with all other parameters and (ii) accounting for *a priori* information. Secondly, we illustrate the method in a multiparameter inversion for 1-D firn properties near the East Greenland Ice Core Project (EastGRIP, see section 3) using multimode Rayleigh wave dispersion measurements. Our analysis will focus on the extent to which firn density, typically estimated from wave speeds with empirical scaling relations (e.g. Kohnen 1972; Diez *et al.* 2014), can be constrained independently.

This paper is organized as follows: We begin in Section 2 with the development of multiparameter versions of the Backus–Gilbert method, with and without consideration of data errors. This leads to a probabilistic reformulation of the method that naturally permits the incorporation of prior knowledge via the conjunction of probability densities. In Section 3, we introduce a seismic distributed acoustic sensing (DAS) data set collected on the Northeast Greenland Ice Stream (NEGIS), near the EastGRIP ice core drilling site. Dispersion curves obtained from the data allow for multiple Rayleigh wave modes to be distinguished. Subsequently, in Section 4, we provide an extensive sensitivity analysis on the basis of the EastGRIP dispersion data, with a focus on our ability to obtain averaging kernels for density that approximate a delta function while reducing trade-offs with *S*- and *P*-wave speeds. Using a specific set of tuning parameters, Section 5 presents an example inversion for density, *S*-wave and *P*-wave structure within the upper 100 m. Finally, in Section 6, we discuss general advantages and limitations of the method, as well as specific results concerning the resolution of 1-D density structure by multimode Rayleigh wave data.

2 THEORETICAL BACKGROUND FOR PROBABILISTIC MULTIPARAMETER BACKUS–GILBERT INVERSION

2.1 A condensed review of classical Backus–Gilbert theory

Before delving into detailed developments, it is useful to establish mathematical context and notation with a brief recapitulation of the single-parameter Backus–Gilbert method for error-free data d_i^{obs} that can be perfectly explained by some continuously distributed model $m(z)$. The index $i = 1, ..., N_{\text{obs}}$ enumerates the measurements, and z represents, in most geophysical cases, depth. The data are linearly related to m via a sensitivity or data kernel G_i,

$$d_i^{\text{obs}} = \int_0^H G_i(z) m(z) \, \mathrm{d}z \,, \tag{1}$$

where $z = 0$ at the Earth's surface and H is some maximum depth. A simple, yet profound, consequence of (1) is that only spatial averages,

$$\langle A, m \rangle = \int_0^H A(z) m(z) \, \mathrm{d}z \,, \tag{2}$$

with some averaging kernel $A(z)$, can be inferred from the observations. The kernel acts as a filter or blurry lens that obscures the details of $m(z)$. Eq. (1) also motivates the representation of $A(z)$ using the kernels G_i as basis functions,

$$A(z) = \sum_{i=1}^{N_{\text{obs}}} a_i G_i(z) \,, \tag{3}$$

where a_i are coefficients that remain to be determined as a function of the desired resolution. Inserting (3) into (2) reveals that the averaging procedure is equivalent to a linear combination of the observations,

$$\langle A, m \rangle = \sum_{i=1}^{N_{\text{obs}}} a_i \int_0^H G_i(z) m(z) \, \mathrm{d}z = \sum_{i=1}^{N_{\text{obs}}} a_i d_i^{\text{obs}} = \mathbf{a}^T \mathbf{d}^{\text{obs}} \,. \tag{4}$$

Eq. (4) has important implications for uniqueness. While (infinitely) many models $m(z)$ may explain the error-free data d_i^{obs} perfectly, the average $\langle A, m \rangle$ is unique because it can be computed via the model-independent scalar product $\mathbf{a}^T \mathbf{d}^{\text{obs}}$. To obtain a sharp image of $m(z)$ at

some target depth z_0, the averaging kernel $A(z)$ should equal the delta function $\delta(z - z_0)$. While this ideal is unachievable in practice, the coefficients a_i may still be chosen such that $A(z)$ is at least close to $\delta(z - z_0)$ in the sense of a deltaness measure $s_0(A, z_0)$. Though the choice of a deltaness measure is to some extent subjective (e.g. Arfken & Weber 2005), the most frequently used and convenient version is

$$s_0(A, z_0) = 12 \int_0^H (z - z_0)^2 \, [A(z)]^2 \, dz \,. \tag{5}$$

In the special case where $A(z)$ is a boxcar function of width w, centred around z_0, we find $s_0(A, z_0) = w$, which justifies the secondary interpretation of $s_0(A, z_0)$ as a position-dependent resolution or averaging length. At this point, the Backus–Gilbert method has taken the form of an optimal-design problem. In fact, minimizing $s_0(A, z_0)$ with the normalization constraint $\int_0^H A(z) \, dz = 1$ leads to a linear system of equations for the coefficients a_i from which the optimal average can be computed via (4).

2.2 Multiparameter inversion of error-free data

To develop a Backus–Gilbert method for multiple parameters we take inspiration from the deep-Earth imaging application of Masters (1979) and begin by neglecting the influence of data errors. To factor in the effect of multiple model parameters on the data d_i^{obs}, we reformulate eq. (1) such that

$$d_i^{\mathrm{obs}} = \int_0^H \sum_{n=0}^{N_p} G_i^{(n)}(z) \, m^{(n)}(z) \, dz = \int_0^H \left[G_i^{(0)}(z) \, m^{(0)}(z) + G_i^{(1)}(z) \, m^{(1)}(z) + G_i^{(2)}(z) \, m^{(2)}(z) + \ldots \right] dz, \tag{6}$$

where N_p is the number of model parameters in addition to the parameter of interest, denoted by $m^{(0)}(z)$. As in Section 2.1, we wish to form a linear combination $\mathbf{a}^T \mathbf{d}^{\mathrm{obs}}$ of the data that has advantageous resolution properties. Substituting (1) for d_i^{obs} gives

$$\mathbf{a}^T \mathbf{d}^{\mathrm{obs}} = \sum_{n=0}^{N_p} \langle A^{(n)}, m^{(n)} \rangle = \langle A, \mathbf{m} \rangle \,, \tag{7}$$

with the components $A^{(n)}$ of the vectorial averaging kernel $\mathbf{A}$ defined as

$$A^{(n)}(z) = \sum_{i=1}^{N_{\mathrm{obs}}} a_i \, G_i^{(n)}(z) \,, \tag{8}$$

for any n between 0 and N_p. Eq. (8) implies that the linear combination of the data, already considered in (4) for the single-parameter Backus–Gilbert method, inevitably involves averages over all model parameters. Unless $A^{(1)}, \ldots, A^{(N_p)}$ are negligibly small compared to $A^{(0)}$, inferences of $m^{(0)}$ are contaminated by the influence of the remaining model parameters.

To maximize resolution of the parameter of interest, while minimizing contamination by other parameters, the optimal set of coefficients a_i should fulfil two requirements: (i) as before, the averaging kernel for the parameter of choice, $m^{(0)}(z)$, should closely approximate $\delta(z - z_0)$. (ii) the averaging kernels of all remaining model parameters $m^{(n)}(z)|_{n>0}$ should simultaneously be close to zero. A balance between these two antagonistic goals can be achieved by using two criteria; the deltaness criterion as in (5) for $m^{(0)}(z)$,

$$s_0(A^{(0)}, z_0) = 12 \int_0^H (z - z_0)^2 \left[A^{(0)}(z) \right]^2 \, dz, \tag{9}$$

and a minimization criterion for $m^{(n)}(z)|_{n>0}$ that forces the L_2 norm of $A^{(n)}(z)|_{n>0}$ towards zero,

$$s_n(A^{(n)}) = \int_0^H \left[A^{(n)}(z) \right]^2 \, dz \,, \quad \text{for} \quad n \in [1, 2, \ldots, N_p] \,. \tag{10}$$

To determine suitable coefficients a_i, we employ the cumulative criterion

$$\begin{aligned} s(\mathbf{A}, z_0) &= s_0(A^{(0)}, z_0) + \frac{1}{2} \sum_{n=1}^{N_p} \beta_n s_n(A^{(n)}) \\ &= 12 \int_0^H (z - z_0)^2 \left[A^{(0)}(z) \right]^2 \, dz + \frac{1}{2} \sum_{n=1}^{N_p} \beta_n \int_0^H \left[A^{(n)}(z) \right]^2 \, dz \,, \end{aligned} \tag{11}$$

where the weights β_n serve to balance the importance of achieving deltaness for $m^{(0)}$ and reducing the influence of $m^{(n)}|_{n>0}$. As we will see in later sections, specific choices of β_n will depend on the trade-offs between resolution and interparameter leakage that are (subjectively)

deemed acceptable in the context of a specific application. To minimize $s(\mathbf{A}, z_0)$ for a fixed z_0, while making sure that $\int_0^H A^{(0)}\, dz = 1$, we construct the Lagrange function

$$L(\mathbf{a}, \lambda) = 12 \int_0^H (z - z_0)^2 \left[\sum_{i=1}^{N_{\text{obs}}} a_i G_i^{(0)}(z) \right]^2 dz + \frac{1}{2} \sum_{n=1}^{N_p} \beta_n \int_0^H \left[\sum_{i=1}^{N_{\text{obs}}} a_i G_i^{(n)}(z) \right]^2 dz - \lambda \left[1 - \sum_{i=1}^{N_{\text{obs}}} a_i \int_0^H G_i^{(0)}(z)\, dz \right], \tag{12}$$

where λ is a Lagrange multiplier. Forcing the partial derivatives of L with respect to the coefficients a_i to zero yields

$$\frac{\partial L}{\partial a_k} = 24 \int_0^H (z - z_0)^2 \sum_{i=1}^{N_{\text{obs}}} a_i G_i^{(0)}(z) G_k^{(0)}(z)\, dz + \sum_{n=1}^{N_p} \beta_n \int_0^H \sum_{i=1}^{N_{\text{obs}}} a_i G_i^{(n)}(z) G_k^{(n)}(z)\, dz + \lambda \int_0^H G_k^{(0)}(z)\, dz = 0. \tag{13}$$

