

A Bayesian framework with geology-informed structural priors for tomographic reconstruction and uncertainty quantification

H. Yazdanian¹, G. Hillers^{2,3}, Y. Lu² and B. Maboudi Afkham¹

¹Research Unit of Mathematical Sciences, University of Oulu, 90014 Oulu, Finland. E-mail: hassan.yazdanian@oulu.fi

²Institute of Seismology, Department of Geosciences and Geography, University of Helsinki, 00014 Helsinki, Finland

³Geothermal Energy and Reservoir Technology, Institute of Applied Geosciences, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

Accepted 2026 September 3. Received 2026 August 19; in original form 2026 March 11

SUMMARY

Tomographic reconstruction involves fundamental trade-offs between resolution, data coverage and parametrization. Conventional approaches commonly rely on regularized pixel-wise or cell-based parametrizations. These regularizations are valuable and can themselves be geology-informed when their assumptions match the target setting. However, they often encode structural information only indirectly through generic penalty terms and regularization weights. As a result, they can be less suited to questions in which the objective is to test a specific structural hypothesis. We develop a Bayesian framework for tomographic reconstruction and uncertainty quantification that formulates such hypotheses explicitly and assesses them through posterior uncertainty. As an illustrative application of the framework, we consider ambient noise tomography. The phase velocity reconstruction is reformulated to address a targeted geological hypothesis by introducing a parametrized structural prior. The prior is constructed from a Whittle–Matérn latent Gaussian field, combined with spectral dimensional reduction via a truncated Karhunen–Loève expansion and a differentiable nonlinear pushforward mapping that promotes approximately piecewise-constant velocity structures. The prior parameters, including the correlation length, smoothness, pushforward slope and velocity bounds, provide interpretable hyperparameters related to expected lateral scale, structural regularity, transition sharpness and plausible phase-velocity contrasts. In this formulation, the prior is used as a parametrized structural assumption informed by geological context, rather than as hard geological knowledge. The resulting parametrization substantially reduces dimensionality, making both maximum *a posteriori* estimation and Hamiltonian Monte Carlo sampling computationally feasible for the synthetic and observed data sets considered here. Synthetic experiments demonstrate accurate recovery of prescribed velocity contrasts and show that posterior uncertainty patterns reflect acquisition geometry and data coverage. Application to an ~ 1000 -sensor dense array data set produces structures consistent with the imposed geology-informed prior assumptions, while providing spatially resolved uncertainty estimates that identify regions of robust inference and remaining ambiguity. The framework offers a proof-of-concept formulation for goal-oriented, uncertainty-aware seismic tomography when prior information supports an explicit structural hypothesis that can be encoded through the proposed Whittle–Matérn latent-field parametrization.

Key words: Bayesian inference; Inverse theory; Seismic noise; Seismic tomography.

1 INTRODUCTION

Imaging the Earth's subsurface is fundamental to a wide range of scientific and societal applications, including understanding Earth's structure and dynamics, natural hazard assessment, resource exploration and environmental monitoring (J.M. Reynolds 2011; M.E. Everett 2013). These objectives are commonly addressed using geophysical imaging methods such as seismic, electromagnetic and gravity tomography, which indirectly infer subsurface structure from surface or near-surface measurements (K. Erkan 2008). This is often formulated as an inverse problem (R.L. Parker 1994), from which a global representation of the subsurface is reconstructed. While many conventional methods for solving the inverse problem can provide broadly plausible subsurface images, they are often primarily designed to recover a global model. Additional analysis may therefore be required to characterize targets of interest that are critical for answering specific scientific questions and supporting decision-making (X. Zhang & A. Curtis 2022; V.C. Tsai 2023). This limitation arises because the relevant features may be subtle, spatially localized or both. For example, the presence, geometry and extent of small-scale anomalies are of primary interest in resource exploration. Subsurface interfaces are central to geological and lithological interpretation, while fault zones provide essential constraints for seismic hazard assessment.

Ambient noise surface wave tomography (ANT) provides a representative example of such inference challenges. By analysing ambient seismic vibrations recorded by arrays of seismic stations, ANT aims to construct the spatial distribution of shear-wave velocities

associated with subsurface structural heterogeneity (N.M. Shapiro *et al.* 2005; G. Bensen *et al.* 2007). The method has been widely applied across diverse research contexts and spatial scales, along with recent methodological developments in data-driven inversion with quantitative error assessment using probabilistic approaches (T. Bodin *et al.* 2012; X. Zhang *et al.* 2018), direct 3-D inversion (H. Fang *et al.* 2015; X. Zhang *et al.* 2018) and wave-equation-based implementations (Y. Lu *et al.* 2020; K. Wang *et al.* 2021). Nevertheless, many ANT studies are primarily designed for exploratory imaging rather than for explicitly testing pre-defined structural hypotheses. As a result, additional interpretation may be required when the scientific objective is to infer specific properties of a target, such as its existence, geometry, or contrast.

V.C. Tsai (2023) has elaborated on these challenges by introducing a conceptual distinction between two imaging modes: discovery-mode and inference-mode. The former prioritizes exploratory visualization of subsurface structures that are consistent with the measurements, whereas the latter focuses on inferring specific features, such as their geometry, contrast, or even existence, together with a quantitative assessment of how strongly such inferences are supported by the measurements. In this sense, inference-mode imaging provides a conceptual framework for reformulating inverse problems in a way that directly addresses specific scientific questions of interest.

A necessary component of formulating inverse problems in an inference-mode setting is the explicit statement of prior assumptions and their relationship to the scientific question being addressed (V.C. Tsai 2023). In classical inverse problem formulations, regularization is often introduced to stabilize the reconstruction and to obtain a well-posed computational problem. Although regularization can be viewed as imposing prior assumptions on the solution, its primary role in many applications is computational and numerical stabilization. Typical examples include Tikhonov regularization and truncated singular value decomposition, which impose smoothness, flattening or energy constraints on the reconstructed images (I. Cabrera-Pérez *et al.* 2021). Sparsity-promoting penalties such as total variation (TV) regularization can also be useful when sharp contrasts or approximately piecewise-constant structures are expected (W. Jiang & J. Zhang 2017). Such assumptions are not intrinsically non-geological: When smooth velocity variations, sharp interfaces or piecewise-constant structures are expected, classical regularization can be appropriate and physically meaningful.

The limitation is rather that, in many standard formulations, the parameters controlling these regularizations are defined mainly in mathematical terms. A Tikhonov or TV regularization parameter, for instance, primarily controls the balance between data misfit and regularity and does not usually correspond uniquely to a geological quantity such as the expected size, geometry, contrast or spatial organization of a target structure. Consequently, their relationship to the specific geological hypothesis under investigation is often indirect, making interpretation in terms of the target feature properties less explicit.

This matters because prior knowledge and regularization should not be treated as conceptually identical, even when they share similar mathematical forms. Prior knowledge denotes information supported independently of the inversion, whereas regularization denotes modelling assumptions introduced within the inversion framework. As said, in discovery-mode imaging, such assumptions are often used primarily to make the inverse problem well-posed and to obtain a plausible global reconstruction. In inference-mode imaging, their role should be made explicit in relation to the geological or structural hypothesis being tested. For example, J.B. Muir & V.C. Tsai (2020) incorporate structural geological information, such as faults, basins and interfaces, directly into the inversion process. This shifts part of the modelling choice from generic stabilizing parameters towards parameters that can be related more directly to geological structure, such as interface geometry, characteristic length scale, roughness or material contrasts, while the remaining algorithmic parameters retain their role in stabilizing and solving the inverse problem.

In addition, regularizing assumptions or prior models in many geophysical inversion frameworks remain closely tied to pixel- or voxel-based parametrizations. In many standard implementations, these assumptions are applied uniformly across the model domain, which may not fully reflect spatially varying data coverage or known geological heterogeneity (A.P. Valentine & M. Sambridge 2018). This can contribute to oversmoothed reconstructions in which fine-scale, geologically or economically significant variations are suppressed (Y. Wang *et al.* 2025), thereby potentially limiting the recovery of localized anomalies of interest.

Several alternative parametrizations have been proposed to address these limitations. Harmonic or spectral representations have long been used in global seismic tomography, often with model complexity controlled through truncation beyond a prescribed maximum wavenumber (J. Trampert & R. Snieder 1996). Irregular-cell and Voronoi-based parametrizations provide another important class of approaches for reducing overparametrization and allowing spatially adaptive model complexity (M. Sambridge & Ó. Guðmundsson 1998; H. Zhang & C. Thurber 2005). Wavelet-based parametrizations have similarly been used to obtain multiscale representations of subsurface structure (R. Hawkins *et al.* 2019). These methods have played an important role in stabilizing inverse problems and improving the flexibility of tomographic models. Nevertheless, in many such formulations, the parameters controlling the representation still primarily describe mathematical or numerical properties, such as model complexity, spatial resolution, basis truncation, cell density or multiscale decomposition, and their connection to a specific geological hypothesis—particularly a localized structure with expected scale, geometry, contrast or interface sharpness—can remain indirect.

Besides moving part of the parametrization towards quantities with clearer geological interpretation, formulating inverse problems within an inference-mode framework also requires the explicit quantification of uncertainty. One way to achieve this is through a probabilistic formulation of the inverse problem, which provides uncertainty estimates for quantities of interest together with the reconstructed images (T. Bodin *et al.* 2012; C. Esteve *et al.* 2025b). Bayesian approaches offer a principled framework for uncertainty quantification (UQ; J.P. Kaipio & E. Somersalo 2005), and recent studies have demonstrated their potential for ANT (Y. Liu *et al.* 2025; X.

Zhao *et al.* 2026). However, the practical use of Bayesian UQ is strongly affected by the dimensionality of the chosen parametrization. This is particularly challenging for pixel- or voxel-based representations of high-resolution velocity fields, although multiscale and adaptive approaches, such as transdimensional Bayesian methods (T. Bodin & M. Sambridge 2009), can partially alleviate this difficulty. Thus, an inference-oriented parametrization should not only make the structural assumptions more explicit, but also provide a reduced representation that enables efficient uncertainty quantification for the target quantities of interest.

In this study, we present a framework for ANT that is designed with inference-mode objectives in mind. The aim is not to replace standard pixel-based or regularized tomography methods in general, but to provide an alternative formulation for cases where the target of interest can be expressed through an explicit structural hypothesis. In this formulation, the prior is used as a parametrized structural assumption informed by geological context, rather than as hard geological knowledge. While several elements of this approach, including prior modelling, reduced-order parametrization and probabilistic inference, have been explored individually in previous studies, here we integrate these components within a single framework that aligns them with clearly defined inferential objectives. Rather than claiming to recover all plausible velocity structures, the approach focuses on estimating targeted structural features from measurements under explicit and transparent modelling assumptions, while indicating a route towards computationally feasible applications at larger scales. Although demonstrated here using ANT, the core components of the framework are not ANT-specific and can be adapted to other tomographic inverse problems with appropriate problem-specific modelling choices. By enabling goal-oriented inference with quantified uncertainty, this formulation provides a more transparent basis for scientific interpretation in geophysical applications. The framework is documented by an open-source software package.

The paper is organized as follows: Section 2 formulates ANT as a Bayesian inverse problem, including the forward model, likelihood and the role of prior assumptions. Section 3 presents the proposed inference-oriented prior construction and dimensionality reduction strategy, detailing the prior model, its implementation and the reduced-order parametric representation. Section 4 describes the Bayesian inference procedures and computational implementation, including single-point estimation, posterior sampling and UQ. Section 5 evaluates the proposed framework through synthetic tests under different station geometries, followed by an application to real data recorded by approximately 1000 seismic sensors in a cratonic ore deposit environment. The real-data results are compared with a reference discovery-mode inversion. Section 6 discusses the implications of the results, the influence of prior hyperparameters and model reduction and the limitations and future extensions of the framework. Finally, Section 7 summarizes the main conclusions of the study.

2 THEORETICAL BACKGROUND

This section provides the theoretical background underlying the proposed inference-mode framework for ANT. In this study, we focus on the first stage of the commonly implemented two-step ANT approach for estimating a pseudo-3-D shear wave velocity model, that is, we obtain 2-D lateral variations of phase or group velocity from seismic measurements. Hereafter, ‘ANT’ refers to this stage. We first formulate ANT as a Bayesian inverse problem and establish the probabilistic foundations required for incorporating geological prior information and UQ.

2.1 ANT as a Bayesian inverse problem

We infer a spatially varying 2-D phase velocity field from measured traveltimes between station pairs. Let $v(\mathbf{x}) : \Omega \rightarrow \mathbb{R}^+$ denote the velocity field where $\Omega \subset \mathbb{R}^2$ is a closed and bounded domain with a Lipschitz-continuous boundary. We collect the measured traveltimes between m station pairs in $\mathbf{y}^{\text{obs}} \in \mathbb{R}^m$. These measurements are often noisy. Assuming additive Gaussian noise, we can model the measurements as

$$\mathbf{y}^{\text{obs}} = G(v) + \boldsymbol{\eta}, \quad (1)$$

where $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_{\text{noise}}^2 I_m)$ is the noise vector with standard deviation σ_{noise} and $G(v)$ is a physical model mapping the velocity field to traveltimes. Although an additive Gaussian noise model is commonly used in geophysical inverse problems, it is a simplifying assumption and may not fully represent measurement errors in practical applications.

A widely used forward model in ANT is (N. Rawlinson & M. Sambridge 2003; G. Nolet 2008)

$$[G(v)]_i := \int_{\Gamma_i} \frac{1}{v(\mathbf{x})} dl, \quad 1 \leq i \leq m. \quad (2)$$

Here i is the station pair index, Γ_i is the straight path connecting the stations, and dl is the differential arc length along the ray path. This forward model assumes the straight-ray approximation, which can be inaccurate in strongly heterogeneous media where ray bending or finite-frequency effects are significant.

The ANT inverse problem now comprises inferring a velocity field $v(\mathbf{x})$ from noisy measurements $\mathbf{y}^{\text{obs}}$. We cast this inverse problem in a Bayesian framework, where all quantities of interest are modelled as random variables and are therefore described by their corresponding probability density functions. Let $p(v)$ and $p(\boldsymbol{\eta})$ represent the probability density functions of the velocity field referred to as the *prior* density and the noise, respectively. The Bayesian solution to the ANT inverse problem is the posterior distribution, that

is, the probability distribution of the velocity field conditioned on measurement data. Using Bayes' rule in density form (A. Tarantola & B. Valette 1982; J.P. Kaipio & E. Somersalo 2005), the posterior density is written as

$$p(v | \mathbf{y}^{\text{obs}}) \propto p(\mathbf{y}^{\text{obs}} | v) p(v), \quad (3)$$

where $p(v | \mathbf{y}^{\text{obs}})$ and $p(\mathbf{y}^{\text{obs}} | v)$ are the posterior and likelihood density functions, respectively. In the following paragraphs, we describe two components of the posterior: likelihood and prior.