Defining the vector $\mathbf{u}$ with components

$$u_k = \int_0^H G_k^{(0)}(z)\, dz, \tag{14}$$

and symmetric matrices $\mathbf{S}$ and $\mathbf{T}^{(n)}$ with elements

$$S_{ik} = 24 \int_0^H (z - z_0)^2 G_i^{(0)}(z) G_k^{(0)}(z)\, dz, \tag{15}$$

and

$$T_{ik}^{(n)} = \int_0^H G_i^{(n)}(z) G_k^{(n)}(z)\, dz, \tag{16}$$

leads to a condensed version of the linear system (13),

$$\left(\mathbf{S} + \sum_{n=1}^{N_p} \beta_n \mathbf{T}^{(n)} \right) \mathbf{a} = -\lambda \mathbf{u}. \tag{17}$$

Eq. (17) determines $\mathbf{a}' = \mathbf{a}/\lambda$. Noting that

$$\int_0^H \sum_{i=1}^{N} a_i G_i(z)\, dz = \mathbf{a}^T \mathbf{u} = 1, \tag{18}$$

we find $\lambda = 1/(\mathbf{a}'^T \mathbf{u})$, which can be used to obtain the coefficients $\mathbf{a} = \mathbf{a}' \lambda$.

2.3 Error-contaminated data

In any real-world application, measurements are contaminated by errors that we assume to be additive and representable as the kth realization of a random variable $\Delta d_{i,k}$. Hence, we may write the components of the data vector as

$$d_{i,k}^{\text{obs}} = \int_0^H \sum_{n=0}^{N_p} G_i^{(n)}(z)\, m^{(n)}(z)\, dz + \Delta d_{i,k}. \tag{19}$$

Computing the linear combination of the data, we find

$$\mathbf{a}^T \mathbf{d}_k^{\text{obs}} = \langle \mathbf{A}, \mathbf{m} \rangle + \varepsilon_k, \tag{20}$$

where the averaging error for the kth realization that we observed is defined as

$$\varepsilon_k = \sum_{i=1}^{N_{\text{obs}}} a_i \Delta d_{i,k}. \tag{21}$$

While it is not possible to work with individual errors $\Delta d_{i,k}$, their statistical properties can be estimated from a sufficiently large number N_s of repeated experiments. In particular, the elements of the data error covariance can be approximated as

$$(\mathbf{C}_D)_{ij} \approx \frac{1}{N_s} \sum_{k=1}^{N_s} \sum_{i,j=1}^{N_{\text{obs}}} \Delta d_{i,k} \Delta d_{j,k}, \tag{22}$$

provided that the average errors vanish. Based on (22), the variance σ_ε^2 of the averaging error (21) can be estimated,

$$
\begin{aligned}
\sigma_\varepsilon^2 &\approx \frac{1}{N_s} \sum_{k=1}^{N_s} \varepsilon_k^2 = \frac{1}{N_s} \sum_{k=1}^{N_s} \left(\sum_{i=1}^{N_{\text{obs}}} a_i \Delta d_{i,k} \right)^2 \\
&= \sum_{i,j=1}^{N_{\text{obs}}} a_i a_j \frac{1}{N_s} \sum_{k=1}^{N_s} \Delta d_{i,k} \Delta d_{j,k} \\
&= \mathbf{a}^T (\mathbf{C}_D) \mathbf{a}\,.
\end{aligned}
\tag{23}
$$

The standard deviation $\sqrt{\sigma_\varepsilon^2} = \sigma_\varepsilon$ serves as a measure of uncertainty in the averages derived from the data. As before, we aim to minimize the averaging length s_0 defined in (9), while also minimizing the L_2 norm of the remaining kernels, $s_{n>0}$, given in (10). Additionally, we now wish to minimize the variance of the averaging error, and we do so by adding σ_ε^2 to the cumulative criterion

$$
\begin{aligned}
s(\mathbf{A}, z_0) &= s_0(A^{(0)}, z_0) + \frac{1}{2} \sum_{n=1}^{N_p} \beta_n s_n(A^{(n)}) + \frac{1}{2} \gamma \sigma_\varepsilon^2 \\
&= 12 \int_0^H (z - z_0)^2 \left[A^{(0)}(z) \right]^2 \, \mathrm{d}z + \frac{1}{2} \sum_{n=1}^{N_p} \beta_n \int_0^H \left[A^{(n)}(z) \right]^2 \, \mathrm{d}z + \frac{1}{2} \gamma \sum_{i,j=1}^{N_{\text{obs}}} a_i (\mathbf{C}_D)_{ij} a_j\,.
\end{aligned}
\tag{24}
$$

The weight γ controls the importance given to the minimization of the averaging error, and its value will need to be chosen such that the trade-offs between deltaness, interparameter leakage and uncertainty are acceptable in the context of a specific application. As in Section 2.2, we determine the coefficients $\mathbf{a}$ by forcing the partial derivatives of the Lagrange function, $\partial/\partial a_i$, to zero. This yields a new linear system of equations,

$$
\left(\mathbf{S} + \sum_{n=1}^{N_p} \beta_n \mathbf{T}^{(n)} + \gamma \mathbf{C}_D \right) \mathbf{a} = -\lambda \mathbf{u}\,, \quad \text{with} \quad \lambda = -\mathbf{a}^T \left(\mathbf{S} + \sum_{n=1}^{N_p} \beta_n \mathbf{T}^{(n)} + \gamma \mathbf{C}_D \right) \mathbf{a}\,.
\tag{25}
$$

2.4 Probabilistic Backus–Gilbert inversion and the incorporation of prior knowledge

The incorporation of prior knowledge on model parameter averages requires a slight generalization of the Backus–Gilbert method for data with errors. For this, we assume that the statistics of the data errors $\Delta \mathbf{d}_k$ can be described by a probability density $\rho_D(\Delta \mathbf{d})$. From this, we can derive the resulting probability density of the averaging errors ε_k, denoted by $\rho_E(\varepsilon)$. In fact, letting $\mathbf{R}$ be a rotation matrix that rotates the coefficient vector $\mathbf{a}$ onto the 1-axis, that is, $\mathbf{Ra} = a\mathbf{e}_1$ with $a = |\mathbf{a}|$ and the unit vector $\mathbf{e}_1 = (1, 0, ..., 0)^T$, we find

$$
\varepsilon = \mathbf{a}^T \Delta \mathbf{d} = \mathbf{a}^T \mathbf{R}^T \mathbf{R} \Delta \mathbf{d} = \mathbf{e}_1^T \boldsymbol{\varepsilon} = \varepsilon_1\,.
\tag{26}
$$

The rotated data error vector is defined as $\boldsymbol{\varepsilon} = a \mathbf{R} \Delta \mathbf{d}$. Eq. (26) implies that the distribution of the scalar ε equals the distribution of the vectorial $\boldsymbol{\varepsilon}$, marginalized over $\varepsilon_2, ..., \varepsilon_N$. To find this marginal distribution, we first apply a coordinate transformation to ρ_D that produces the distribution of $\boldsymbol{\varepsilon}$,

$$
\rho(\boldsymbol{\varepsilon}) = \rho_D \left(\frac{1}{a} \mathbf{R}^T \boldsymbol{\varepsilon} \right) \left| \det \left(\frac{1}{a} \mathbf{R} \right) \right| = \frac{1}{a^N} \rho_D \left(\frac{1}{a} \mathbf{R}^T \boldsymbol{\varepsilon} \right)\,.
\tag{27}
$$

Marginalizing over all components of $\boldsymbol{\varepsilon}$ except ε_1, yields

$$
\rho_E(\varepsilon) = \frac{1}{a^N} \int \rho_D \left(\frac{1}{a} \mathbf{R}^T \boldsymbol{\varepsilon} \right) \, \mathrm{d}\varepsilon_2 ... \mathrm{d}\varepsilon_N\,.
\tag{28}
$$

In the special case where the distribution of the data errors is a Gaussian with covariance $\mathbf{C}_D = \sigma^2 \mathbf{I}$, eq. (28) simplifies to

$$
\rho_E(\varepsilon) = \frac{1}{\sqrt{2\pi}\, a\sigma} e^{-\frac{1}{2} \left(\frac{\varepsilon}{a\sigma} \right)^2}\,.
\tag{29}
$$

Using (20), we can transform the distribution of the averaging errors, $\rho_E(\varepsilon)$, into the distribution of the actual model parameter averages,

$$
\rho_M(\langle \mathbf{A}, \mathbf{m} \rangle) = \rho_E(\mathbf{a}^T \mathbf{d}^{\text{obs}} - \langle \mathbf{A}, \mathbf{m} \rangle)\,.
\tag{30}
$$

Eq. (30) is the probabilistic variant of Backus–Gilbert inversion. It expresses data-derived averages in terms of a probability density that describes model uncertainty in terms of the underlying data errors. On the basis of (30), prior knowledge on plausible averages, encoded in a prior probability density $\rho_M^{\text{prior}}(\langle \mathbf{A}, \mathbf{m} \rangle)$, can easily be included via the conjunction of probability densities,

$$
\rho_M^{\text{post}}(\langle \mathbf{A}, \mathbf{m} \rangle) = q\, \rho_M(\langle \mathbf{A}, \mathbf{m} \rangle)\, \rho_M^{\text{prior}}(\langle \mathbf{A}, \mathbf{m} \rangle)\,,
\tag{31}
$$

with a normalization constant q, which ensures that the posterior probability density ρ_M^{post} integrates to 1. Using these theoretical developments, we continue to provide examples of multiparameter averaging kernels, a Backus–Gilbert inversion of the data introduced in Section 3, and the incorporation of prior knowledge.

(a) geographic and experimental setup
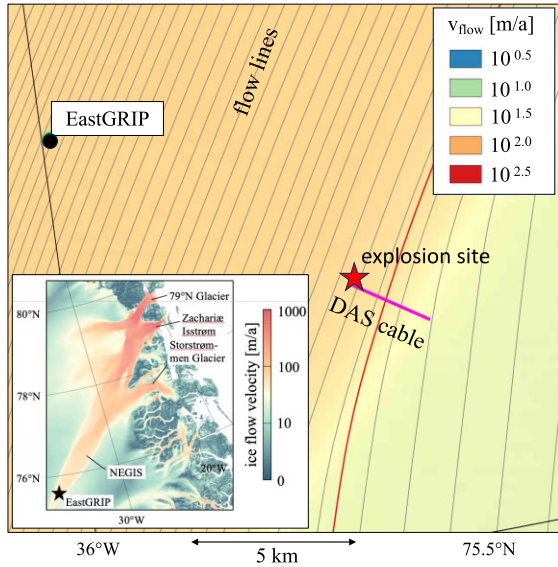
(b) data in frequency – phase velocity domain
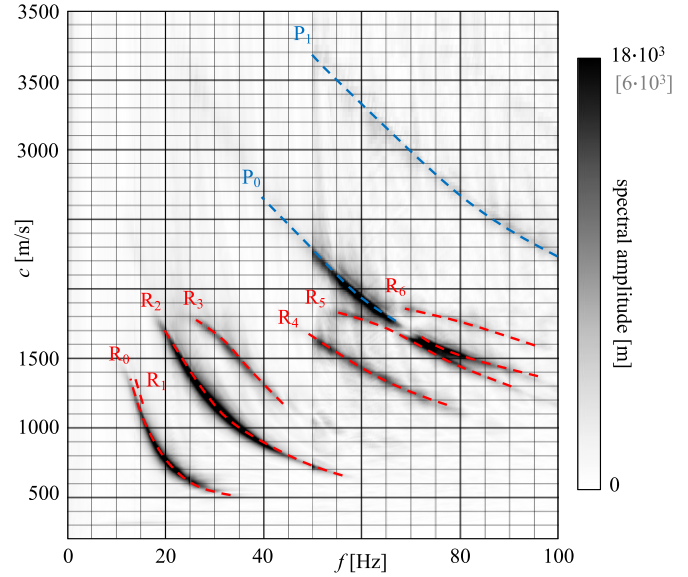
Figure 1. Setup and data. (a) Location of the DAS cable and the explosion site on the Northeast Greenland Ice Stream (NEGIS). The EastGRIP ice core drill site is shown for reference. Ice flow velocities and flow lines are from Joughin *et al.* (2018). (b) DAS data produced by the explosion, visualized in the frequency-phase velocity domain. Rayleigh wave modes are marked in red, and pseudo-acoustic modes in blue. For better visibility of lower amplitude modes, the spectral amplitudes at frequencies above 50 Hz are amplified by a factor of 3 (grey colour bar label).

3 SEISMIC DATA FROM THE NORTHEAST GREENLAND ICE STREAM

To bring our previous developments of a probabilistic multiparameter Backus–Gilbert method to life, we include examples based on seismic data collected near the EastGRIP ice core drilling site on the Northeast Greenland Ice Stream (NEGIS) in summer 2022 (Fig. 1a). With a length of ~600 km, NEGIS is the largest ice stream of the Greenland Ice Sheet, accounting for around 12 per cent of its total mass discharge (Rignot & Mouginot 2012). The acquisition system consisted of a Solifos BRUfield fibre-optic cable deployed over a length of 3000 m across the eastern shear margin of NEGIS. One of the four single-mode fibres in the cable was attached to a Silixa iDAS v2.0 interrogator, which collected strain rate data with 10 m gauge length, 2 m channel spacing and 1 kHz sampling rate. Our example data originate from a detonation of 100 g PETN (20 m detonation cord placed at the surface) close to the eastern end of the cable. The frequency–phase velocity representation in Fig. 1(b) reveals an unusually large number of clearly distinguishable Rayleigh and pseudo-acoustic modes. In subsequent examples, the data d_i^{obs} will be the phase velocities at selected frequencies of the Rayleigh fundamental mode R_0 and the Rayleigh overtones R_2, R_3, R_4 and R_5. Being only weakly excited and close to R_0, the first overtone R_1 will be excluded from the analyses.

The data closely resemble those produced by an airplane landing near the drill site, which served to constrain *S*-wave speed in the firn layer using a single-parameter Backus–Gilbert inversion (Fichtner *et al.* 2023b). In contrast to the underlying ice, firn is permeable to air and easily compressible (e.g. Paterson 1994), making it an essential ingredient of ice sheet mass balance calculations (e.g. Helsen *et al.* 2008), ice core climatology (e.g. Parrenin *et al.* 2012), estimates of surface melt (e.g. van den Broeke 2005) and the correction of seismic reflection and transmission data for near-surface effects (e.g. Schlegel *et al.* 2019; Fichtner *et al.* 2023a).

4 SENSITIVITY ANALYSIS

In the first series of applications, we perform various sensitivity analyses on modal dispersion results to investigate the resolution of density structure, the influence of higher surface wave modes and the effect of including data errors. The relevant model parameters are the *S*-wave speed v_s, the *P*-wave speed v_p and density ρ. To make averaging kernels for different parameters comparable, we work with the relative model parameters $\rho' = \rho/\rho^{\mathrm{init}}$, $v_s' = v_s/v_s^{\mathrm{init}}$ and $v_p' = v_p/v_p^{\mathrm{init}}$, where the initial model parameters $\mathbf{m}^{\mathrm{init}} = (\rho^{\mathrm{init}}, v_s^{\mathrm{init}}, v_p^{\mathrm{init}})$ were found by manually fitting dispersion data of the fundamental Rayleigh and pseudo-acoustic modes from a nearby experiment (Fichtner *et al.* 2023b). Their distributions are displayed in Fig. 2 together with a comparison of observed and computed phase velocities. For the calculation of sensitivity kernels $G_i^{(n)}(\mathbf{m})$ for surface wave dispersion measurements, we employ the classic approach of Takeuchi & Saito (1972). Motivated by the general difficulty of constraining density with seismic data, we select ρ as the target parameter that we wish to resolve, that is, $\rho(z) = m^{(0)}(z)$. Simultaneously, we seek to minimize leakage from $v_s = m^{(1)}(z)$ and $v_p = m^{(2)}(z)$.
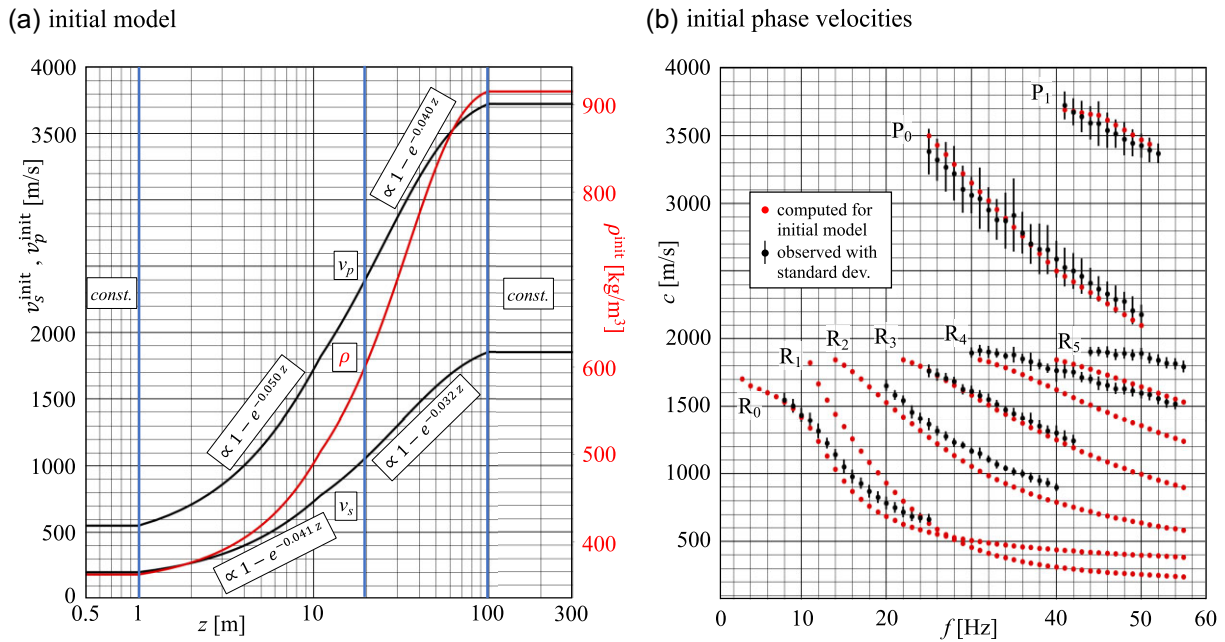
Figure 2. Initial model and data fit. (a) Depth distributions and functional forms of the initial density (red), *S*-wave speed and *P*-wave speed (black). (b) Comparison of computed phase velocities (red) and observed phase velocities c (black) with estimated standard deviations of their errors (black bars). Phase velocity measurements were made in intervals of 1 Hz. Figure adapted from Fichtner *et al.* (2023b).