The likelihood density function $p(\mathbf{y}^{\text{obs}} | v)$ quantifies the agreement between the measured traveltimes and those computed by the forward model. Considering eq. (1), we can write

$$p(\mathbf{y}^{\text{obs}} | v) = p(\mathbf{y}^{\text{obs}} - G(v)) = p(\eta) \propto \exp\left(-\frac{1}{2\sigma_{\text{noise}}^2} \|\mathbf{y}^{\text{obs}} - G(v)\|^2\right), \quad (4)$$

where $\|\cdot\|$ is the Euclidean norm.

The prior density function $p(v)$ encodes our assumptions or geological knowledge about v . Designing a suitable prior that can be expressed as a probability distribution is often the most important but also the most challenging aspect of Bayesian inversion (J.P. Kaipio & E. Somersalo 2005). When geological interpretation is the objective, it is useful for the prior assumptions to be stated in terms that can be related to geological information. In this work, the prior is constructed to represent a simplified structural hypothesis about the subsurface by encoding spatial correlation, expected smoothness of geological interfaces and the presence of inhomogeneities with controlled spatial scale and texture. In applications targeting environments characterized by distinct geological units separated by sharp interfaces, the prior can be designed to promote an approximately piecewise-constant velocity field, corresponding to internally relatively homogeneous geological units separated by sharp boundaries. This modelling choice is not intended to be universally appropriate; rather, it is useful when such structures are supported by independent geological knowledge or by the scientific objective of the study. In this case, the prior makes the assumed structural hypothesis explicit, instead of leaving it implicit in a generic smoothness penalty or in subsequent interpretation of a reconstructed image.

The resulting posterior distribution $p(v | \mathbf{y}^{\text{obs}})$ combines data and prior information into a full probability distribution over admissible velocity fields. This probabilistic formulation enables quantitative UQ, which is a necessary component of inference-mode imaging. Posterior summaries such as variances and credibility intervals provide spatially resolved measures of confidence in the inferred structures. This explicit treatment of uncertainty allows subsurface interpretations to be assessed not only by visual plausibility, but also by posterior uncertainty conditional on the assumed model, prior and noise structure.

3 PRIOR MODELLING AND DIMENSIONAL REDUCTION

In this section, we introduce a prior distribution for the unknown velocity field v . Our objective is to construct a reduced-order structural prior model whose parameters can be related to the target feature being inferred, while remaining computationally practical. For the ore-deposit setting considered here and at the analysed period, the working structural hypothesis is that the lateral phase-velocity field can be approximated by a small number of relatively homogeneous regions separated by comparatively sharp transitions. This assumption is motivated by the expected presence of distinct formation units and intrusive blocks, but it remains a modelling hypothesis rather than a general claim about all ANT problems. This assumption enables us to focus on the interfaces between crustal regions rather than on the entire domain, which substantially reduces the effective number of degrees of freedom while retaining a direct connection between the prior parameters and the imposed structural hypothesis. When this type of structural hypothesis is not justified, smoother prior models or conventional smoothness-based regularization can provide a more appropriate representation of the expected velocity field.

To realize this modelling assumption, we adopt a multilayered prior construction based on a low-dimensional latent variable u that parametrizes the overall behaviour of the velocity field. The physical velocity field is then obtained by applying differentiable nonlinear pushforward mappings to u , which implicitly induce sharp transitions between internally homogeneous regions. This approach allows interface characteristics, such as smoothness or roughness, to be controlled indirectly through the prior and mapping parameters, without explicitly parametrizing the interfaces themselves.

3.1 Whittle–Matérn prior for geology-informed structural modelling

The latent variable u is a smooth representation of the velocity field v . A common mathematical model for representing spatially varying random fields is a Gaussian process (GP; C.E. Rasmussen & C.K.I. Williams 2005). A GP is characterized by a mean function and a covariance structure, which is often represented by a covariance kernel, a symmetric and positive definite function describing statistical dependence between spatial locations. Let $u(\mathbf{x}) : \Omega \rightarrow \mathbb{R}$ be a random function, and let $\mathbf{x}, \mathbf{x}' \in \Omega$ denote two points in the domain. The covariance kernel $C(\mathbf{x}, \mathbf{x}')$ can be defined such that $C(\mathbf{x}, \mathbf{x}') = \text{Cov}(u(\mathbf{x}), u(\mathbf{x}'))$, quantifying how strongly the values of u at those two points are correlated. If a random function u is fully characterized by a covariance kernel and a mean function $m(\mathbf{x})$, we write $u \sim \text{GP}(m, C)$.

In a finite-dimensional representation, that is, when u is discretized on a uniform grid to obtain a vector $\mathbf{u}$, the covariance kernel C can be discretized using standard kernel approximation techniques (e.g. the Nyström method, M.S. Tong & W.C. Chew 2020) to obtain a covariance matrix $\mathbf{C}$ for $\mathbf{u}$. The probability density function of $\mathbf{u}$ is then

$$p(\mathbf{u}) \propto \exp\left(-\frac{1}{2}(\mathbf{u} - \mathbf{m})^T \mathbf{C}^{-1}(\mathbf{u} - \mathbf{m})\right), \quad (5)$$

where $\mathbf{m}$ is a discretization of m on the same grid.

While the finite-dimensional formulation in eq. (5) clarifies the structure of the Gaussian prior and its associated probability density, explicit construction and factorization of the covariance matrix $\mathbf{C}$ become computationally prohibitive for large-scale problems. We therefore return to the continuous setting and specify the prior through its covariance kernel, enabling alternative representations that avoid an explicit covariance matrix formation.

A flexible covariance kernel whose parameters admit useful structural interpretations is the Whittle–Matérn (WM) kernel (J.B. Muir & V.C. Tsai 2020)

$$C(\mathbf{x}, \mathbf{x}') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\|\mathbf{x} - \mathbf{x}'\|}{\ell} \right)^\nu K_\nu \left(\frac{\|\mathbf{x} - \mathbf{x}'\|}{\ell} \right), \quad (6)$$

where Γ is the gamma function, K_ν is the modified Bessel function of the second kind (F.W. Olver 2010), σ^2 is an amplitude scale, ℓ is the correlation length controlling the range over which points of u retain strong correlation and ν is the smoothness parameter that interpolates between very smooth fields associated with large ν values and rough, exponential-type fields controlled by small ν . This parametrization provides direct control over the expected spatial scale and regularity of the latent structural field. These quantities can be related to geological expectations, such as characteristic feature size or interface roughness, making the WM kernel useful for inference-mode imaging when such assumptions are stated explicitly.

As said, direct covariance-based sampling of such priors is computationally demanding due to the need to factorize large covariance matrices (L. Roininen *et al.* 2014). To overcome this limitation, the WM prior can be approximately represented through a stochastic partial differential equation (SPDE), whose solution has a similar covariance structure (L. Roininen *et al.* 2014; M.M. Dunlop *et al.* 2017)

$$(\tau^2 - \Delta)^{(v+1)/2} u(\mathbf{x}) = w(\mathbf{x}), \quad (7)$$

where $\tau = 1/\ell$, $w(\mathbf{x})$ represents a Gaussian white noise stochastic process specified by (9) and Δ is the Laplacian operator. The SPDE formulation generates a sample by drawing a realization of white noise and solving the resulting elliptic problem, rather than sampling u directly from a mean function and covariance kernel. This approach is substantially more robust under discretization and benefits from decades of research on efficient and accurate numerical solvers for elliptic partial differential equations (A. Quarteroni & A. Valli 1994), making it well-suited for large-scale problems.

We propose using a Fourier spectral method (A. Quarteroni & A. Valli 1994) to solve eq. (7). The domain of interest is the unit square $\Omega = [0, 1]^2$ with periodic boundary conditions, under which the eigenfunctions of the Laplacian are the Fourier modes $\phi_{\mathbf{k}}(\mathbf{x}) = \exp(i\mathbf{k} \cdot \mathbf{x})$, with 2-D angular wavenumber vector $\mathbf{k} = (k_1, k_2) \in 2\pi\mathbb{Z}^2$. Although periodic boundaries may not be physically realistic, they greatly simplify the analysis and enable an efficient spectral representation of the SPDE.

In the numerical implementation, the continuous field $u(\mathbf{x})$ is evaluated on a uniform $N \times N$ grid, yielding a discrete vector $\mathbf{u}$. This discretization implicitly determines the spectral truncation, since only Fourier modes representable on the grid can be resolved. Specifically, for even N , the admissible frequencies are $k_i \in 2\pi\{-N/2, \dots, N/2 - 1\}$, yielding N^2 retained modes. Let $\hat{\mathbf{u}} \in \mathbb{C}^{N^2}$ denote the vector of Fourier coefficients of u and let $\mathbf{w} \sim \mathcal{N}(0, I_{N^2})$ be a realization of a standard normal random vector, where I_{N^2} denotes the identity matrix of size N^2 . The SPDE eq. (7) decouples in Fourier space as

$$(\tau^2 + \|\mathbf{k}\|^2)^{(v+1)/2} [\hat{\mathbf{u}}]_{\mathbf{k}} = [\mathbf{w}]_{\mathbf{k}} := w_{\mathbf{k}}, \quad (8)$$

where we implicitly assumed that the definition of $w(\mathbf{x})$ in eq. (7) is

$$w(\mathbf{x}) \approx \sum_{\mathbf{k}} w_{\mathbf{k}} \phi_{\mathbf{k}}(\mathbf{x}). \quad (9)$$

Here, $[\cdot]_{\mathbf{k}}$ denotes the component corresponding to the mode $\mathbf{k}$. The WM covariance implied by eq. (7) has the spectral density

$$S(\mathbf{k}) := (\tau^2 + \|\mathbf{k}\|^2)^{-(v+1)}. \quad (10)$$

A discrete realization of the latent field is obtained by filtering white noise in Fourier space,

$$[\hat{\mathbf{u}}]_{\mathbf{k}} = S(\mathbf{k})^{1/2} w_{\mathbf{k}},$$

followed by an inverse discrete Fourier transform,

$$\mathbf{u} = \mathcal{F}^{-1}(S^{1/2} \mathbf{w}),$$

where $\mathcal{F}^{-1}$ is efficiently implemented using the fast Fourier transform (FFT), yielding grid-based approximations of $u(\mathbf{x})$ (A. Quarteroni & A. Valli 1994). In the implementation, angular wavenumbers and τ are scaled by $1/\pi$ before evaluating the spectral weights, which are then normalized. The reported numerical input values of τ use this scaled convention.

Although this approach removes the need for matrix factorization, the prior stochastic complexity still depends on the resolution of the Fourier discretization, that is, the number of Fourier modes used to represent the random field. This can lead to a high-dimensional stochastic space. In the following section, we revisit the Karhunen–Loève (KL) expansion of $u(\mathbf{x})$, which provides a means to identify the most important subspaces for such priors effectively.

3.2 Parametric prior representation

3.2.1 Latent Gaussian parametrization

In the previous section, we introduced an SPDE that removes the need for matrix factorization of the covariance of $u(\mathbf{x})$. However, numerical exploration of this prior still requires working in a high-dimensional stochastic space that is determined by the resolution used to represent $w(\mathbf{x})$. We now describe the construction of an effective low-dimensional approximation by retaining only the most significant components of $\mathbf{w}$ in eq. (8).

Let us expand $u(\mathbf{x})$ in the Fourier basis $\phi_k(\mathbf{x})$, yielding

$$u(\mathbf{x}) = m(\mathbf{x}) + \sum_{\mathbf{k}} S(\mathbf{k})^{1/2} w_{\mathbf{k}} \phi_{\mathbf{k}}(\mathbf{x}), \quad (11)$$

where $S(\mathbf{k})$ is the WM spectral density defined in eq. (10). This expansion is referred to as the KL expansion of u . Let us truncate this expansion to the $N_{\text{KL}} \in \mathbb{N}$ modes with the largest spectral weights, indexed by $\mathcal{K}_{\text{KL}}$, and denote this approximation by $u^{\text{KL}}(\mathbf{x})$. We can express the expected squared L^2 error for this approximation as (I.A. Ibragimov & Y.A. Rozanov 2012)

$$\begin{aligned} \mathbb{E} \|u(\mathbf{x}) - u^{\text{KL}}(\mathbf{x})\|_{L^2}^2 &= \mathbb{E} \left\| \sum_{\mathbf{k} \notin \mathcal{K}_{\text{KL}}} S(\mathbf{k})^{1/2} w_{\mathbf{k}} \phi_{\mathbf{k}}(\mathbf{x}) \right\|_{L^2}^2 \\ &= \sum_{\mathbf{k} \notin \mathcal{K}_{\text{KL}}} S(\mathbf{k}) \mathbb{E} (w_{\mathbf{k}}^2) \|\phi_{\mathbf{k}}(\mathbf{x})\|_{L^2}^2 \\ &= \sum_{\mathbf{k} \notin \mathcal{K}_{\text{KL}}} S(\mathbf{k}), \end{aligned} \quad (12)$$

where equalities are due to orthogonality of the Fourier basis, and $\mathbb{E}(w_{\mathbf{k}}^2) = 1$ since $w_{\mathbf{k}} \sim \mathcal{N}(0, 1)$. Because the series $\sum_{\mathbf{k}} S(\mathbf{k})$ is convergent, N_{KL} can be chosen such that the series' tail in eq. (12) is smaller than a tolerance δ_{KL} . A typical choice for this tolerance is $\delta_{\text{KL}} = (1 - \rho)\sigma_u^2$, where $\sigma_u^2 := \sum_{\mathbf{k}} S(\mathbf{k})$ is the total variance of the random field u , and $0 < \rho < 1$. This Fourier-based KL reduction provides the low-dimensional parameter vector $\mathbf{w}^{\text{KL}}$ used in the Bayesian inversion in Section 4.

Since the coefficients $w_{\mathbf{k}}$ in eq. (11) are real-valued while the basis functions $\phi_{\mathbf{k}}$ are complex, eq. (11) generally produces complex valued $u(\mathbf{x})$. These imaginary components have no physical meaning and arise solely from the lack of Hermitian symmetry in the spectral coefficients. To obtain a real-valued field and to get a full expansion in a complete sin-cos basis, in contrast to a sine or cosine basis only, we define $\tilde{u}^{\text{KL}} := \sqrt{2} (\Re(u^{\text{KL}}) + \Im(u^{\text{KL}}))$, where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of u^{KL} , respectively. It can be shown that $\tilde{u}^{\text{KL}}$ has the same WM covariance structure as u^{KL} , ensuring that the resulting prior is physically meaningful (B. Maboudi Afkham *et al.* 2024, 2026).

Fig. 1 shows example prior realizations generated with the WM model using our Fourier-based sampling method. The examples illustrate how the WM hyperparameters affect the latent Gaussian field. Increasing τ , that is, decreasing the correlation length, yields smaller and less spatially coherent structures. In contrast, increasing ν increases the regularity of the latent field, producing smoother realizations with less small-scale roughness.