4.1 Examples for error-free data

Fig. 3 displays averaging kernels focused on a target depth of $z_0 = 50$ m. The differential phase-velocity data $d_i^{\text{obs}} = c^{\text{obs}}(f_i) - c(\mathbf{m}^{\text{init}})$ for a range of frequencies f_i include the measurements marked in Fig. 2(b) for Rayleigh modes 0, 2, 3, 4 and 5. For now, we neglect the error term in eq. (25) by setting $\gamma = 0$.

Varying the weights β_1 and β_2 reveals their role in balancing the deltaness and minimization criteria in eqs (9) and (10). Choosing small values, $\beta_1 = \beta_2 = 10^{-3}$, produces a narrow averaging kernel for density, $A_{\rho'}$, with an averaging length $s_0 = 62$ m. However, averaging kernels for *S*-wave speed, $A_{v'_s}$ and *P*-wave speed, $A_{v'_p}$, are around two orders of magnitude larger.

Increasing the minimization weight for v_p to $\beta_2 = 10^3$ has the desired effect of suppressing the *P*-wave averaging kernel to values that are nearly five times smaller than the maximum of the density averaging kernel. While the width of the peak around $z_0 = 50$ m in the density averaging kernel $A_{\rho'}$ is nearly unchanged, the formal averaging length s_0 has increased by nearly a factor of ten to 454 m. On the one hand, this result reflects the presence of a long tail at greater depths. On the other hand, however, it illustrates that the deltaness criterion (5) is useful for optimization but not necessarily for measuring the width of an averaging kernel, because the quadratic term gives excessive weight to even small contributions far from the target depth z_0. This deficiency of the deltaness criterion, discussed in more detail in Section 6, must be kept in mind when interpreting actual values of s_0.

A similar effect occurs when β_1 is increased to 10^3, while keeping $\beta_2 = 10^{-3}$. Finally, increasing both β_1 and β_2 to 10^3 reduces the amplitude of the v_s and v_p averaging kernels below that of the density kernel. However, the averaging length scale for ρ increases to several kilometres.

4.2 Influence of higher surface-wave modes

The EastGRIP data set shown in Fig. 1(b) is distinguished by the presence of an unusually large number of clearly visible Rayleigh wave overtones. While the benefit of overtones for improved depth sensitivity and resolution of *S*-wave speed is well-known (e.g. Takeuchi & Saito 1972; Stutzmann & Montagner 1993), their ability to aid in the estimation of density structure has not been explored systematically.

Fig. 4 presents such an analysis for averaging kernels based on an increasing number of overtones and two choices of the minimization weights ($\beta_{1,2} = 10^{-3}$ in black, $\beta_{1,2} = 10^3$ in red). Using fundamental-mode measurements alone produces density averaging kernels $A_{\rho'}$ that generally do not resemble a positive-definite single-peak distribution. When the second overtone is included, such a distribution emerges, but only when the averaging kernels for v_s and v_p are not simultaneously forced to zero, that is, for the small minimization weight $\beta_{1,2} = 10^{-3}$. From the fourth overtone onwards, it becomes possible to obtain a bell-shaped averaging kernel for density around $z_0 = 50$ m, while reducing the amplitudes of the v_s and v_p kernels below the amplitude of $A_{\rho'}$.
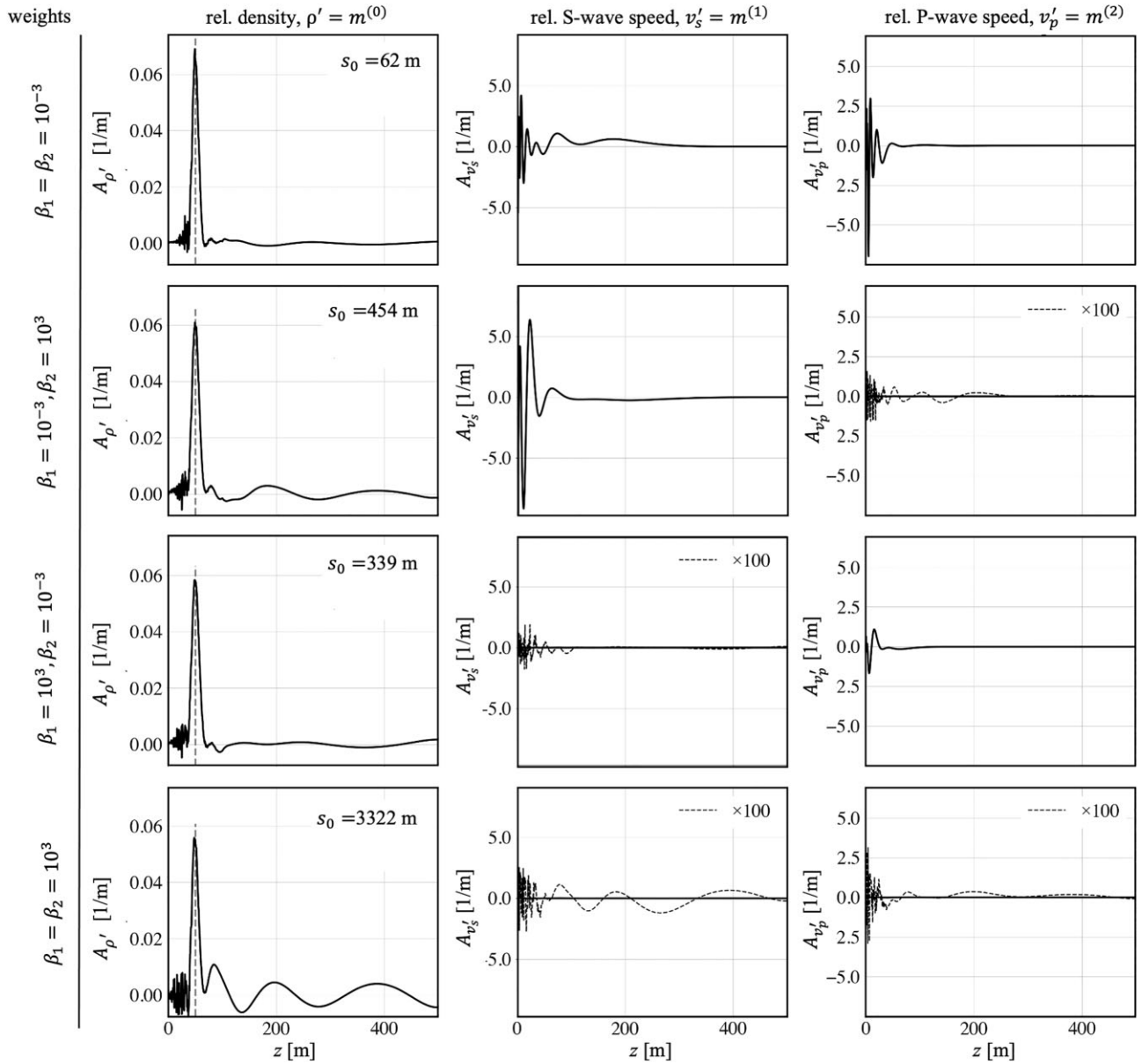
Figure 3. Averaging kernels for relative density ρ' (left), relative S-wave speed v'_s (centre) and relative P-wave speed v'_p (right) based on phase velocity measurements of the Rayleigh modes 0, 2, 3, 4 and 5 shown in Figs 1(b) and 2(b). The influence of measurement errors is ignored by setting $\gamma = 0$. Different rows correspond to different choices of the minimization weights β_1 and β_2 that balance the effects of focusing the density kernel $A_{\rho'}$ around the target depth $z_0 = 50$ m and minimizing the S-wave and P-wave kernels, $A_{v'_s}$ and $A_{v'_p}$. In the top row ($\beta_1 = \beta_2 = 10^{-3}$), the emphasis is on focusing $A_{\rho'}$, which produces high-amplitude $A_{v'_s}$ and $A_{v'_p}$, that is, good resolution and strong interparameter trade-offs. In the bottom row, the weights are increased to $\beta_1 = \beta_2 = 10^3$, thereby producing small $A_{v'_s}$ and $A_{v'_p}$ at the expense of poor spatial resolution for ρ. For better visibility, amplified versions of some kernels are plotted as dashed curves, with amplification factor shown in top right in the second and third column from left.

4.3 Influence of the error term

Including the error term in eq. (24) introduces another factor, in addition to interparameter leakage, that needs to be balanced against deltaness. Choosing large values for the error weight γ gives preference to achieving small errors in the inferred averages, at the expense of reducing deltaness and increasing trade-offs between model parameters.