Importantly, the figure also highlights the effect of dimensional reduction. In this example, the physical velocity field $\mathbf{v}$ is defined on a 512×512 grid, so that the latent field $\mathbf{u} \in \mathbb{R}^{262144}$. The prior is parametrized by a reduced latent Gaussian vector $\mathbf{w}^{\text{KL}} \in \mathbb{R}^{N_{\text{KL}}}$, where N_{KL} is chosen to retain a prescribed fraction $\rho = 0.99$ of the total prior variance σ_u^2 . Although the full velocity field is discretized on 262 144 pixels, only a comparatively small number of KL modes N_{KL} is required to represent it. Depending on the hyperparameter choice, N_{KL} ranges from as few as 46 to at most 10 423. This corresponds to a reduction of more than one to three orders of magnitude in stochastic dimension, while preserving the essential spatial structure implied by the prior.

3.2.2 Pushforward mapping for approximate piecewise-constant structure

In the previous sections, we introduced Gaussian processes as prior distributions for smooth spatially varying fields. We also presented an efficient sampling strategy based on an SPDE formulation of the prior, along with an effective approximation method that truncates the associated stochastic space. This provides a low-dimensional latent representation of the velocity field in ANT. In this section, we

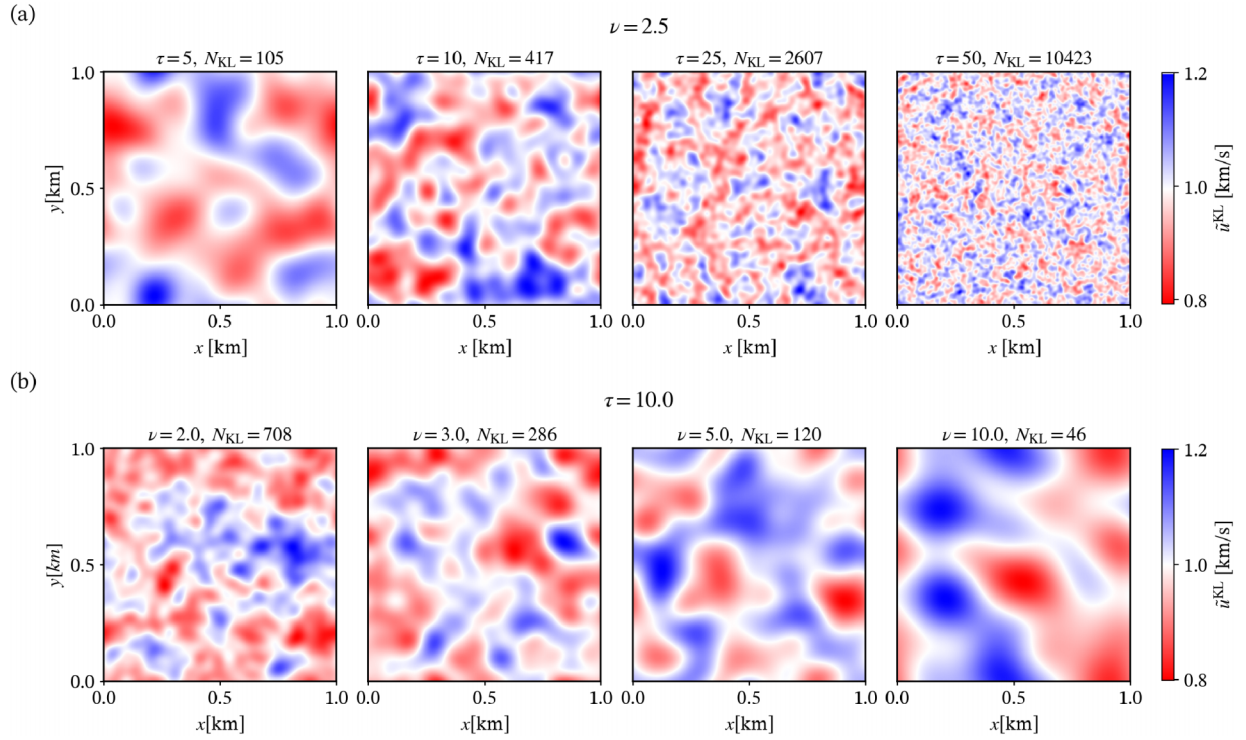


Figure 1. Influence of Whittle–Matérn hyperparameters ν and τ on spatial structure and effective dimensionality of Gaussian random-field realizations. (a) Effect of correlation length (inverse τ) on field structure at fixed smoothness $\nu = 2.5$. Smaller τ (longer correlation length) produces larger, more continuous features, while larger τ yields finer, less correlated patterns. The corresponding KL truncation levels retaining more than 99 per cent of the prior variance increase from $N_{\text{KL}} = 105$ to 10 423 (from left to right). (b) Effect of smoothness parameter ν at fixed inverse correlation length $\tau = 10$. Increasing ν produces smoother and more regular latent Gaussian fields, while smaller ν allows rougher field realizations with more small-scale variability. The retained KL modes decrease from $N_{\text{KL}} = 708$ to 46 (from left to right). Together, these results illustrate how the prior hyperparameters control both spatial structure and the degree of dimensional reduction relative to the full 512×512 discretization.

introduce pushforward maps, that is, nonlinear transformations, that map this latent space to velocity fields capable of representing geological features of the crust.

Many tectonically and economically interesting structures, including fault zones, volcanoes, rift systems and mineral resource formations, exhibit strong lateral velocity variations. Within the latent-variable framework introduced above, such structures can be synthesized by mapping the smooth latent field u to a piecewise-constant velocity field. A simple yet effective approach is to use the Heaviside function H (J.B. Muir & V.C. Tsai 2020)

$$v_{\text{HS}}(\mathbf{x}) = v_{\min} + (v_{\max} - v_{\min})H(\tilde{u}^{\text{KL}}(\mathbf{x})), \quad (13)$$

where $v_{\max}, v_{\min} \in \mathbb{R}^+$ denote the velocity levels for a bi-level field. Here, $v_{\min}$ and $v_{\max}$ are interpreted as effective phase-velocity levels at the considered period, not exact intrinsic lithological velocities, that can absorb unresolved layering and subgrid variability. This notation can also be naturally generalized to more levels of the velocity field (H. Antil & A.K. Saibaba 2024). However, in gradient-based inference strategies, including those considered in this work, the velocity field must be differentiable with respect to the latent variables. Since the Heaviside function is non-differentiable, we instead adopt a smooth approximation that preserves the piecewise-constant character of $v(\mathbf{x})$ while enabling gradient-based inference.

A convenient choice is the sigmoid mapping with a tunable slope parameter α (A. Iliev *et al.* 2017)

$$v(\mathbf{x}) = v_{\min} + (v_{\max} - v_{\min})\sigma_{\alpha}(\tilde{u}^{\text{KL}}(\mathbf{x})), \quad \alpha \gg 1, \quad (14)$$

where

$$\sigma_{\alpha}(\tilde{u}^{\text{KL}}(\mathbf{x})) = \frac{1}{1 + \exp(-\alpha \tilde{u}^{\text{KL}})}, \quad (15)$$

and α controls transition sharpness. As $\alpha \rightarrow \infty$, σ_{α} approaches H in eq. (13), producing nearly discontinuous interfaces between regions, while for finite α it remains smooth and differentiable. This yields a piecewise-smooth prior that approximates piecewise-constant geological units with sharp transitions, while preserving the computational advantages of Gaussian latent models. Large values of α localize the gradient of the sigmoid map near the zero level set of the latent field. This is consistent with the intended piecewise-constant representation, since gradients away from the transition zones are not needed to update the plateau values. However, α should not be chosen so large that the transition zone becomes unresolved on the computational grid.

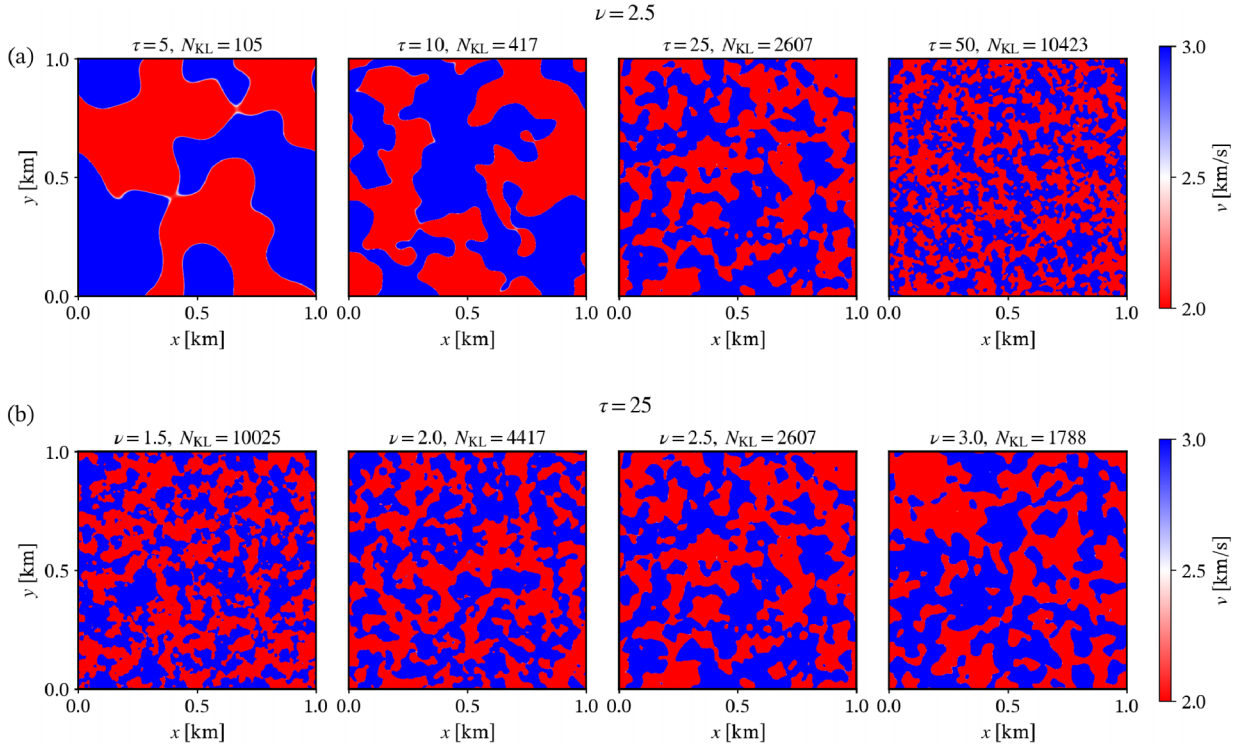


Figure 2. Effect of the pushforward mapping on WM prior realizations. Each panel shows a velocity-field realization obtained by transforming a WM latent field using the sigmoid pushforward map eq. (14) with $\alpha = 10^7$ and bounds $v_{\min} = 2 \text{ km s}^{-1}$, $v_{\max} = 3 \text{ km s}^{-1}$. (a) Effect of the inverse correlation length τ at fixed smoothness $\nu = 2.5$. Increasing τ reduces the correlation length and produces smaller, spatially less coherent velocity units, while also increasing the number of retained KL modes. (b) Effect of the smoothness parameter ν at fixed inverse correlation length $\tau = 25$. Increasing ν produces more regular latent fields and therefore smoother, less tortuous boundaries of the induced velocity units, while smaller ν yields rougher and more fragmented boundaries. The pushforward mapping converts the continuous WM latent field into near piecewise-constant velocity regions separated by sharp, numerically resolvable transitions.

Fig. 2 illustrates the effect of the sigmoid pushforward mapping with $\alpha = 10^7$ on WM prior realizations used in the inference-mode ANT framework. The first row shows the effect of the inverse correlation length τ for fixed smoothness $\nu = 2.5$. Increasing τ reduces the correlation length and produces smaller, spatially less coherent velocity units. The second row shows the effect of the smoothness parameter ν for fixed $\tau = 25$. Although changing ν can also affect the apparent correlation structure, its main role after the pushforward is to control the regularity and roughness of the induced unit boundaries. Smaller ν yields more irregular and fragmented boundaries, whereas larger ν produces smoother and more coherent boundaries.

These realizations exhibit sharp yet controlled transitions representative of near piecewise-constant structures. These examples show that τ and ν have related but distinct effects: τ primarily controls the characteristic lateral scale of the units, while ν controls the regularity of the latent field and, after pushforward, the roughness of the level-set boundaries. The corresponding values of N_{KL} are reported in each panel, illustrating how the effective stochastic dimension changes with the WM hyperparameters.

The resulting model provides both interpretational and computational benefits: It makes the imposed structural assumptions explicit through the WM covariance and pushforward mapping, enables efficient sampling and optimization in a reduced space and allows the prior family to be adjusted to represent a range of geology-informed structural hypotheses.

3.3 Choosing hyperparameters and discretization

3.3.1 WM prior hyperparameters

Recall that the WM covariance family associated with eq. (7) is characterized by a smoothness parameter ν and an inverse correlation length τ . In this section, we briefly discuss parameter ranges that are appropriate for geological applications and explain how these parameters interact with the spatial discretization of the velocity field.

When designing a WM prior, we must ensure both mathematical robustness of the hyperparameters and compatibility with the spatial discretization of the computational domain for the velocity field. Regarding robustness, the parameter choices of ν and τ must support the convergence of the KL expansion eq. (11) as required by eq. (12). This condition is satisfied for any $\nu > 0$ (A.M. Stuart 2010; M.M. Dunlop *et al.* 2017).

To address the compatibility of the hyperparameter choice with the computational discretization, we consider a uniform 2-D computational domain with mesh spacing $h_{\max}$, defined as the diagonal length of a discretization cell. A mesh with spacing $h_{\max}$ cannot resolve wavenumbers above $k_{\text{thresh}} := \pi/h_{\max}$ (A. Quarteroni & A. Valli 1994). We therefore require the fraction of spectral variance contained in frequencies $\|\mathbf{k}\| > k_{\text{thresh}}$ to be sufficiently small, that is,

$$\frac{\sum_{\|\mathbf{k}\| > k_{\text{thresh}}} S(\mathbf{k})}{\sum_{\mathbf{k}} S(\mathbf{k})} \leq \varepsilon, \quad (16)$$

where $\varepsilon > 0$ is a prescribed tolerance. Assuming that the mesh is sufficiently fine so that $k_{\text{thresh}} \gg \tau$, the ratio can be approximated asymptotically by

$$\frac{\sum_{\|\mathbf{k}\| > k_{\text{thresh}}} S(\mathbf{k})}{\sum_{\mathbf{k}} S(\mathbf{k})} \approx \left(\frac{\tau}{k_{\text{thresh}}} \right)^{2\nu}. \quad (17)$$

Substituting the frequency resolution condition $k_{\text{thresh}} = \pi/h_{\max}$ yields the constraint on τ

$$\left(\frac{\tau h_{\max}}{\pi} \right)^{2\nu} \leq \varepsilon, \quad \text{or equivalently} \quad (\tau h_{\max})^{2\nu} \leq \varepsilon, \quad (18)$$

where the constant factor $\pi^{-2\nu}$ has been absorbed into ε . This criterion is used here only as a spectral resolution guideline for the uniform-grid Fourier representation of the WM prior.

This relationship provides a principled way to balance computational complexity with frequency resolution. In practice, the smoothness parameter ν can be chosen based on external geological constraints, expert interpretation, or sensitivity analysis. For example, if independent information suggests relatively smooth or highly irregular interfaces, ν can be selected to reflect this expected regularity. When such information is unavailable, ν should be regarded as a modelling hyperparameter rather than as independently established prior knowledge. Given a fixed computational budget, reflected in the mesh refinement parameter $h_{\max}$, the parameter τ can then be selected to accommodate a prescribed minimum correlation length. Conversely, if a minimum correlation length is suggested by independent geological information or by the target structural hypothesis, the estimate above provides an upper bound on the admissible mesh spacing and thus caps the computational cost by prescribing an appropriate discretization resolution.

We remark that, in this work, the WM hyperparameters are assumed to be fixed and known. In practice, however, identifying appropriate values for these parameters is a challenging task and typically requires prior field investigations. Results by M.M. Dunlop *et al.* (2017) and B. Maboudi Afkham *et al.* (2024) suggest that inference schemes can be developed to simultaneously estimate both the unknown fields and the WM hyperparameters. Such approaches, however, remain at the frontier of current research and are beyond the scope of this paper. We leave the implementation of these methods into an ANT framework to future work.

3.3.2 Choosing pushforward hyperparameters

The pushforward mapping introduced in Section 3.2.2 involves additional hyperparameters α , $v_{\min}$ and $v_{\max}$ that control how the smooth latent field is transformed into a near piecewise-constant velocity field. The sigmoid slope parameter α governs the width of the transition zone between velocity levels and must be chosen to balance geological interpretability with numerical robustness for gradient-based inference schemes introduced in Section 4. Meaningful use of the pushforward mapping therefore requires consistency between this transition width and the spatial discretization.

The effective lateral transition width induced by the sigmoid mapping is inversely proportional to the product $\alpha |\tilde{u}^{\text{KL}}|$, where $|\tilde{u}^{\text{KL}}|$ denotes a representative amplitude of the latent field under the chosen WM hyperparameters τ and ν . To ensure that the reconstructed velocity transitions are both geologically sharp and numerically resolvable for gradient computation, this lateral transition width should include several mesh cells of size $h_{\max}$ in order to avoid discretization artefacts, leading to the practical condition

$$h_{\max} \lesssim \frac{1}{\alpha |\tilde{u}^{\text{KL}}|}.$$

In addition, the lateral transition width should remain small relative to the characteristic correlation length of the latent field, $1/\tau$, so that interfaces are sharp compared to the dominant geological length scales. Together, these conditions enforce a clear separation between numerical resolution, the numerical transition width of the sigmoid mapping, and the lateral geological structure represented in the phase-velocity map.

In addition to α , the velocity bounds $v_{\min}$ and $v_{\max}$ define the admissible range of seismic velocities and therefore impose strong prior constraints about the reconstructed field. In principle, these parameters could be inferred jointly with the latent variables within a hierarchical Bayesian framework. However, doing so requires explicit prior models for the velocity levels and introduces additional coupling between velocity contrast and the latent structural field. Under the finite-slope sigmoid parametrization used here, the traveltime likelihood also depends nonlinearly on $v_{\min}$ and $v_{\max}$, so these parameters cannot in general be eliminated analytically. We therefore fix them using external information and the calibration reported in Supplementary Material S2.4, while leaving their hierarchical inference for future work.

4 BAYESIAN INFERENCE AND COMPUTATION

Having the forward model, prior structure and reduced latent parametrization, we now describe the Bayesian inference procedures used to estimate the velocity field and quantify its uncertainty. The twofold goal is (i) to obtain a point estimate of the velocity field via maximum *a posteriori* (MAP) optimization, and (ii) to explore the posterior distribution to assess the range of models consistent with the data and prior assumptions. Operating in the latent space allows both optimization and sampling to be performed efficiently, enabling UQ for large-scale ANT data sets.

4.1 Parametrized posterior

With the parametrized, nearly piecewise-constant prior introduced in Section 3, we return to the Bayesian formulation in eq. (3). The velocity field v is represented as a deterministic transformation $\mathcal{T} : \mathbb{R}^{N_{\text{KL}}} \rightarrow \mathbb{R}^{N^2}$ of a low-dimensional latent variable $\mathbf{w}^{\text{KL}} \in \mathbb{R}^{N_{\text{KL}}}$ through the prior construction

$$\mathcal{T} : \mathbf{w}^{\text{KL}} \xrightarrow{f_{\text{KL}}} \tilde{\mathbf{u}}^{\text{KL}} \xrightarrow{g} v,$$

where f_{KL} denotes the truncated spectral construction of the WM latent field and g is the nonlinear pushforward mapping that produces the velocity field. Here, N_{KL} denotes the number of retained KL coefficients, and N^2 corresponds to the number of grid points used to discretize the 2-D velocity field. The prior is specified directly on the latent variable as $\mathbf{w}^{\text{KL}} \sim \mathcal{N}(\mathbf{0}, I_{N_{\text{KL}}})$, with $I_{N_{\text{KL}}}$ being the identity matrix of size N_{KL} , which induces the corresponding prior distribution on v through the transformation $\mathcal{T}$. The posterior distribution is therefore naturally expressed in terms of the latent variables as

$$p(\mathbf{w}^{\text{KL}} | \mathbf{y}^{\text{obs}}) \propto p(\mathbf{y}^{\text{obs}} | \mathcal{T}(\mathbf{w}^{\text{KL}})) p(\mathbf{w}^{\text{KL}}). \quad (19)$$

Using the likelihood defined in eq. (4), this posterior can be written explicitly as

$$p(\mathbf{w}^{\text{KL}} | \mathbf{y}^{\text{obs}}) \propto \exp\left(-\frac{1}{2\sigma_{\text{noise}}^2} \|\mathbf{y}^{\text{obs}} - G(\mathcal{T}(\mathbf{w}^{\text{KL}}))\|^2 - \frac{1}{2} \|\mathbf{w}^{\text{KL}}\|^2\right). \quad (20)$$

The posterior distribution over velocity fields $p(v | \mathbf{y}^{\text{obs}})$ is obtained as the transformation of $p(\mathbf{w}^{\text{KL}} | \mathbf{y}^{\text{obs}})$ under the mapping $\mathcal{T}$. This latent formulation forms the basis for the low-dimensional MAP estimation and the UQ in Sections 4.2 and 4.3, respectively.

4.2 Maximum *a posteriori* (MAP) estimation

To obtain a representative reconstruction of the velocity field, we compute the MAP estimate, that is, the maximizer of the posterior density eq. (20). Let $\Phi(\mathbf{w})$ denote the log-posterior up to an additive constant

$$\Phi(\mathbf{w}) := -\frac{1}{2\sigma_{\text{noise}}^2} \|\mathbf{y}^{\text{obs}} - G(\mathcal{T}(\mathbf{w}^{\text{KL}}))\|_2^2 - \frac{1}{2} \|\mathbf{w}^{\text{KL}}\|_2^2.$$

The MAP estimate is then given by the maximization problem

$$\mathbf{w}_{\text{MAP}} = \underset{\mathbf{w}}{\operatorname{argmax}} \Phi(\mathbf{w}^{\text{KL}}). \quad (21)$$

Since Φ is the log-posterior, its maximizer corresponds to the most probable latent representation of the velocity field under the posterior.

We solve the optimization problem eq. (21) using the limited-memory Broyden Fletcher Goldfarb Shanno (L-BFGS) algorithm, which requires only first derivatives, is well-suited for problems with hundreds of unknowns (R.H. Byrd *et al.* 1995), and is available in open-source software, for example, in the PyTorch library (A. Paszke *et al.* 2019). It is possible to explicitly define the gradient of the map $\mathbf{w}^{\text{KL}} \mapsto \Phi$ using the chain rule. However, gradients, in this work, are computed by the automatic differentiation of the PyTorch library (A. Paszke *et al.* 2019). As a stopping criterion for the L-BFGS method, we set a relative tolerance 10^{-6} for the gradient norm.

4.3 Posterior sampling and UQ

While the MAP estimate provides a single velocity field, it does not capture the range of alternative fields that are also consistent with the posterior. To quantify uncertainty and assess the credibility of reconstructed features, we compute posterior summaries such as the posterior mean and variance. These quantities are obtained through ergodic averages of samples drawn from the posterior distribution $p(\mathbf{w}^{\text{KL}} | \mathbf{y}^{\text{obs}})$. If $\{\mathbf{w}^{(s), \text{KL}}\}_{s=1}^{N_s}$ denotes a sequence of approximately independent posterior samples, the corresponding velocity-field samples are obtained through the transformation $v^{(s)} = \mathcal{T}(\mathbf{w}^{(s), \text{KL}})$. Posterior moments are then approximated by

$$\mathbb{E}[v(\mathbf{x})] \approx \frac{1}{N_s} \sum_{s=1}^{N_s} v^{(s)}(\mathbf{x}), \quad \operatorname{Var}[v(\mathbf{x})] \approx \frac{1}{N_s - 1} \sum_{s=1}^{N_s} (v^{(s)}(\mathbf{x}) - \mathbb{E}[v(\mathbf{x})])^2.$$

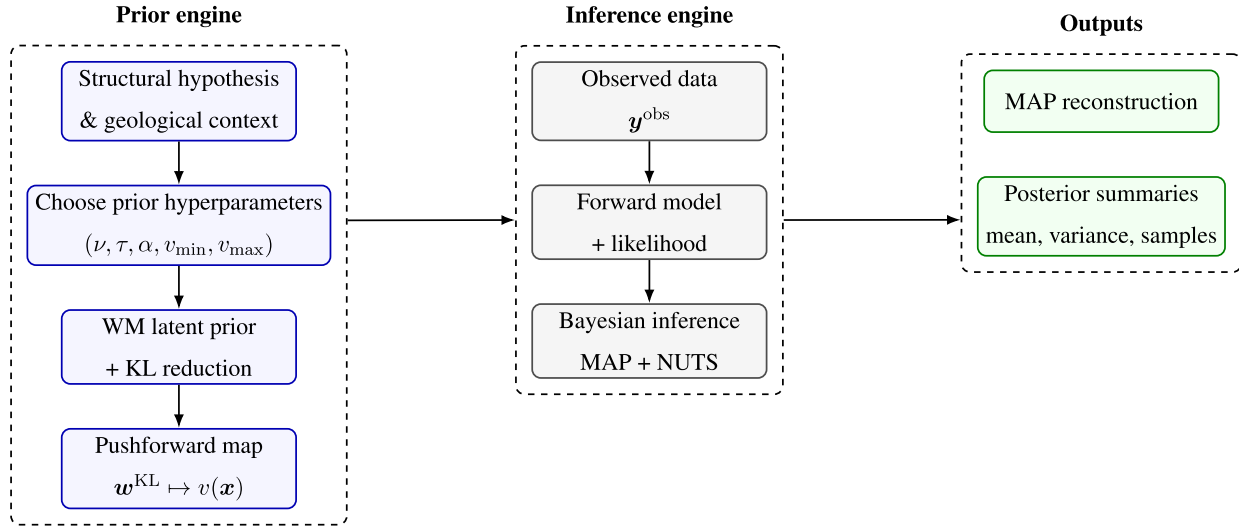


Figure 3. Schematic workflow of the proposed inference-mode framework, demonstrated here for ANT. The prior engine contains the method-specific structural prior model, which combines geological context, prior hyperparameter selection, a WM latent prior, KL reduction and a pushforward mapping to define the velocity parametrization. Once this parametrization is specified, the inference engine follows a modular Bayesian workflow based on the observed data, forward model, likelihood, MAP estimation and NUTS sampling. The outputs include both MAP reconstructions and posterior summaries, thereby supporting structural interpretation and uncertainty-aware decision-making.

These summaries provide spatially resolved measures of uncertainty, distinguishing regions strongly constrained by the data from those dominated by prior information.

To obtain the samples $\{\mathbf{w}^{(s),\text{KL}}\}$, Markov Chain Monte Carlo (MCMC) methods are used (J.P. Kaipio & E. Somersalo 2005). MCMC methods construct a Markov chain whose states approximate independent samples from a posterior distribution. The most basic of these methods is the Metropolis–Hastings algorithm, which is based on a Metropolis proposal step, followed by a Hastings acceptance/rejection scheme to create a Markov chain. Although this method is effective for posterior exploration, it is often computationally inefficient for the complex and nonlinear posteriors encountered in this paper.

To overcome these challenges, we propose using the *No-U-Turn Sampler* (NUTS; M.D. Hoffman & A. Gelman 2014), an adaptive variant of the Hamiltonian Monte Carlo (HMC; A. Fichtner *et al.* 2019). HMC augments the stochastic space with auxiliary momentum variables and interprets the resulting joint distribution through Hamiltonian dynamics. Proposed samples are generated by approximately simulating trajectories of constant Hamiltonian energy, for which efficient numerical integrators such as leapfrog schemes are available (E. Hairer *et al.* 2006). NUTS automatically adapts the trajectory length by detecting when the simulated path begins to double back on itself, a ‘U-turn’, thereby removing the need to manually tune the number of leapfrog steps and improving sampling efficiency. While NUTS is not specifically designed for infinite-dimensional posterior distributions, our numerical experiments suggest that it performs robustly for the experiments in the paper.

Applying NUTS requires only the negative log-posterior $-\Phi(\mathbf{w})$ and its gradient $-\nabla\Phi(\mathbf{w})$, where Φ is defined in Section 4.2 and the gradient is obtained via automatic differentiation. The resulting samples $\mathbf{w}^{(s)}$ are then mapped through $\mathcal{T}$ to yield corresponding velocity-field realizations $\mathbf{v}^{(s)}$, which form the basis of UQ.

4.4 Algorithm summary and implementation details

The complete workflow is summarized in Algorithm 1. This implementation uses automatic differentiation for both MAP estimation and NUTS trajectory integration. Both MAP estimation and NUTS sampling require initial values in the reduced KL coefficient space, denoted by $\mathbf{w}_0^{\text{MAP}}$ and $\mathbf{w}_0^{\text{NUTS}}$ in Algorithm 1, respectively. The specific choices used in each numerical experiment are stated in Section 5.