This behaviour is illustrated in Fig. 5 using the EastGRIP data up to the fifth Rayleigh overtone. Again, the target depth in the examples is $z_0 = 50$ m, and the minimization weights are set to $\beta_{1,2} = 10^{-3}$ (black) and $\beta_{1,2} = 10^3$ (red). The case $\gamma = 10^{-8}$ is nearly identical to the error-free scenarios in Figs 3 and 4. As before, the density averaging kernel is dominated by a bell-shaped peak centred around ~ 50 m depth. For $\beta_{1,2} = 10^3$, the v_s and v_p kernels are oscillatory with amplitudes below the maximum of $A_{\rho'}$. As γ increases, the reduction of errors in the density averages becomes more important. Consequently, the peak of $A_{\rho'}$ becomes successively broader and skewed, and the amplitudes
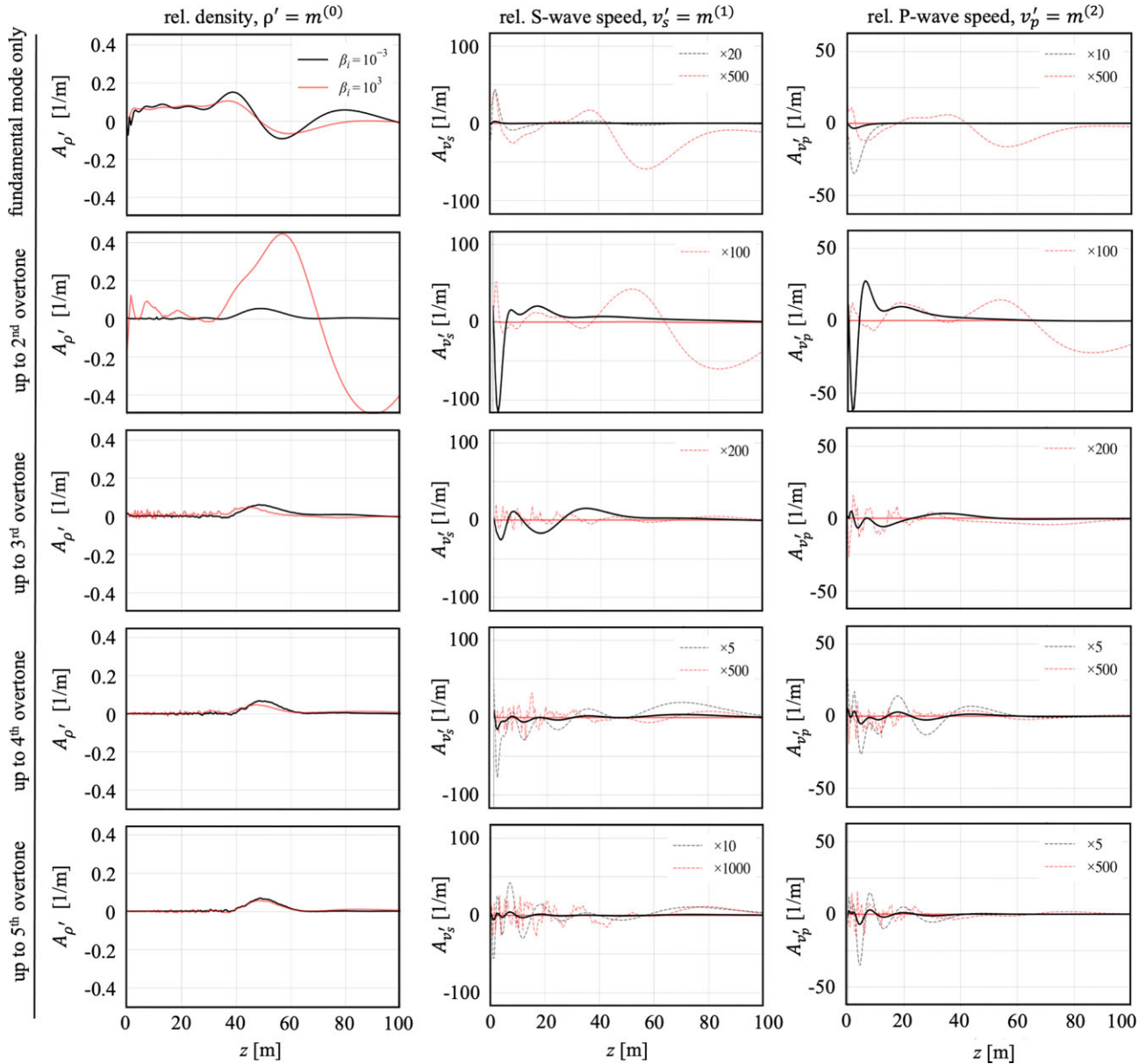
Figure 4. Effect of higher Rayleigh wave modes on the resolvability of density structure. Averaging kernels are shown for an increasing number of overtones from top to bottom and for two different choices of the minimization weights ($\beta_{1,2} = 10^{-3}$ in black, $\beta_{1,2} = 10^3$ in red). For better visibility, amplified versions of some kernels are plotted as dashed curves, with the amplification indicated in the top right.

of $A_{v_s'}$ and $A_{v_p'}$ increase. At $\gamma = 1$, $A_{\rho'}$ has acquired a complex shape that is not useful for the inference of density averages near 50 m depth, and interparameter leakage would dominate the computed averages.

Fig. 6 complements the kernel examples in Figs 3 and 5 by showing the dependence of the averaging length in $A_{\rho'}$ on the minimization weights $\beta_{1,2}$ and the error weight γ. Increasing any of these parameters is compensated by an increased averaging length, that is, reduced deltaness. The freedom to choose specific values for these parameters can either be used as a tool to produce a suite of useful models with different resolution properties, or to achieve a pre-defined error level or resolution length.

5 INVERSION OF EASTGRIP SURFACE-WAVE DISPERSION DATA

5.1 Inversion without prior knowledge

The tuning parameters β_1, β_2 and γ generate a continuum of model averages $\langle \mathbf{A}(\beta_1, \beta_2, \gamma), \mathbf{m} \rangle$, with different balances between resolution, errors and interparameter trade-offs. While none of the infinitely many alternatives are *per se* wrong, some will be more useful than others in the context of a specific application. The notion of useful may, for instance, be defined by specific target values for resolution and errors or by the requirement that averaging kernels are approximately bell-shaped.
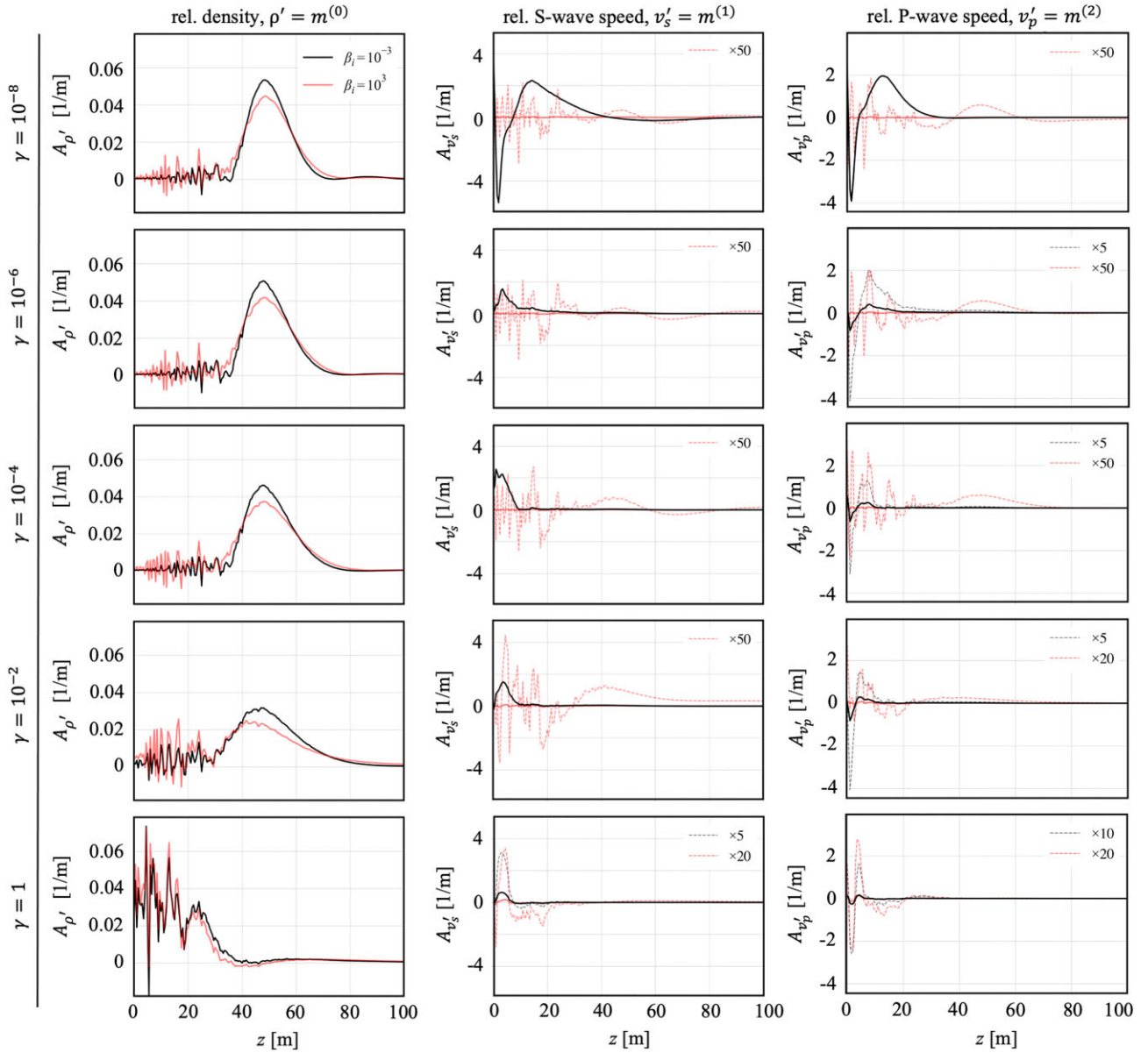
Figure 5. Illustration of averaging kernels for varying values of the tuning parameter γ that balances deltaness and interparameter leakage against the error standard deviation of the inferred averages. Large values of γ give preference to smaller uncertainties, and vice versa. All averaging kernels are for the target depth $z_0 = 50$ m and all modes up to the fifth overtone. Black curves correspond to $\beta_{1,2} = 10^{-3}$ and red curves to $\beta_{1,2} = 10^3$.