To complement the algorithmic description, Fig. 3 summarizes the modular structure of the proposed workflow. The prior engine contains the method-specific modelling component, where geological context and the target structural hypothesis are translated into a reduced latent parametrization of the velocity field. Once this parametrization is specified, the subsequent inference engine follows a standard Bayesian workflow based on forward modelling, likelihood evaluation, MAP estimation and posterior sampling. This separation emphasizes that the added modelling structure is concentrated in the prior construction, while the remaining inference workflow can in principle be combined with other prior or regularizing models.

All computations were performed on a MacBook Pro, Apple M-series GPU, 24 GB RAM, using Python 3.12.2, PyTorch 2.2.2 and Pyro 1.9.1 for posterior sampling. Four independent chains were run for each experiment, with N_{KL} latent dimensions determined by the retained spectral energy fraction ρ introduced in Section 3.2.1.

Algorithm 1: Inference-mode ANT

Input: Observed travel times $\mathbf{y}^{\text{obs}}$, station pairs $\{(i, j)\}$, prior hyperparameters $(\tau, \nu, \alpha, v_{\min}, v_{\max})$ based on prior geological knowledge, spectral energy fraction ρ , noise variance σ_{noise}^2 , MAP initialisation $\mathbf{w}_0^{\text{MAP}}$, NUTS initialisation $\mathbf{w}_0^{\text{NUTS}}$

Output: MAP velocity field $v_{\text{MAP}}(\mathbf{x})$, posterior mean and variance

1. WM prior: Construct the spectral density $S(\mathbf{k}) = (\tau^2 + \|\mathbf{k}\|^2)^{-(\nu+1)}$

2. KL dimensional reduction: Select the smallest index set $\mathcal{K}_{\text{KL}}$ of leading modes retaining energy fraction $\geq \rho$, and set $N_{\text{KL}} = |\mathcal{K}_{\text{KL}}|$.

3. Latent parameterisation: Construct the KL expansion

$$u^{\text{KL}}(\mathbf{x}) = m(\mathbf{x}) + \sum_{\mathbf{k} \in \mathcal{K}_{\text{KL}}} S(\mathbf{k})^{1/2} w_{\mathbf{k}} \phi_{\mathbf{k}}(\mathbf{x}), \quad w_{\mathbf{k}} \sim \mathcal{N}(0, 1).$$

4. Real-valued latent field: Construct $\tilde{u}^{\text{KL}} = \sqrt{2}(\Re(u^{\text{KL}}) + \Im(u^{\text{KL}}))$

5. Pushforward mapping: Compute $v(\mathbf{x}) = v_{\min} + (v_{\max} - v_{\min})\sigma_{\alpha}(\tilde{u}^{\text{KL}}(\mathbf{x}))$

6. Forward model: For each station pair (i, j) , compute predicted travel times

$$G(v)_{ij} = \int_{\Gamma_{ij}} v^{-1}(\mathbf{x}) dl$$

7. MAP estimation: Starting from $\mathbf{w}_0^{\text{MAP}}$, maximise the log-posterior

$$\Phi(\mathbf{w}^{\text{KL}}) = -\frac{1}{2\sigma_{\text{noise}}^2} \|\mathbf{y}^{\text{obs}} - G(\mathcal{T}(\mathbf{w}^{\text{KL}}))\|_2^2 - \frac{1}{2} \|\mathbf{w}^{\text{KL}}\|_2^2$$

using L-BFGS to obtain $\mathbf{w}_{\text{MAP}}^{\text{KL}}$ and $v_{\text{MAP}}(\mathbf{x}) = \mathcal{T}(\mathbf{w}_{\text{MAP}}^{\text{KL}})(\mathbf{x})$

8. Posterior sampling: Initialise NUTS from $\mathbf{w}_0^{\text{NUTS}}$ and draw samples $\{\mathbf{w}^{(s), \text{KL}}\}_{s=1}^{N_s}$ and compute $v^{(s)}(\mathbf{x}) = \mathcal{T}(\mathbf{w}^{(s), \text{KL}})(\mathbf{x})$

9. Posterior summaries: Compute posterior mean and variance from $\{v^{(s)}(\mathbf{x})\}_{s=1}^{N_s}$

5 NUMERICAL AND EXPERIMENTAL VALIDATION

This section evaluates the proposed inference-mode ANT framework through a sequence of increasingly realistic tests. We begin by introducing the experimental ambient-noise data set used in this study. We assess velocity-field reconstruction accuracy and posterior uncertainty through controlled synthetic experiments. We first consider a regular station geometry and then, inspired by the experimental setup, examine a realistic station geometry to study the effect of station distribution on reconstruction quality and UQ under controlled conditions. In Section 5.3, we apply the inference-mode ANT framework to the experimental data set and compare the results with those obtained from a discovery-mode inversion.

5.1 The Kylylahti data set and pre-processing

Active seismic surveys are conventionally used to image reflectors of lithological boundaries associated with mineralized rock formations, often in combination with other geophysical methods, but passive imaging can also constrain indicative velocity variations (P. Dales *et al.* 2020; T. Ryberg *et al.* 2022; D. Teodor *et al.* 2022). Here we use ambient-noise recordings acquired in the Kylylahti polymetallic mine area in the Outokumpu district, eastern Finland (M. Chamarczuk 2021), for our inference-mode surface wave imaging. The Kylylahti area hosts a volcanogenic massive sulphide deposit embedded within a heterogeneous sequence of black schist, greywacke, quartz rock, serpentinite, skarn and talc schist. The coloured lithological units in Fig. 4 are the geologically mapped surface expressions of the formation. The near-vertically extended blocks of the intrusions dip steeply to the south, and they have well-defined lateral boundaries to the host rock environment. This study site presents an opportunity to test inference-mode inversion because the mapped lithological units motivate a simplified structural hypothesis in which the laterally variable phase-velocity field is approximated by a small number of relatively homogeneous regions separated by sharp transitions. The lithological map is therefore used as external geological context for designing the structural prior model used in Sections 5.2 and 5.3, rather than as direct or complete prior knowledge of the subsurface velocity field.

Between August and September 2016, 994 vertical-component receivers were deployed over an area of approximately 3.5 km × 3 km, with 200 m line spacing and 50 m station spacing (Fig. 4). Each station consisted of a Geospace GSR recorder attached to six 10 Hz geophones, recording continuous ambient noise at a sampling interval of 2 ms for approximately one month. The continuous recordings were pre-processed considering the workflow of M. Chamarczuk *et al.* (2022). Active-source shots and other spurious transients were removed, the traces were detrended and bandpass filtered and interstation cross-correlations were computed to obtain empirical Green's functions (V. Hopyavuori *et al.* 2025). Rayleigh wave phase velocities were measured in the frequency domain following the approach of G. Ekström *et al.* (2009), based on the formulation of K. Aki (1957) for the cross-correlation of stochastic wavefields. Phase velocities were determined at the zero-crossing frequencies from the real part of the cross-correlation spectrum and then interpolated

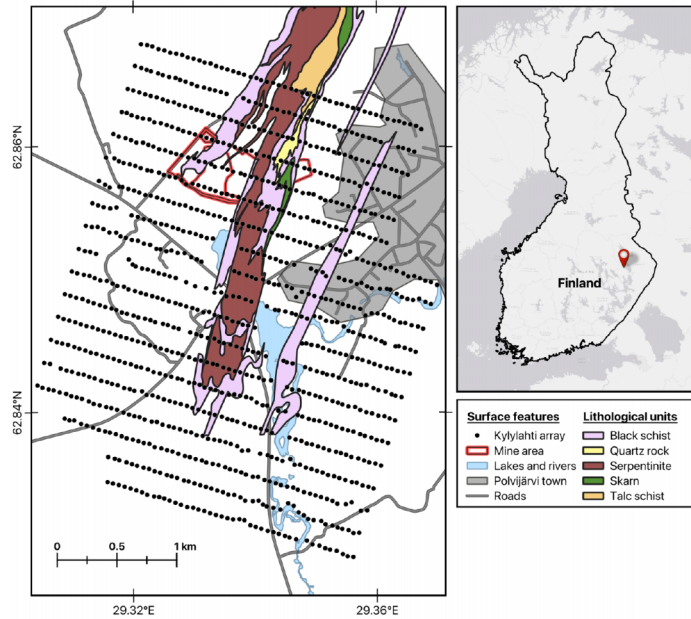


Figure 4. The Kylylahti dense array deployment in eastern Finland. The 994-station array covers an area of $\sim 10.5 \text{ km}^2$ and is organized in 19 rows with 200 m spacing. The inline interstation distance is 50 m.

to regularly spaced frequencies in the range of 1-10 Hz. Quality control was applied, including a signal-to-noise ratio threshold, a minimum interstation distance of one wavelength, continuity of phase velocity dispersion across frequencies and a physically reasonable velocity range. This yielded 37 399 measurements for input into ANT inversion.

5.2 Synthetic experiments

To assess the performance of the proposed framework under controlled conditions, we design synthetic experiments that approximate the Kylylahti setting while providing access to a known true velocity model. The aim of these experiments is (i) to quantify reconstruction accuracy and the recovery of prescribed structural features under different station geometries, and (ii) to examine how posterior uncertainty and MCMC performance respond to changes in data coverage.

5.2.1 Setup

True velocity model and domain. Fig. 5(a) shows the reference phase-velocity field for the synthetic tests. It is a piecewise-constant model motivated by the main lithological patterns shown in the mapped geology (Fig. 4), with velocity values bounded between $v_{\min} = 2 \text{ km s}^{-1}$ and $v_{\max} = 3 \text{ km s}^{-1}$. The model is discretized on a uniform grid with $N_{\text{points}} = 310$ points in each direction, resulting in a $310 \times 310 = 96\,100$ representation of the true velocity field. This piecewise-constant model is the synthetic ground truth against which reconstruction accuracy is quantitatively assessed.

Although the synthetic domain considered here is $1 \text{ km} \times 1 \text{ km}$, which is smaller than the spatial extent of the Kylylahti data set, this choice does not affect the validity of the seismic model investigated in this study. The forward model and the prior parametrization are scale-invariant with respect to the domain size, so the synthetic experiments serve primarily as a controlled benchmark for assessing reconstruction behaviour and UQ.

Station geometries. Synthetic traveltimes are generated for two acquisition scenarios. First, the regular geometry is an idealized, uniformly spaced station layout on a 14×14 grid. This configuration produces a nearly uniform ray coverage across the domain, with all interstation paths within the array used, including both station orders, resulting in a total of 38 220 synthetic traveltimes. Fig. 5(b) shows the corresponding ray-density map. Secondly, the real geometry of the actual Kylylahti station configuration is used. Owing to irregular station spacing, interstation paths and array boundaries, this setup yields a strongly heterogeneous ray coverage, with 37 399 interstation paths (Fig. 5c). By comparing these two acquisition scenarios, we explicitly evaluate the impact of data coverage on both velocity reconstruction quality and UQ within the proposed inference-mode ANT framework.

Synthetic data generation. For each acquisition geometry, synthetic traveltimes are generated by applying the forward operator (eq. 2) to the reference velocity model. The line integrals are evaluated using $N_{\text{quad}} = 450 > \sqrt{2} \times 310$ quadrature points per ray to ensure accurate numerical integration. Zero-mean Gaussian noise is then added to the noise-free traveltimes. Let $\mathbf{y}^{\text{true}} \in \mathbb{R}^m$ denote the vector

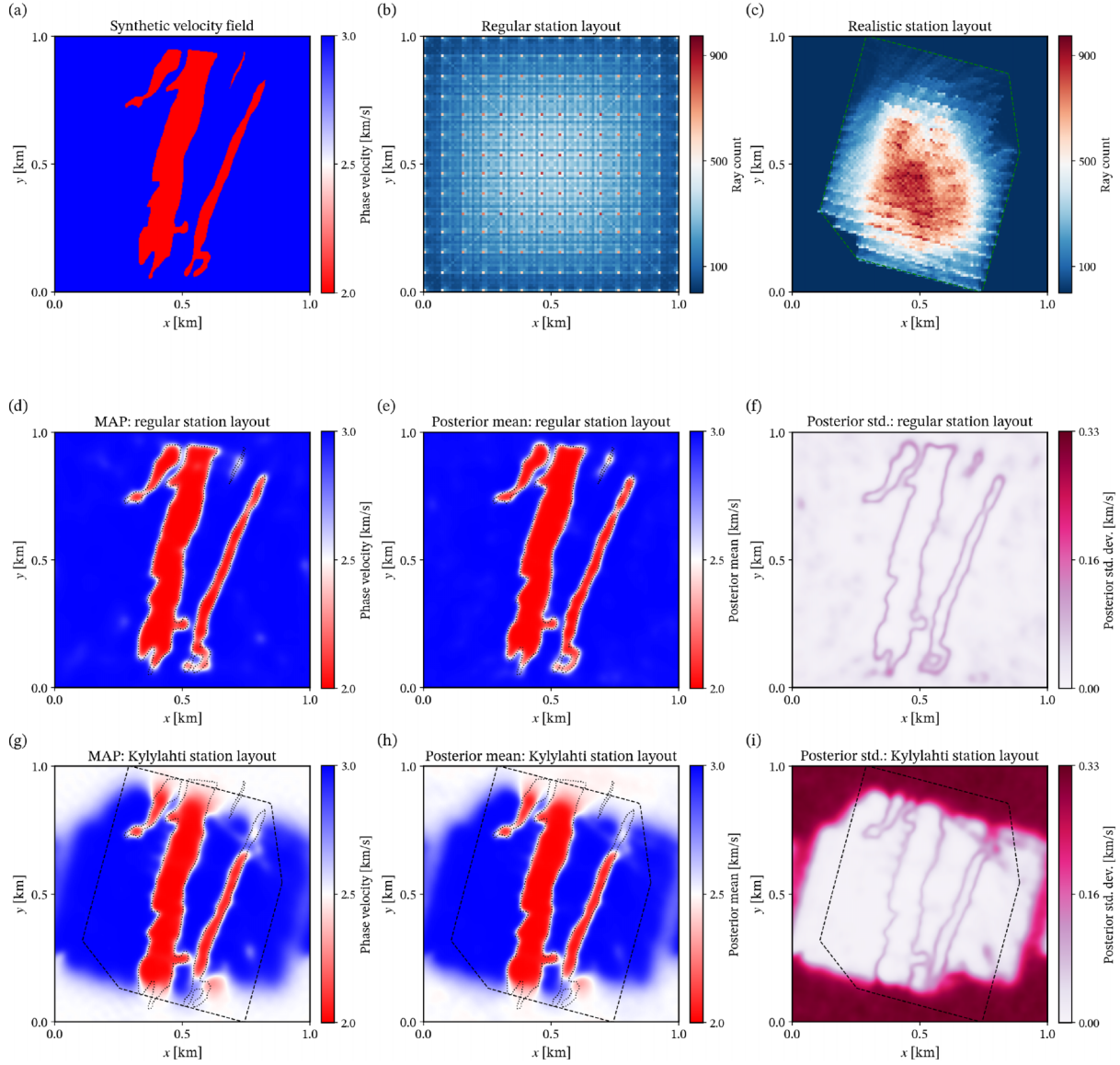


Figure 5. Synthetic experiment setup, MAP reconstructions and posterior uncertainty for two station layouts. (a) True synthetic phase-velocity field. (b, c) Ray-density maps for the regular and Kylylahti-inspired station layouts, respectively. (d)–(f) Results for the regular station layout: MAP reconstruction, posterior mean and posterior standard deviation. (g)–(i) Corresponding results for the Kylylahti-inspired station layout. The dotted contours overlaid on the reconstruction panels indicate the outlines of the true low-velocity structures. The regular layout provides more uniform ray coverage and sharper reconstructions, whereas the Kylylahti-inspired layout leads to spatially varying resolution and increased posterior uncertainty in regions of limited coverage. The black dashed line in (g)–(i) indicates the station coverage area.

of noise-free synthetic traveltimes, where m is the number of traveltimes measurements. A noise vector $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_{\text{noise}}^2 I_m)$ is drawn with

$$\sigma_{\text{noise}} = d_{\text{noise}} \frac{\|\mathbf{y}^{\text{true}}\|_2}{\sqrt{m}}, \quad (22)$$

where I_m is the $m \times m$ identity matrix and d_{noise} denotes the relative noise level (RNL). In this study, we choose $d_{\text{noise}} = 0.05$, corresponding in expectation to a root-mean-square (RMS) noise level equal to 5 per cent of the RMS signal amplitude. This choice is consistent with previous noise tomography studies that consider noise levels in the range of 3–8 per cent to account for data uncertainty (C. Esteve *et al.* 2025a, b). We additionally tested RNLs ranging from 1 to 15 per cent for synthetic experiments and found the proposed framework to be robust across this range. The noisy synthetic observations are then $\mathbf{y}^{\text{obs}} = \mathbf{y}^{\text{true}} + \boldsymbol{\eta}$.

Prior hyperparameters. The prior hyperparameters are chosen in accordance with the guidelines discussed in Section 3.3. The WM prior used in the synthetic experiments is parametrized as

$$\alpha = 5 \times 10^4, \quad \tau = 15, \quad \nu = 2, \quad v_{\min} = 2 \text{ km s}^{-1}, \quad v_{\max} = 3 \text{ km s}^{-1}.$$

The pushforward parameter α is chosen relative to the amplitude scale of the latent Gaussian field on this grid. For the range of WM hyperparameters considered in the diagnostic sweeps, the RMS amplitude of $\tilde{u}^{\text{KL}}$ remains of order 10^{-5} , with typical values between 2×10^{-5} and 7×10^{-5} . Hence, the choice $\alpha = 5 \times 10^4$ gives $\alpha \text{RMS}(\tilde{u}^{\text{KL}})$ of order unity. This produces a sufficiently sharp sigmoid pushforward while keeping the map differentiable over the typical amplitude range of the latent field. Larger values of α make the map closer to a Heaviside function, but also localize the gradients more strongly near the zero level set and were found to reduce numerical robustness.

The value $\tau = 15$ is chosen to be consistent with the characteristic lateral scale of the structures in the example, while also satisfying the discretization requirement. Since the synthetic experiments are performed on the unit square, this corresponds to a prior correlation length $\ell = 1/(\pi\tau) \approx 0.0212$. For the 310×310 computational grid, the diagonal cell size is $h_{\text{max}} \approx 4.56 \times 10^{-3}$, so that $\ell/h_{\text{max}} \approx 14.6\pi^{-1}$. Thus, the prior correlation length is well resolved by the discretization and is comparable to the widths and separations of the main inclusions in the example. This makes $\tau = 15$ a structurally motivated modelling choice rather than a purely numerical tuning parameter. As an additional consistency check, Supplementary Material S2.1 shows diagnostic sweeps over τ , monitoring the trade-off between data misfit and the prior norm. These diagnostics support the selected value of τ .

The smoothness parameter $\nu = 2$ is chosen to allow sufficiently regular latent fields while still permitting spatial variability in the interface geometry. With these hyperparameters, the latent Gaussian field is represented using a truncated KL expansion retaining $N_{\text{KL}} = 1591$ modes, corresponding to a spectral energy fraction $\rho = 0.99$. This reduces the number of unknowns from 96 100 grid values to 1591 latent coefficients, corresponding to more than 98 per cent dimensional reduction, while preserving the dominant variability implied by the prior.

The velocity bounds v_{min} and v_{max} are fixed to represent a controlled contrast between low- and high-velocity regions in the synthetic experiment, so that the numerical study focuses on reconstructing the interface geometry and quantifying its uncertainty.

Quality metrics. For the synthetic experiments, where the reference velocity field is known, image-domain reconstruction accuracy is quantified using two complementary metrics. We compute the percentage of the relative error (RE),

$$\text{RE}(\text{per cent}) = 100 \times \frac{\|v_{\text{rec}} - v_{\text{true}}\|_2}{\|v_{\text{true}}\|_2},$$

which measures the global amplitude mismatch between the reconstructed and reference velocity fields. We also compute the structural similarity index (SSIM),

$$\text{SSIM}(v_{\text{rec}}, v_{\text{true}}) = \frac{(2\mu_{\text{rec}}\mu_{\text{true}} + C_1)(2\sigma_{\text{rec},\text{true}} + C_2)}{(\mu_{\text{rec}}^2 + \mu_{\text{true}}^2 + C_1)(\sigma_{\text{rec}}^2 + \sigma_{\text{true}}^2 + C_2)},$$

where μ_{rec} and μ_{true} denote the mean values of the reconstructed and reference velocity fields, σ_{rec}^2 and σ_{true}^2 their variances and $\sigma_{\text{rec},\text{true}}$ their covariance. The constants C_1 and C_2 are small positive stabilization parameters. The SSIM measures structural similarity by jointly comparing luminance, contrast and structural information (Z. Wang *et al.* 2004). Its values typically lie in $[-1, 1]$, with $\text{SSIM} = 1$ indicating perfect structural agreement between the compared fields.

For both synthetic and real-data experiments, we quantify the traveltime data fit using the noise-normalized residual RMS (NRMS_{res}),

$$\text{NRMS}_{\text{res}} := \frac{\|y^{\text{obs}} - G(v_{\text{rec}})\|_2}{\sqrt{m} \sigma_{\text{noise}}}.$$

Values near one indicate that the residual RMS is comparable to the assumed noise level, whereas values substantially larger than one indicate a poorer fit relative to that noise level. For synthetic data, this compares the residual with the imposed measurement noise; for real data, it compares the residual with the assumed noise level used in the likelihood.

Sampling diagnostics. For experiments involving Bayesian UQ, posterior samples are generated using NUTS in the reduced KL space. Sampling performance is assessed using standard diagnostics, including trace plots of representative KL coefficients, the effective sample size (ESS), and the Gelman–Rubin statistic $\hat{R}$ (A.B. Owen 2013). High ESS values and $\hat{R} \approx 1$ across all retained modes indicate good mixing and cross-chain convergence, while trace plots verify stable and stationary exploration of the posterior. Together, these diagnostics confirm that the posterior uncertainty estimates reported in Sections 5.2.3 and 5.3.3 are statistically reliable.

5.2.2 MAP reconstructions and effect of station geometry

We first examine how data coverage influences the reconstruction quality of the MAP estimate. Figs 5(d) and (g) compare MAP velocity fields obtained using a regular, uniformly spaced station layout with those obtained using the Kylylahti deployment geometry, respectively. For the MAP estimation, the L-BFGS optimization is initialized from a random draw in the reduced KL coefficient space, $w_0^{\text{MAP}} \sim \mathcal{N}(\mathbf{0}, I_{N_{\text{KL}}})$.

With uniformly distributed stations, ray coverage is dense and approximately uniform across the domain (Fig. 5b). The resulting MAP estimate, shown in Fig. 5(d), accurately recovers the major high- and low-velocity regions of the reference model, with sharp interfaces and correct amplitude contrasts. Quantitatively, the reconstruction achieves $\text{RE} = 4.10$ per cent and $\text{SSIM} = 0.83$ in the velocity field. The traveltime data fit is $\text{NRMS}_{\text{res}} = 1.003$, indicating that the residual RMS is consistent with the imposed synthetic

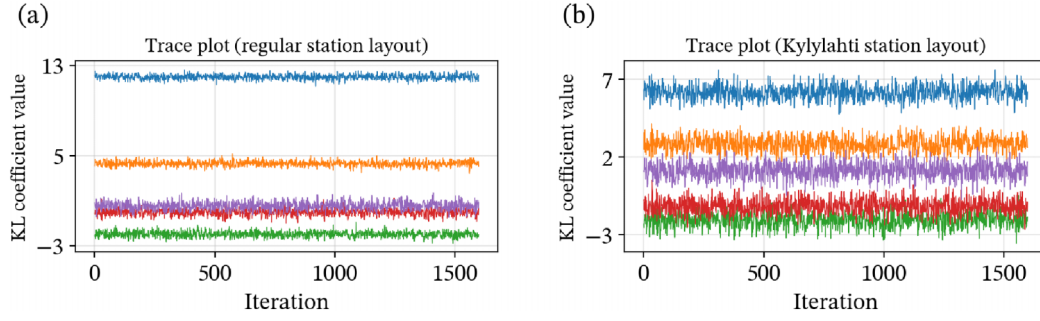


Figure 6. Trace plots of the first five KL coefficients for the synthetic experiments with (a) a regular station layout and (b) a Kylahti-inspired station layout. Each line shows one KL coefficient across all four MCMC chains, concatenated for visualization (400 samples per chain; 1600 samples in total). The chains exhibit stationary behaviour without visible drift, consistent with the high ESS and $\hat{R} \approx 1$.

noise level. These values indicate good global reconstruction accuracy and a data fit consistent with the assumed noise model. Overall, the regular geometry provides a controlled and spatially uniform benchmark, illustrating the performance achievable under relatively homogeneous and well-distributed data coverage.

When the irregular and anisotropic field station distribution is used, ray coverage becomes uneven, with sparse sampling in several parts of the domain (Fig. 5c). The corresponding MAP estimate, shown in Fig. 5(g), still reproduces the large-scale structure of the true model, but interfaces appear locally smeared and smaller-scale features are partially lost. This degradation in resolution is reflected in the metrics, with RE = 10.44 per cent and SSIM = 0.67, indicating reduced structural similarity despite the overall geometry of the anomalies remaining consistent with the ground truth. At the same time, the traveltime data fit remains consistent with the imposed noise level, with $\text{NRMS}_{\text{res}} = 0.995$. This indicates that the degraded image quality is mainly associated with the reduced resolving power of the irregular acquisition geometry rather than a poor fit to the traveltime data.

This comparison highlights the role of station geometry in controlling the achievable resolution in ANT, even within an inference-mode framework. Regular station spacing facilitates more uniform and sharper reconstructions, while realistic acquisition geometries naturally lead to anisotropic resolution patterns and spatially varying uncertainty. These effects reflect intrinsic properties of the data coverage rather than limitations of the inversion methodology, and they become even more evident when examining posterior uncertainty in the following subsection.

5.2.3 Uncertainty quantification

We now analyse the posterior distributions obtained using the NUTS in the reduced KL space for both station geometries. Four independent NUTS chains are initialized from the zero vector in the reduced KL coefficient space, $\mathbf{w}_0^{\text{NUTS}} = \mathbf{0}$, and run with 250 warm-up iterations followed by 400 sampling iterations, yielding a total of 1 600 posterior samples. Figs 5(e), (f), (h) and (i) show the posterior mean and standard deviation fields for both acquisition geometries, which together characterize the inferred velocity structures and their associated uncertainty.

Figs 5(e) and (f) show the posterior mean and standard deviation fields, respectively, for the case of dense and approximately uniform ray coverage (Fig. 5b). The posterior mean exhibits a high similarity with the reference distribution. Posterior standard deviation values are almost uniformly low throughout the interior of the domain and are large mostly along velocity interfaces. This spatial pattern indicates that the data constrain the model well across most of the domain, with elevated uncertainty limited to regions associated with sharp contrasts. A small localized region of elevated uncertainty near the upper-right corner suggests limited sensitivity of the data in this area, which can potentially be reduced through increased data coverage or alternative prior hyperparameter choices.

Quantitatively, the reconstruction achieves RE = 3.98 per cent and SSIM = 0.84. The traveltime data fit is $\text{NRMS}_{\text{res}} = 1.005$, indicating that the residual RMS is consistent with the imposed synthetic noise level. The sampler shows good sampling efficiency and stable convergence behaviour. The minimum ESS across the 1 591 retained KL coefficients is 582, and the maximum Gelman–Rubin diagnostic is $\hat{R}_{\text{max}} = 1.012$, consistent with effective mixing and cross-chain convergence. The trace plots shown in Fig. 6(a) exhibit stable and stationary behaviour with no visible drift or stickiness, consistent with the high ESS values. The total runtime for all four chains executed sequentially is approximately 106 min on a MacBook Pro with an Apple M-series GPU and 24 GB RAM.

Figs 5(h) and (i) show the posterior mean and standard deviation fields, respectively, for the Kylahti-inspired station layout (Fig. 5c). The posterior mean continues to recover the dominant high- and low-velocity regions of the true model, but with reduced sharpness and more pronounced smoothing in areas with sparse ray coverage. The posterior standard deviation reveals substantially increased uncertainty, as before, in localized regions with strong velocity gradients in the well-imaged central parts of the domain, and then where ray coverage is sparse or strongly anisotropic. These regions correspond closely to areas where the MAP estimate exhibits blurred or distorted interfaces, confirming that the posterior variance appropriately captures the loss of information associated with limited data coverage. Values outside the black dashed line should be interpreted with caution, as these regions lie outside the effective data coverage and are not constrained by observations.

Quantitatively, the reconstruction achieves $RE = 10.53$ per cent and $SSIM = 0.68$. The corresponding traveltime fit is $NRMS_{res} = 0.995$, indicating that the residual RMS remains consistent with the imposed synthetic noise level. The increase in image-domain error therefore reflects the limited spatial resolving power of the irregular station geometry, rather than a poor fit to the traveltime data. Sampling performance remains robust despite the more complex posterior geometry. Across the 1591 retained KL coefficients, the minimum ESS is 752 and the maximum Gelman–Rubin statistic is $\hat{R}_{max} = 1.007$, consistent with reliable cross-chain convergence and adequate posterior exploration. The trace plots for the real-geometry experiment (Fig. 6b) again exhibit stable, stationary behaviour without visible drift or stickiness, consistent with the ESS and $\hat{R}$ diagnostics. The total runtime for all four chains is 190 min, showing that NUTS remains practical under uneven and sparse ray coverage.