For illustration, we present an inversion of the EastGRIP dispersion data in Fig. 2(b) up to the fifth Rayleigh overtone, initially without the incorporation of any prior knowledge. We fix $\beta_1 = \beta_2 = 10^3$ and $\gamma = 10^{-2}$, loop from $z_0 = 0$ to $z_0 = 100$ m and repeat the kernel optimization and average computation for each depth and each model parameter individually. The complete suite of average models for arbitrary choices of tuning parameters can be constructed using the dispersion data and inversion codes found on the software webpages of the ETH Seismology & Wave Physics Group (www.swp.ethz.ch).

Fig. 7 shows the resulting depth-dependent averages for ρ, v_s and v_p, complemented by their respective error standard deviations and averaging lengths. For our choice of tuning parameters, averaging lengths are nearly constant in the metre range below $\sim$10 m depth but then increase rapidly to more than 100 m at 100 m depth. The error standard deviation shows a similar behaviour, indicating that density is not well constrained by the data alone below $\sim$10 m. In contrast, the v_s averaging length at a certain depth z is usually around $z/2$, similar to inversions where trade-offs with ρ and v_p are not minimized (Fichtner *et al.* 2023b). Error standard deviations for v_s are on the order of 10 m s^{-1}, that is, a few percent of the inferred v_s averages. Also for v_p error standard deviations are at the percent level, but averaging lengths are as unfavourable as for ρ below $\sim$10 m depth, which can be associated with more rapid evanescence of the P-wave content of the modes with depth.

A more detailed and informative image of spatial resolution for density is shown in Fig. 8 in the form of resolution kernels $A_{\rho'}$ for different target depths z_0. To depths of $\sim$75 m, approximately bell-shaped averaging kernels can be constructed. At 50 m depth, the half-width
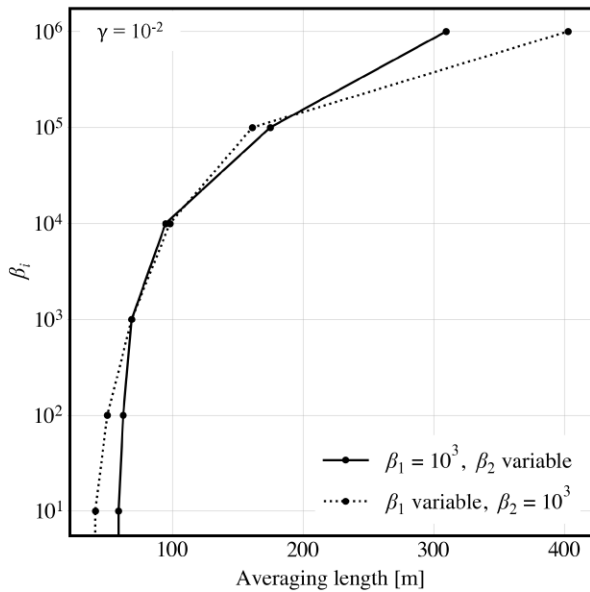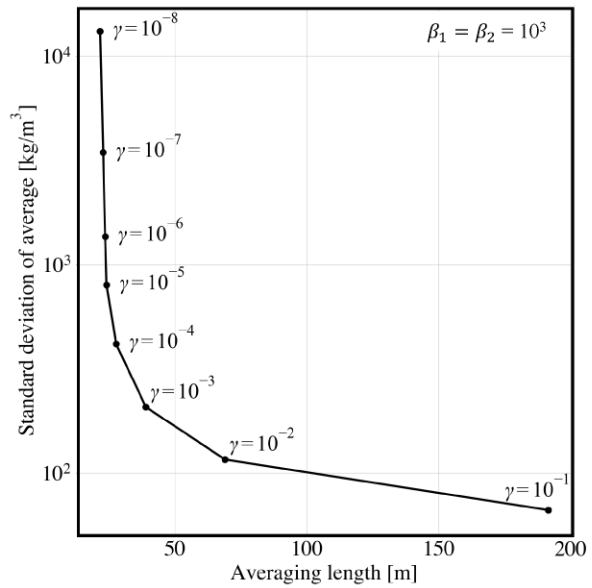
Figure 6. Dependence of averaging length on tuning parameters $\beta_{1,2}$ and γ. (a) Variation of the minimization weights β_1 (solid) and β_2 (dashed). Increasing $\beta_{1,2}$ gives more weight to the reduction of interparameter leakage, which is compensated by an increasing averaging length in $A_{\rho'}$. (b) Variation of the error weight γ for fixed $\beta_1 = \beta_2 = 10^3$. Forcing errors in the density averages to become smaller by increasing γ also leads to an increased averaging length.
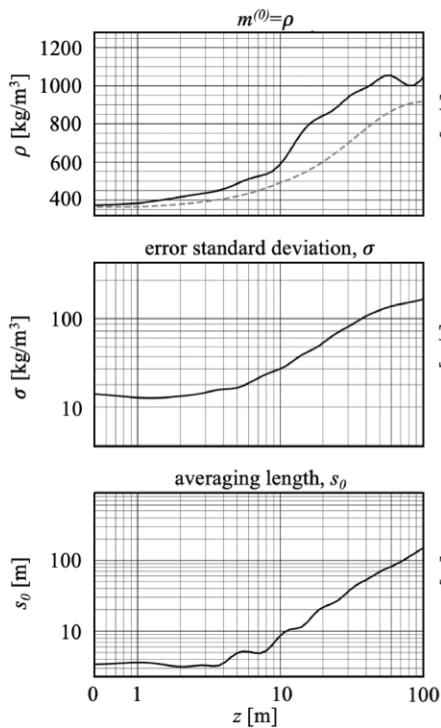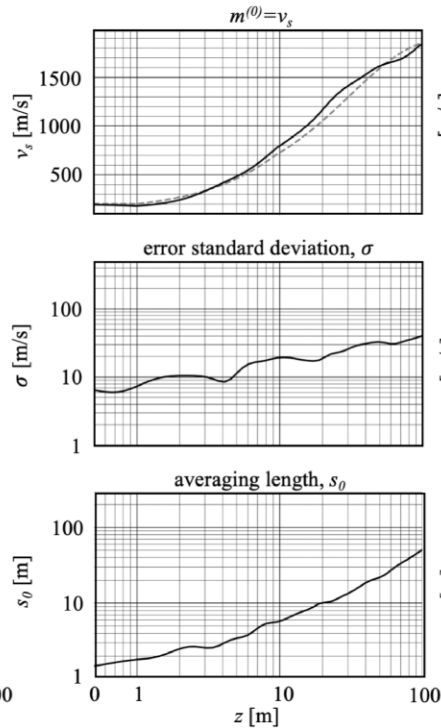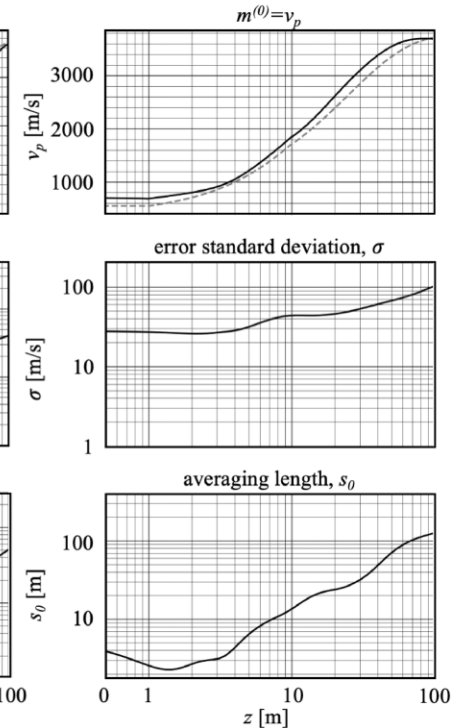

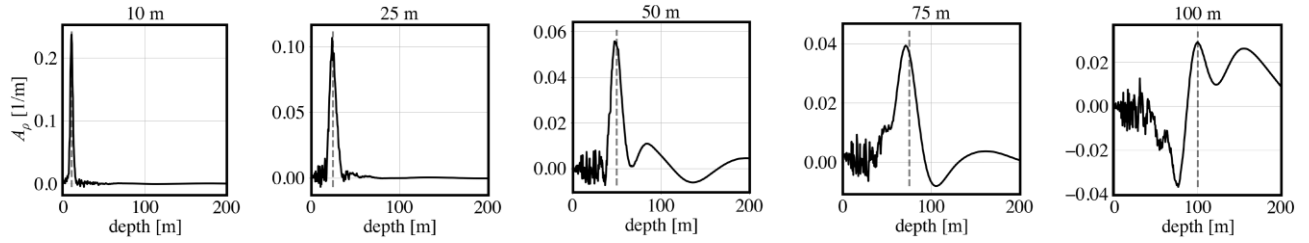**Figure 7.** Depth-dependent averages, error standard deviations σ and averaging lengths s_0 for (a) ρ, (b) v_s and (c) v_p and the tuning parameter values $\beta_1 = \beta_2 = 10^3$ and $\gamma = 10^{-2}$. The initial model, around which sensitivity kernels are linearized, is shown as grey dashed curves.