The posterior analysis demonstrates that the proposed inference-mode ANT framework produces reliable uncertainty estimates that reflect the data coverage and inversion geometry. Regular station spacing yields sharply resolved anomalies with low uncertainty, whereas the field station configuration produces a posterior variance pattern that clearly identifies regions where the inversion is weakly constrained. These results emphasize the importance of UQ. Even when two MAP models appear visually similar, the posterior variance quantifies the degree of confidence in different parts of the reconstructed structure.

5.3 Array-data inversion and comparison with a discovery-mode example

In this section, we evaluate the performance of the proposed inference-mode ANT framework on data from the Kylylahti passive seismic experiment and compare the results with a conventional discovery-mode inversion.

5.3.1 Setup

We apply the inference-mode ANT framework to the 6 Hz Rayleigh-wave phase-velocity measurements introduced in Section 5.1. Before inversion, the station coordinates are transformed from geographic coordinates to Universal Transverse Mercator coordinates (in kilometres) and subsequently normalized to the unit square. The inversion and posterior sampling are performed in this normalized computational domain, while the final reconstructed velocity fields are mapped back to physical coordinates for visualization and interpretation.

The overall inversion setup follows that of the synthetic experiments, with a small number of modifications. In particular, the inverse correlation length of the WM prior is reduced to $\tau_{norm} = 8$ in the normalized computational domain, while the smoothness parameter is kept fixed at $\nu = 2$.

This choice reflects a modelling expectation of smoother large-scale velocity variations in the acquired data compared to the synthetic models. Since the station coordinates are normalized using the physical length scale $N_F = 4.23$ km, the normalized value $\tau_{norm} = 8$ corresponds to a physical inverse correlation length $\tau_{phys} = \pi \tau_{norm} / N_F \approx 1.89\pi$ km⁻¹, or equivalently to a physical correlation length $\ell_{phys} \approx 0.53\pi$ km. For the 310×310 computational grid, the physical diagonal cell size is $h_{max} \approx 0.0193$ km, which yields $\ell_{phys} / h_{max} \approx 27.4\pi$. Thus, the physical correlation length associated with τ_{norm} is well resolved by the discretization. As an additional consistency check, Supplementary Material S2.2 reports diagnostic sweeps over τ_{norm} , monitoring the trade-off between data misfit and the prior norm. These diagnostics support the selected value of τ_{norm} .

The pushforward slope parameter is kept fixed at $\alpha = 5 \times 10^4$, as in the synthetic experiments. This value was chosen to balance interface sharpness and numerical stability. With these hyperparameters, the latent Gaussian field is represented using a truncated KL expansion retaining $N_{KL} = 453$ modes, which still corresponds to a spectral energy fraction $\rho = 0.99$. This represents a dimensionality reduction of more than 99.5 per cent while preserving the dominant prior-implied variability. The relative noise level is set to $d_{noise} = 5$ per cent, consistent with the data uncertainty levels assumed in the synthetic experiments and previous noise tomography studies (C. Esteve *et al.* 2025a, b).

The pushforward mapping is defined in the normalized velocity representation, while the corresponding physical velocity bounds after inverse normalization are

$$v_{min} = 2.55 \text{ km s}^{-1}, \quad v_{max} = 2.80 \text{ km s}^{-1}.$$

These values should be interpreted as effective phase-velocity levels at the considered period, rather than as exact intrinsic velocities of two homogeneous lithologies. They are consistent with the velocity range suggested by the discovery-mode reconstruction (see Section 5.3.4) and were further supported by a grid-search calibration over admissible values of v_{min} and v_{max} (see Supplementary Material S2.4).

5.3.2 MAP reconstruction

Fig. 7(a) shows the MAP phase-velocity field. The reconstruction reveals a set of coherent high- and low-velocity domains with contrasts of 0.25 km s^{-1} , forming elongated bands and isolated patches rather than pixel-scale artefacts. These patterns are spatially consistent with the dominant structural trends inferred from the surface geology of the Kylylahti area. The MAP estimate yields $NRMS_{res} = 0.714$ with respect to the measured traveltime data, indicating that the residual RMS is below the noise level assumed in the likelihood model.

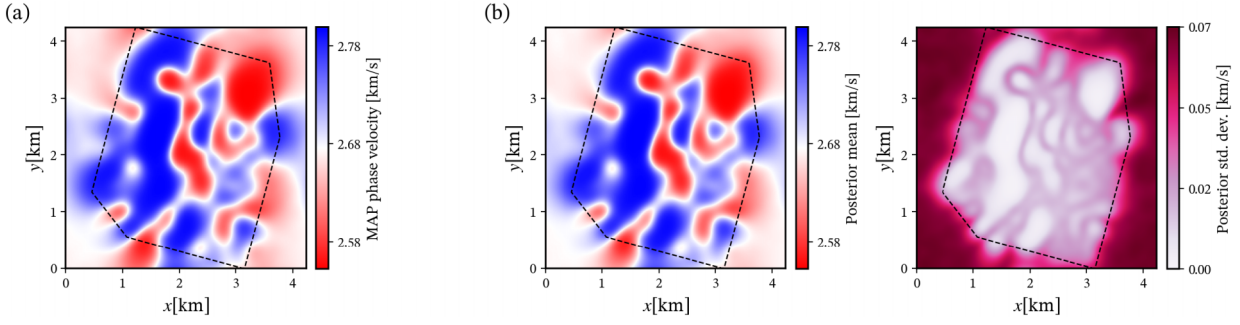


Figure 7. Inference-mode inversion of 6 Hz Rayleigh-wave phase-velocity data from the Kylylahti experiment. (a) MAP phase-velocity model showing elongated high- and low-velocity domains. (b) Posterior mean and standard deviation fields obtained from NUTS, illustrating regions that are well constrained by the data (low variance) and areas of elevated uncertainty, primarily along anomaly boundaries. The black dashed line indicates the station coverage area.

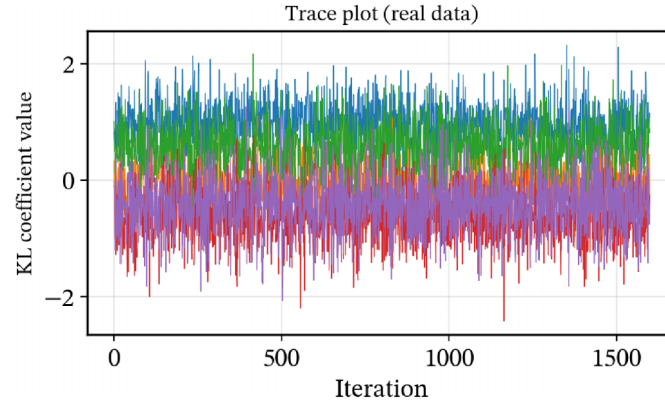


Figure 8. Trace plot of the first five KL coefficients for the real-data inversion. Each line shows one KL coefficient across all four MCMC chains, concatenated for visualization (400 samples per chain; 1600 samples in total). The chains exhibit stationary behaviour without visible drift, consistent with the high ESS and $\hat{R} \approx 1$.

5.3.3 Uncertainty quantification

Fig. 7(b) shows the posterior mean and posterior standard deviation fields. The posterior mean is consistent with the structure observed in the MAP estimate, indicating that the major velocity variations are strongly supported by the data. Moreover, the posterior mean gives a comparable traveltime data fit to the MAP estimate, with $\text{NRMS}_{\text{res}} = 0.714$ relative to the noise level assumed in the likelihood model. The posterior standard deviation highlights regions where the inversion is less well constrained, primarily along the inferred unit boundaries and in areas with sparse or uneven ray coverage. These uncertainty patterns indicate that the location and overall extent of the main reconstructed velocity domains are comparatively robust within the assumed model, while their fine-scale geometry has a larger degree of uncertainty.

The sampler achieves high sampling efficiency with rapid convergence. Across all retained KL coefficients, the maximum Gelman–Rubin diagnostic is $\hat{R}_{\text{max}} = 1.002$, and the minimum ESS is $\text{ESS}_{\text{min}} = 1565$, indicating efficient mixing and near-independent posterior samples. The trace plots for the first five KL coefficients (Fig. 8) exhibit stable and stationary behaviour without evidence of drift, consistent with the high ESS values and $\hat{R} \approx 1$ convergence diagnostics. The total runtime for all four chains is 104 min.

5.3.4 Comparison with discovery-mode inversion

To assess the performance of the proposed inference-mode framework, we compare the reconstructed 6 Hz phase-velocity image with a discovery-mode inversion result shown in Fig. 9. Both reconstructions are based on the Kylylahti traveltime data set, but they differ fundamentally in their formulation of the inverse problem. For this comparison, we employ a conventional deterministic ANT approach, as a representative discovery-mode inversion. The velocity domain is discretized similarly to the inference-mode inversion described in Section 5.2.1, and we use a forward simulator under the same assumption of straight rays. The misfit function is formulated following Y. Lu *et al.* (2018) and F. Korostelev *et al.* (2022), incorporating Tikhonov-type regularization

$$F(\mathbf{v}) = \|\mathbf{y}^{\text{obs}} - G(\mathbf{v})\|^2 + \lambda \|\mathbf{L}\mathbf{v}\|^2, \quad (23)$$

where $G(\mathbf{v})$ denotes the forward model defined in eq. (2). Here, we assume uniform data covariance across all measurements, that is, each measurement is weighted equally in the misfit function. The second term represents a flattening regularization term, where

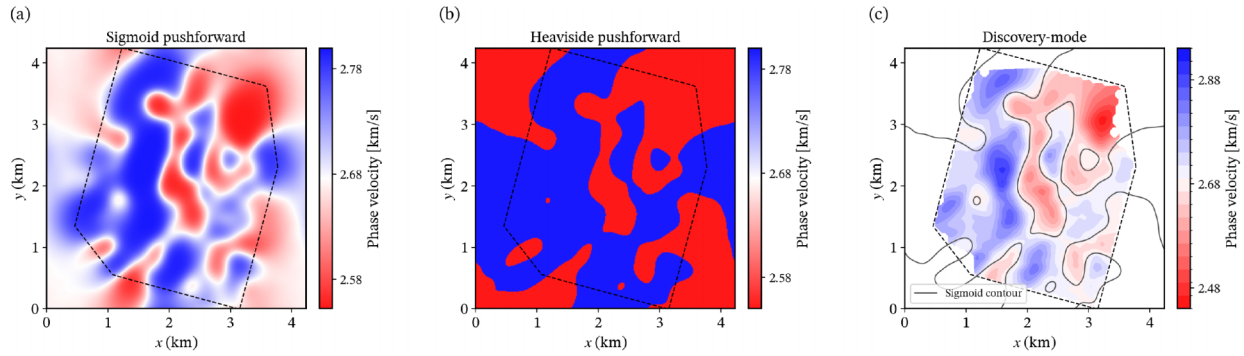


Figure 9. Comparison between inference-mode and discovery-mode reconstructions of the 6 Hz Rayleigh-wave phase-velocity data from the Kylylahti experiment. (a) MAP estimate obtained using the proposed inference-mode reconstruction with the sigmoid pushforward. (b) Heaviside transformation of the reconstruction shown in panel (a), included to visualize the corresponding segmented representation. (c) Discovery-mode inversion. The black dashed line indicates the station coverage area. The grey contour in panel (c) represents the boundary inferred from the inference-mode reconstruction in panel (a), overlaid to facilitate comparison between the large-scale structures recovered by the two approaches.

L is the first-order differential operator of the model parameter v with respect to spatial location, and λ is the weight assigned to the regularization constraint. This penalty suppresses large spatial gradients and therefore favours spatially gradual variations in the reconstructed velocity field. The inversion starts from a homogeneous initial velocity model, set to the median value of all phase velocity measurements. The solution is approximated using the iterative least-squares QR (LSQR) algorithm (C.C. Paige & M.A. Saunders 1982). The regularization coefficient is set to $\lambda = 2$, close to the point of maximum curvature on the L-curve, balancing data fit and model regularity (P. Hansen 2001); see Supplementary Material S2.3. The resulting phase velocity map is shown in Fig. 9(c), omitting areas without sufficient data coverage. For visual comparison, the same figure also shows the sigmoid-pushforward inference-mode MAP reconstruction in panel (a) and its Heaviside-pushforward (eq. 13) representation in panel (b). Panel (b) is included to separate the inferred smooth sigmoid transition from the corresponding segmented interpretation and to make explicit the connection with thresholded or post-processed discovery-mode images. The boundary inferred from the inference-mode reconstruction is overlaid as a grey contour on the discovery-mode result in panel (c).

The discovery-mode inversion, while also reproducing the expected geological patterns, yields a single smooth deterministic model for which uncertainty estimates are not computed in the present comparison. We acknowledge that uncertainty estimates can in principle be derived for regularized least-squares inversions, for example, through linearized error propagation from measurement uncertainties or through an equivalent Bayesian Gaussian formulation (A. Tarantola & B. Valette 1982). However, the interpretation of such uncertainty estimates depends strongly on the assumed regularization or prior covariance structure.

Sharper segmented representations can also be obtained from smooth discovery-mode models by applying additional post-processing, for example, through thresholding or level-set transformations, or by integrating seismic velocities with complementary geophysical data, such as resistivity, using unsupervised classification tools (T. Ryberg *et al.* 2022). Indeed, Fig. 9 shows that the sigmoid-pushforward MAP reconstruction and the discovery-mode inversion share broadly similar large-scale structures.

The difference between the two approaches is therefore not simply the visual sharpness of the final image, but how the structural assumptions are introduced. In the proposed approach, the expected piecewise nature of the velocity field is incorporated directly into the WM prior and pushforward mapping. The resulting MAP reconstruction should therefore be interpreted as a reconstruction conditional on the imposed geology-informed structural assumptions, rather than as an objectively more interpretable image in all settings. The proposed framework provides posterior uncertainty information within the same Bayesian formulation used to obtain the reconstruction. These variance fields indicate which reconstructed structural features are well constrained by the data and where ambiguity remains, for example, near sharp velocity contrasts and in regions of sparse ray coverage. We acknowledge that these uncertainty estimates are also conditional on the assumed forward model, noise model, and prior structure.

Although not assessed in this experiment, our framework can reduce the computational burden relative to probabilistic discovery-mode ANT formulations that sample all grid values directly. Rather than inverting for all grid values, the velocity field is represented in a reduced KL space with 453 coefficients, retaining more than 99 per cent of the prior variance. This compact representation stabilizes the inversion and enables efficient UQ.

6 DISCUSSION

Our results illustrate both the reconstruction performance and the associated uncertainty patterns of the proposed framework. We discuss their implications from an inference-mode perspective.