Figure 8. Density averaging kernels $A_{\rho'}$ for different target depths, marked by vertical dashed lines. The tuning parameters are equal to the ones used for the inversion shown in Fig. 7.
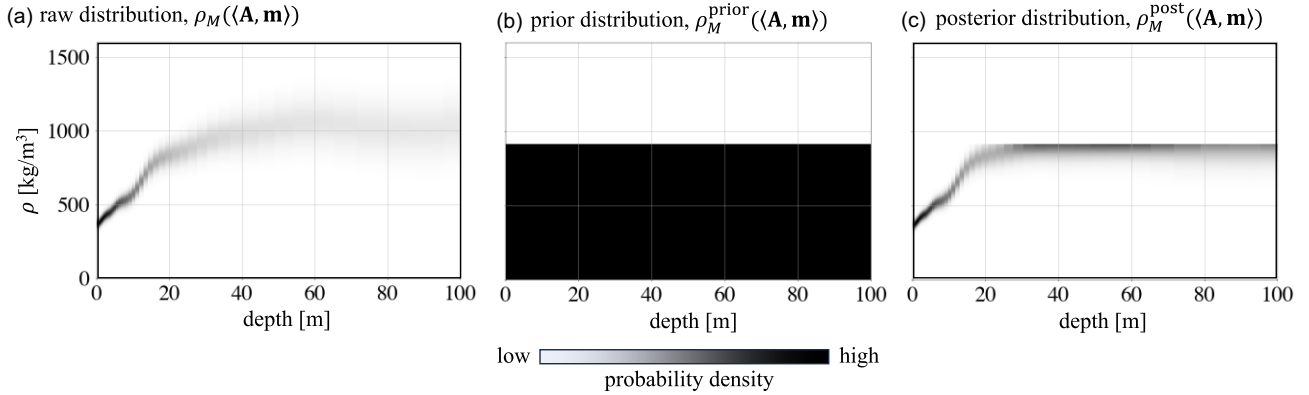
Figure 9. Incorporation of prior information on density. (a) The raw distribution $\rho_M(\langle \mathbf{A}, \mathbf{m} \rangle)$ is the result of plain Backus–Gilbert inversion, cast in the form of a probability density. (b) The prior distribution $\rho_M^{\mathrm{prior}}(\langle \mathbf{A}, \mathbf{m} \rangle)$ encodes the hard constraint that the density of ice must not exceed 917 kg m^{-3}. (c) The posterior distribution $\rho_M^{\mathrm{post}}(\langle \mathbf{A}, \mathbf{m} \rangle)$ is the conjunction of $\rho_M(\langle \mathbf{A}, \mathbf{m} \rangle)$ and $\rho_M^{\mathrm{prior}}(\langle \mathbf{A}, \mathbf{m} \rangle)$. Note that the region of high probability density located between 25 and 65 m is due to integrated averaging kernels for each depth being re-normalized to one following the conjunction of (a) and (b).

of $A_{\rho'}$ is on the order of 10 m, that is, significantly smaller than the resolution length $s_0 \approx 50$ m derived from the deltaness criterion. Similarly, at 100 m depth, the kernel half-width is around 20 m, in contrast to $s_0 \approx 100$ m.

5.2 Incorporating prior knowledge

While providing depth-dependent averaged densities and their uncertainties in the form of standard deviations, the results from Fig. 7(a) contradict hard prior knowledge that the density of ice does not exceed 917 kg m^{-3} (e.g. Petrenko & Whitworth 2002). The formalism developed in Section 2.4 allows us to incorporate this non-Gaussian *a priori* information via the conjunction of two probability densities, the raw distribution $\rho_M(\langle \mathbf{A}, \mathbf{m} \rangle)$ and the prior distribution $\rho_M^{\mathrm{prior}}(\langle \mathbf{A}, \mathbf{m} \rangle)$. This process is illustrated in Fig. 9.

Under the assumption of Gaussian data errors, $\rho_M(\langle \mathbf{A}, \mathbf{m} \rangle)$ is a Gaussian with the mean shown in the top panel and the standard deviation shown in the middle panel of Fig. 7(a). The prior $\rho_M^{\mathrm{prior}}(\langle \mathbf{A}, \mathbf{m} \rangle)$ takes the form of a step function, independent of depth. Combining both yields the posterior $\rho_M^{\mathrm{post}}(\langle \mathbf{A}, \mathbf{m} \rangle)$, which places the maximum-likelihood values near 917 kg m^{-3} for depths below 30 m. This result is a consequence of the poorly focused averaging kernels below this depth, as shown in Fig. 8.

6 DISCUSSION

We present a probabilistic multiparameter variant of Backus–Gilbert inversion that enables the incorporation of prior knowledge about inferred model properties. It rests on (i) the solution of an optimization problem that maximizes deltaness for the parameter of interest, while minimizing trade-offs with all other parameters, and (ii) the derivation of a probability density of inferred averages based on a probability density of observational errors. To illustrate the way the method works, we applied it to a multimode Rayleigh wave dispersion data set collected on the Northeast Greenland Ice Stream. In the following we discuss remaining issues concerning available seismic constraints on firn density structure, the translation of the method to the Earth's crust and mantle and limitations of the methods.

6.1 Constraints on firn density structure

In addition to the method itself, the most important result of this work is that averaging kernels for 1-D density structure that approximate a delta function can be obtained when Rayleigh wave dispersion measurements of at least the first three overtones are available. This, in

principle, enables the replacement of empirical conversions from wave speed to density for firn (e.g. Kohnen 1972; Diez *et al.* 2014) by direct inferences of the density distribution using seismic data.

What still limits this replacement in practice are the data errors, encoded in the data covariance matrix $\mathbf{C}_D$. For the specific example of the EastGRIP data, a useful resolution of 1-D density structure can be achieved only in the upper 10 m, a depth that is still easily accessible by drilling. As illustrated by eq. (23), the error standard deviation σ of the inferred average is directly proportional to $\mathbf{C}_D$. Hence, keeping the averaging coefficients $\mathbf{a}$ fixed, we may reduce σ without compromising resolution, provided that the entries of $\mathbf{C}_D$ can be reduced. This may be achieved, for instance, through repeated experiments, though at the expense of a longer and more costly data acquisition.

The inversion results in Section 5 highlights the need to incorporate prior information, as plausible densities may be beyond one standard deviation of the density averages inferred only from the data. Translating the classic Backus–Gilbert method into a Bayesian framework allows us to include prior information easily via the conjunction of probability densities. While the formalism is clean, the construction of prior probability densities often has a subjective component that must be acknowledged explicitly.

Naturally, we are faced with a large number of potential solutions dependent on the subjective selection of tuning parameters in the inversion. The question then arises of how to better exploit this wide range of model outcomes. While the example in Section 5 takes $\beta_1 = \beta_2 = 10^3$ and $\gamma = 10^{-2}$ to produce a single model for illustration, the range of solutions could be explored by systematically varying the tuning parameters and producing a suite of models. Model solutions may be also steered by pre-defining targets for resolution or error levels.

6.2 Translation to density inversion in the Earth's crust and mantle

Inversions for 1-D and 3-D density structure in the Earth's mantle have been a topic of research and debate for several decades (e.g. Dziewoński & Anderson 1981; Kennett 1998; Ishii & Tromp 1999; Romanowicz 2001; Resovsky & Trampert 2002). Hopes that full-waveform inversion (Blom *et al.* 2017) or gradiometry (Faber & Curtis 2024) may improve our knowledge of density heterogeneities have so far not materialized in practice, though they are known to have significant effects on crustal wave propagation (Plonka *et al.* 2016). The EastGRIP example benefits from the easy observation of strong Rayleigh wave overtones. It is a consequence of the nearly exponential wave speed increase within the firn layer, which traps waves much more efficiently than the Earth's crust, where wave speeds increase with depth only by some tens of percent (e.g. Bassin *et al.* 2000). Hence, replacing classical scaling relations for crustal rocks, such as Gardner's law (Gardner *et al.* 1974), may be more challenging. At regional to global scales, Rayleigh wave observations up to the sixth overtone are available (e.g. Moulik *et al.* 2021), but their usefulness in constraining crustal and upper-mantle density structure remains to be explored.

6.3 Methodological limitation

Our application assumes that lateral variations in medium properties are negligible over the 3 km length of the DAS cable. While the large number of narrowly defined modes, shown in Fig. 1(b), supports this assumption, ice lenses created by melt water or firn thickness variations across the shear margin cannot be excluded. All variants of the Backus–Gilbert method rest on the assumption of a linear forward problem. Hence, when the forward problem is nonlinear, the governing equations must be linearized around a suitable initial model. It follows that the range of applicability of Backus–Gilbert methods is limited to problems where such a model can actually be found. Hence, this issue must be analysed on a case-by-case basis. For the EastGRIP data, an initial model that closely matches the observations could be estimated manually with some prior information on plausible firn structures (Fichtner *et al.* 2023b).

The quantification of spatial resolution in the form of a single number is generally challenging. While the deltaness measure s_0 defined in (5) may also be used as a measure of resolution length, it provides useful results only when the averaging kernel is similar to a boxcar function. In the presence of side lobes far from the target depth z_0, the numerical values of s_0 are unhelpful. Though numerous alternatives have been proposed (e.g. Backus & Gilbert 1970; Masters & Gubbins 2003), the generally impossible task of condensing a complicated function into one representative number suggests that a visual analysis of averaging kernels for different depths and sets of tuning parameters is likely to be the most powerful and useful approach.

Eq. (31) assumes the availability of prior knowledge on parameter averages and not on the actual parameters. As shown in the example in Section 5.2, prior knowledge on upper (and equivalently lower) bounds is easy to implement, as it translates one-to-one from the parameters to their averages. In cases where the averaging length is small compared to plausible spatial variations of the parameters, prior distributions for the parameters may be used to approximate prior distributions of parameter averages. For more complicated scenarios, the link between Backus–Gilbert inversion and Gaussian processes (e.g. Valentine & Sambridge 2019) may offer useful solutions.

7 CONCLUSIONS

We demonstrate an application of a Backus–Gilbert method that enables inversions for multiple parameter classes and the incorporation of arbitrary prior knowledge. The adapted method is based on a reformulation of standard Backus–Gilbert inversion results in the form of a

probability density that translates arbitrary error distributions in the data into a distribution of spatial model parameter averages. We apply the method to surface wave dispersion data collected by DAS on the Northeast Greenland Ice Stream, and demonstrate that density can be independently constrained to depths of around 10 m, provided that data include at least the third overtone. Resolution at greater depths is theoretically possible but practically limited by current data quality.

ACKNOWLEDGMENTS

DATA AVAILABILITY

All codes and data used in this research are openly and freely available on the software webpages of the ETH Seismology & Wave Physics Group (www.swp.ethz.ch).

REFERENCES

Amiri, S., Maggi, A., Tatar, M., Zigone, D. & Zaroli, C., 2023. Rayleigh-wave group velocities in Northwest Iran: SOLA Backus-Gilbert vs. Fast Marching tomographic methods, *Seismica,* **2,** doi:10.26443/seismica.v2i2.1011.

Arfken, G.B. & Weber, H.J., 2005. *Mathematical Methods for Physicists,* 6th edn, Elsevier.

Backus, G.E., 1970a. Inference from inadequate and inaccurate data, I, *Proc. Nat. Acad. Sci.,* **65,** 1–7.

Backus, G.E., 1970b. Inference from inadequate and inaccurate data, III, *Proc. Nat. Acad. Sci.,* **67,** 282–289.

Backus, G.E. & Gilbert, F., 1967. Numerical application of a formalism for geophysical inverse problems, *Geophys. J. R. Astron. Soc.,* **13,** 247–276.

Backus, G.E. & Gilbert, F., 1968. The resolving power of gross Earth data, *Geophys. J. R. Astron. Soc.,* **16,** 169–205.

Backus, G.E. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross Earth data, *Phil. Trans. R. Soc. London, A,* **266,** 123–192.

Bassin, C., Laske, G. & Masters, G., 2000. The current limits of resolution for surface wave tomography in North America, *EOS, Trans. Am. geophys. Un.,* **81,** F897, doi:10.1029/2017GC007370.

Blom, N., Boehm, C. & Fichtner, A., 2017. Synthetic inversions for density using seismic and gravity data, *Geophys. J. Int.,* **209,** 1204–1220.

Cobden, L.J., Trampert, J. & Fichtner, A., 2018. Insights on upper mantle melting, rheology, and anelastic behavior from seismic shear wave tomography, *Geochem. Geophys. Geosys.,* **19,** 3892–3916

Diez, A., Eisen, O., I. Weikusat, J.E., Hofstede, C., Bohleber, P., Bohlen, T. & Polom, U., 2014. Influence of ice crystal anisotropy on seismic velocity analysis, *Ann. Glaciol.,* **55,** 97–106.

Dziewoński, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.,* **25,** 297–356.

Faber, M. & Curtis, A., 2024. On seismic gradiometric wave equation inversion for density, *Geophys. J. Int.,* **237,** 1459–1489

Fichtner, A., 2021. *Lecture Notes on Inverse Theory,* Cambridge Open Engage.

Fichtner, A., Hofstede, C., Gebraad, L., Zunino, A., Zigone, D. & Eisen, O., 2023a. Borehole fibre-optic seismology inside the Northeast Greenland Ice Stream, *Geophys. J. Int.,* **235,** 2430–2441.

Fichtner, A., Hofstede, C., Kennett, B.L.N., Nymand, N.F., Lauritzen, M.L., Zigone, D. & Eisen, O., 2023b. Fiber-optic airplane seismology on the Northeast Greenland Ice Stream, *The Seismic Record,* **3,** 125–133.

Fichtner, A. et al., 2024. Seismic Tomography 2024, *Bull. seism. Soc. Am.,* **114,** 1185–1213.

Gardner, G.H.F., Gardner, L.W. & Gregory, A.R., 1974. Formation velocity and density–the diagnostic basics for stratigraphic traps, *Geophysics,* **39,** 770–780.

Helsen, M.M., van den Broeke, M.R., van de Wal, R.S.W., van de Berg, W.J., van Meigaard, E., Davis, C.H., Li, Y. & Goodwin, I., 2008. Elevation changes in Antarctica mainly determined by accumulation variability, *Science,* **320,** 1626–1629.

Ishii, M. & Tromp, J., 1999. Normal-mode and free-air gravity constraints on lateral variations in velocity and density of Earth's mantle, *Science,* **285,** 1231–1236.

Joughin, I., Smith, B.E. & Howat, I.M., 2018. A complete map of Greenland ice velocity derived from satellite data collected over 20 years, *J. Glaciol.,* **64,** 1–11.

Kennett, B.L.N., 1998. On the density distribution within the Earth, *Geophys. J. Int.,* **132,** 374–382.

Kennett, B.L.N. & Sambridge, M.S., 1998. Inversion for multiple parameter classes, *Geophys. J. Int.,* **135,** 304–306.

Kennett, B.L.N., Sambridge, M.S. & Williamson, P.R., 1988. Subspace methods for large inverse problems with multiple parameter classes, *Geophys. J. Int.,* **94,** 237–247.

Koelemeijer, P., Deuss, A. & Ritsema, J., 2017. Density structure of earth's lowermost mantle from stoneley mode splitting observations, *Nat. Commun.,* **8,** 15241, doi: 10.1038/ncomms15241.

Kohnen, H., 1972. Über die Beziehung zwischen seismischen Geschwindigkeiten und der Dichte in Firn und Eis, *Z. Geophys.,* **38,** 925–935.

Kosovichev, A.G., 1999. Inversion methods in helioseismology and solar tomography, *J. Comp. Appl. Math.,* **109,** 1–39.

Masters, G., 1979. Observational constraints on the chemical and thermal structure of the earth's deep interior, *Geophys. J. R. Astr. Soc.,* **57,** 507–534.

Masters, G. & Gubbins, D., 2003. On the resolution of density within the Earth, *Phys. Earth Planet. Int.,* **140,** 159–167.

Moulik, P. *et al.*, 2021. Global reference seismological data sets: multimode surface wave dispersion, *Geophys. J. Int.,* **228,** 1808–1849.

Nocedal, J. & Wright, S.J., 1999. *Numerical Optimization,* Springer.

Parrenin, F. *et al.*, 2012. On the gas-ice depth difference (Δdepth) along the EPICA Dome C ice core, *Clim. Past.,* **8,** 1239–1255.

Paterson, W., 1994. *The Physics of Glaciers,* 3rd edn, Pergamon.

Petrenko, V.F. & Whitworth, R.W., 2002. *Physics of Ice,* Oxford Univ. Press.

Pijpers, F.P., 1995. Inversions in astronomy using the SOLA method, in *Inverse Problems in Wave Propagation,* eds, Chavent, G., Sacks, P., Papanicolaou, G. & Symes, W. W., Springer.

Pijpers, F.P. & Thompson, M.J., 1992. Faster formulations of the optimally localized averages method for helioseismic inversions, *Astron. Astrophys.,* **262,** 33–36.

Plonka, A., Blom, N. & Fichtner, A., 2016. The imprint of crustal density heterogeneities on regional seismic wave propagation, *Solid Earth,* **76,** 1591–1608.

Resovsky, J. & Trampert, J., 2002. Reliable mantle density error bars: an application of the Neighbourhood Algorithm to normal-mode and surface wave data, *Geophys. J. Int.,* **150**(3), 665–672.

Rignot, E. & Mouginot, J., 2012. Ice flow in Greenland for the international polar year 2008-2009, *Geophys. Res. Lett.,* **39,** doi:10.1029/2012GL051634.

Romanowicz, B., 2001. Can we resolve 3D density heterogeneity in the lower mantle?, *Geophys. Res. Lett.,* **28,** 1107–1110.

Schlegel, R. *et al.*, 2019. Comparison of elastic moduli from seismic diving-wave and ice-core microstructure analysis in antarctic polar firn, *Ann. Glaciol.,* **60,** 220–230.

Snieder, R., 1991. A extension of Backus-Gilbert theory to nonlinear inverse problems, *Inv. Prob.,* **7,** 409–433.

Stutzmann, E. & Montagner, J.-P., 1993. An inverse technique for retrieving higher-mode phase velocity and mantle structure, *Geophys. J. Int.,* **113,** 669–683.

Takeuchi, H. & Saito, M., 1972. Seismic surface waves, *Methods Comput. Phys.,* **11,** 217–295.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* 2nd edition., Society for Industrial and Applied Mathematics.

Trampert, J., Deschamps, F., Resovsky, J. & Yuen, D., 2004. Probabilistic tomography maps chemical heterogeneities throughout the lower mantle., *Science,* **306,** 853–856.

Valentine, A.P. & Sambridge, M.S., 2019. Gaussian process models–I. A framework for probabilistic continuous inverse theory, *Geophys. J. Int.,* **220,** 1632–1647.

van den Broeke, M., 2005. Depth and density of the Arctic firn layer, *Geophys. Res. Lett.,* **32,** doi:10.1029/2005GL023247.

Yanovskaya, T.B., 2000. Development of methods for surface wave tomography based on the Backus-Gilbert approach, in *Selected Papers From Volume 32 of Vychislitel'naya Seysmologiya,* eds, Chowdhury D.K., Nyland E., Odom R., Sen M., Keilis-Borok V.I., Levshin A.L., Molchan G.M. & Naimark B.M., *Computational Seismology and Geodynamics,* vol. **7,** pp. 5–16, American Geophysical Union.

Zaroli, C., 2016. Global seismic tomography using Backus-Gilbert inversion, *Geophys. J. Int.,* **207,** 876–888.

Zaroli, C., 2019. Seismic tomography using parameter-free Backus-Gilbert inversion, *Geophys. J. Int.,* **218,** 619–630.

Zaroli, C., Koelemeijer, P. & Lambotte, S., 2017. Towards seeing the Earth's interior through unbiased tomographic lenses, *Geophys. Res. Lett.,* **44,** doi:10.1002/2017GL074996.