6.1 Inference-mode versus discovery-mode interpretation

The proposed framework follows the inference-mode perspective articulated by V.C. Tsai (2023), specifying the assumed structural hypothesis and quantifying the uncertainty associated with it. The aim is not to replace exploratory tomography, which remains essential for many problems, but to provide a complementary formulation for cases where the target structures and relevant geological questions are specified in advance. The framework combines three key ingredients: the structural prior, reduced-order parametrization and posterior UQ. The WM prior and pushforward mapping used here encode explicit modelling assumptions about the structural class considered in this study, namely internally relatively homogeneous regions separated by sharp lateral transitions. Dimensional reduction via the KL expansion restricts the inversion to spatial modes considered relevant under the chosen prior. Posterior UQ is central to interpreting the inversion results. In particular, posterior variance identifies which structures are relatively well-constrained under the assumed forward model, noise model and prior, and which remain ambiguous, even when MAP reconstructions appear visually similar. This shifts the emphasis from visual inspection of a single reconstruction towards answering pre-defined inferential questions under an explicit probabilistic model.

The proposed framework imposes stronger assumptions about the structure of the model than conventional discovery-mode tomography, so these additional assumptions should be justified by practical benefit. In the present setting, this added structure has four main benefits. First, the structural prior model makes the imposed modelling assumptions more explicit and interpretable. Secondly, the KL reduction lowers the effective stochastic dimension of the inverse problem, making optimization and posterior sampling more tractable. Thirdly, the Bayesian formulation provides uncertainty estimates in addition to a point reconstruction. Fourthly, these uncertainty estimates help distinguish robust reconstructed features from features that should be interpreted with caution. Although the prior construction is more elaborate than a single smoothing penalty, the workflow remains modular: The prior construction, forward model and posterior inference can be treated as separate components. This modularity is important beyond the ANT example considered here. The prior construction can be redesigned for other target structures or imaging modalities, while the Bayesian inference workflow remains largely unchanged once the forward model and likelihood have been specified. In this respect, ANT serves as a concrete example of a broader inference-mode strategy.

6.2 Role of model reduction and parametrization

Model reduction via the KL expansion is central to the inference-mode ANT formulation. Representing the velocity field in a low-dimensional latent space reduces computational cost while preserving the dominant variability implied by the chosen prior model. This reduction makes gradient-based MAP optimization and NUTS sampling feasible for the problem sizes considered here. It also makes the imposed structural assumptions easier to interpret, because the inversion is restricted to coherent spatial modes rather than pixel-scale fluctuations. Dimensional reduction is therefore not merely a numerical convenience, but a modelling choice aligned with the inference goals. By suppressing poorly informed high-frequency degrees of freedom, the framework can limit the risk of fitting data noise and focus UQ on structures represented by the chosen prior model. The resulting posterior summaries should therefore be interpreted within the assumed structural prior model and can support geological interpretation when the target hypothesis is appropriate for the site.

This reduction is also consistent with the spatial-resolution limits imposed by the computational discretization. As discussed in Section 3.3, the WM hyperparameters are selected so that the prior places negligible spectral variance above the mesh-resolvable frequency threshold. This helps prevent unresolved high-frequency components from influencing the inversion.

6.3 Limitations and future work

This study should be interpreted as a proof-of-concept demonstration of an inference-mode prior construction for ANT, rather than as a field-ready formulation for general geophysical imaging problems. The numerical experiments were designed to isolate the effect of the WM-pushforward prior, KL dimensional reduction and posterior sampling under controlled modelling assumptions. The most important limitations for practical ANT applications are the straight-ray forward model, the fixed effective velocity levels, the fixed prior hyperparameters and the restriction to a 2-D single-frequency phase-velocity map.

The most immediate limitation is the use of the straight-ray traveltime approximation in eq. (2). This approximation makes the forward operator inexpensive and allows us to focus on the prior and sampling aspects of the method, but it does not account for ray bending caused by lateral velocity contrasts. This is not only a technical refinement: For the synthetic example, the imposed velocities of 2 and 3 km s⁻¹ correspond to a strong contrast for which ray paths will deviate significantly from straight lines. Therefore, the synthetic results should not be interpreted as evidence that ray bending is negligible. Rather, they demonstrate that, conditional on a straight-ray forward model, the proposed prior can recover the prescribed structural class and that the posterior uncertainty reflects the imposed acquisition geometry.

For practical ANT applications, the first extension of the forward model should therefore be to replace the straight-ray line-integral model with an eikonal-equation-based traveltime solver or a nonlinear ray tracing formulation (N. Rawlinson & M. Sambridge 2003), so that ray paths are updated consistently with the evolving velocity field. This retains the reduced KL-pushforward parametrization,

but the computational bottleneck shifts towards repeatedly solving the nonlinear forward and adjoint problems. Efficient MAP estimation and gradient-based posterior sampling then require adjoint-state derivatives of the traveltime misfit with respect to the latent KL coefficients. Similar adjoint-based Hamiltonian Monte Carlo strategies have been demonstrated for nonlinear traveltime tomography (A. Fichtner *et al.* 2019; A. Zunino *et al.* 2023). Finite-frequency kernels or waveform-based likelihoods would be further extensions, but ray bending is the most immediate source of forward-modelling bias for the examples considered here.

A second important limitation is the fixed bi-level velocity representation. The parameters $v_{\min}$ and $v_{\max}$ do not merely define plotting limits; they constrain the class of admissible velocity fields and therefore have a direct influence on both the MAP reconstruction and the posterior uncertainty. In the Kylylahti example, these values are interpreted as effective phase-velocity levels at 6 Hz, not as intrinsic velocities of two perfectly homogeneous lithologies. They can absorb unresolved effects of vertical layering, finite-frequency sensitivity and subgrid lithological variability. This interpretation is plausible for the present site, where the surface geology contains steeply dipping units with clear lateral trends, but it is not generally valid for ANT in more complex settings.

The pushforward construction itself is not intrinsically restricted to two levels. Supplementary Material S3 provides an illustrative three-level synthetic example using the same principle. We nevertheless keep the main experiments bi-level to focus on the primary prior construction, dimensional reduction and posterior sampling framework. Beyond the number of levels, a related limitation is that their velocity values are fixed. A more complete formulation would treat these levels as unknown parameters and infer them jointly with the latent structural field. For practical applications, joint or hierarchical inference of the velocity levels would be especially important when the velocity contrast is not well constrained by independent geological or geophysical information.

The remaining hyperparameters also influence the posterior and should not be interpreted as purely numerical tuning constants. The inverse correlation length τ , smoothness parameter ν , pushforward slope α and assumed noise level σ_{noise} control, respectively, the characteristic lateral scale, interface regularity, transition width and relative weighting of the data and prior terms. In this study, these quantities are fixed so that the inference problem remains focused on the latent structural field and the synthetic and real-data examples remain computationally tractable. The reported posterior uncertainties should therefore be interpreted as conditional uncertainties under the chosen forward model, noise model, prior structure and hyperparameters. A natural extension is to infer selected hyperparameters hierarchically or to propagate their effect through model comparison and sensitivity analysis.

Finally, the present implementation is restricted to a 2-D lateral phase-velocity map at a single frequency. The 6 Hz Rayleigh-wave phase-velocity field should therefore be interpreted as an effective lateral model over the corresponding sensitivity range, not as a complete depth-dependent description of the subsurface. Extending the framework to multifrequency ANT is important for practical subsurface interpretation, either by applying the reconstruction frequency by frequency or by coupling several frequencies through a shared structural prior and frequency-dependent velocity levels. A subsequent depth inversion, or a full 3-D tomographic formulation, will then be required to relate the inferred phase-velocity structures to shear-wave velocity variations at depth.

Such extensions raise computational and scalability issues. The WM prior construction, KL reduction and pushforward mapping can, in principle, be formulated for 3-D or multiparameter fields. However, the number of spatial degrees of freedom, retained KL modes and forward-model evaluations increase substantially. A direct full-grid FFT implementation may become costly in both computation and memory. Alternative implementations could discretize the corresponding SPDE operator, compute the retained eigenpairs once in an offline stage and store the reduced KL basis for repeated use during MAP optimization or posterior sampling. Multiparameter extensions, involving, for example, several velocity fields, attenuation or other material properties, require additional choices such as independent structural priors, coupled latent fields or shared-interface representations. These extensions define natural directions towards practical inference-mode tomography.

7 CONCLUSIONS

We have introduced an inference-mode framework for ANT that integrates geology-informed structural prior modelling, reduced-order parametrization and Bayesian UQ to support targeted geological interpretation. Rather than producing a single deterministic image, the approach infers velocity structure together with spatially resolved measures of confidence.

Synthetic experiments demonstrate that the framework can recover the main prescribed velocity contrasts under favourable data coverage, and that posterior uncertainty is consistent with the acquisition geometry and data-coverage limitations. Application to the Kylylahti dense array data set shows that the inference-mode MAP reconstruction is consistent with the imposed geology-informed structural assumptions, while posterior variance fields indicate regions that are relatively well constrained under the assumed model and regions of remaining ambiguity. The Kylylahti polymetallic mine area in the Outokumpu mineral district, Finland, contains steeply dipping intrusive blocks with well-defined lateral boundaries, making it a suitable test case for inference-mode inversion using a piecewise-constant prior with two velocity levels. These levels should be interpreted as effective phase-velocity bounds at the analysed period, rather than as exact intrinsic velocities of two homogeneous lithologies. In this setting, the seismic phase-velocity field can be approximated through a simplified structural hypothesis motivated by the mapped lithological context.

Dimensional reduction via the KL representation is a key modelling choice that enables efficient MAP optimization and MCMC sampling while restricting inference to the dominant spatial modes implied by the chosen prior model. The current formulation is

limited to 2-D phase-velocity maps at a single frequency; extending the framework to multifrequency and 3-D ANT, together with hierarchical hyperparameter inference and more realistic forward models, is an important direction for future work. Overall, the proposed inference-mode ANT framework provides a potential complement to discovery-mode inversion in settings where specific structural hypotheses are available and UQ is required for quantitative subsurface interpretation.

ACKNOWLEDGMENTS

This work was supported by the Research Council of Finland (RCF; Flagship of Advanced Mathematics for Sensing Imaging and Modelling grant 359182 and grant 359186). BMA was supported by RCF under grant number 371523. This study is part of the Geoenergy subtopic within the MTET-Materials and Technologies for the Energy Transition program of the Helmholtz Association. We thank V. Hopiavuori for providing Fig. 4. We thank the editor H. Yao and reviewers A. Fichtner and F. Tilmann for comments that helped to improve the manuscript. Language editing was partially supported using ChatGPT (OpenAI 2026); the responsibility for the content remains with the authors.

SUPPORTING INFORMATION

Supplementary data are available at [GJIRAS](https://www.gjiras.com) online.

suppl_data

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

DATA AVAILABILITY

The passive seismic data set used in this study is openly available from the Finnish Fairdata Etsin service. It contains the unprocessed continuous recordings from the passive imaging component of the COGITO-MIN project, conducted in the Kylylahti polymetallic mine area, eastern Finland, between August and September 2016. The data set is described in the associated metadata record (M. Chamarczuk 2021). The processed Rayleigh-wave phase-velocity measurements at 6 Hz used as input for the tomography inversion are provided in the online Supplementary Material as Supplementary Table S1 in the machine-readable ASCII file `Supplementary_Table_S1_input_6Hz.txt`. The source code used to perform the synthetic experiments with the uniformly spaced station layout is available at <https://github.com/hassanyazdanian/inference-mode-ant>.

REFERENCES

- Aki, K., 1957. Space and time spectra of stationary stochastic waves, with special reference to microtremors, *Bull. Earthquake Res. Inst.*, **35**, 415–456.
- Antil, H. & Saibaba, A.K., 2024. Efficient algorithms for Bayesian inverse problems with Whittle–Matérn priors, *SIAM J. Sci. Comput.*, **46**(2), S176–S198.
- Bensen, G., Ritzwoller, M., Barmin, M., Levshin, A.L., Lin, F., Moschetti, M., Shapiro, N. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, **169**(3), 1239–1260.
- Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, **178**(3), 1411–1436.
- Bodin, T., Sambridge, M., Rawlinson, N. & Arroucau, P., 2012. Transdimensional tomography with unknown data noise, *Geophys. J. Int.*, **189**(3), 1536–1556.
- Byrd, R.H., Lu, P., Nocedal, J. & Zhu, C., 1995. A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.*, **16**(5), 1190–1208.
- Cabrera-Pérez, I., D’Auria, L., Soubestre, J., Barrancos, J., Padilla, G.D. & Pérez, N.M., 2021. A nonlinear multiscale inversion approach for ambient noise tomography, *Geophys. J. Int.*, **225**(2), 1158–1173.
- Chamarczuk, M., 2021. Kylylahti 3D virtual seismic survey (unprocessed ambient-noise recordings), Institute of Geophysics PAS. [10.23729/48acb337-3be0-4e76-94f9-5e60779c26fe](https://doi.org/10.23729/48acb337-3be0-4e76-94f9-5e60779c26fe)
- Chamarczuk, M., Malinowski, M., Draganov, D., Koivisto, E., Heinonen, S. & Röttsä, S., 2022. Reflection imaging of complex geology in a crystalline environment using virtual-source seismology: case study from the Kylylahti polymetallic mine, Finland, *Solid Earth*, **13**(3), 705–723.
- Dales, P. et al., 2020. Virtual sources of body waves from noise correlations in a mineral exploration context, *Seismol. Res. Lett.*, **91**(4), 2278–2286.
- Dunlop, M.M., Iglesias, M.A. & Stuart, A.M., 2017. Hierarchical Bayesian level set inversion, *Stat. Comput.*, **27**(6), 1555–1584.
- Ekström, G., Abers, G.A. & Webb, S.C., 2009. Determination of surface-wave phase velocities across USArray from noise and Aki’s spectral formulation, *Geophys. Res. Lett.*, **36**(18), L18301.
- Erkan, K., 2008. A comparative overview of geophysical methods, *Geodetic Science and Surveying*, The Ohio State University, Columbus, Ohio.
- Esteve, C., Lu, Y., Gosselin, J.M., Kramer, R., Aiman, Y. & Bokelmann, G., 2025a. The seismic signature and geothermal potential of the Schwechat Depression in the Vienna Basin, Austria, from ambient noise tomography, *Geothermics*, **127**, 103–211.
- Esteve, C., Lu, Y., Gosselin, J.M., Kramer, R., Bokelmann, G. & Götzl, G., 2025b. Seismic imaging of the Southern Vienna Basin (Austria) using probabilistic ambient-noise tomography, *Geophys. Prospect.*, **73**(7), e70074.
- Everett, M.E., 2013. *Near-Surface Applied Geophysics*, Cambridge University Press.
- Fang, H., Yao, H., Zhang, H., Huang, Y.C. & van der Hilst, R.D., 2015. Direct inversion of surface wave dispersion for three-dimensional shallow crustal structure based on ray tracing: methodology and application, *Geophys. J. Int.*, **201**(3), 1251–1263.

- Fichtner, A., Zunino, A. & Gebraad, L., 2019. Hamiltonian Monte Carlo solution of tomographic inverse problems, *Geophys. J. Int.*, **216**(2), 1344–1363.
- Hairer, E., Hochbruck, M., Iserles, A. & Lubich, C., 2006. Geometric numerical integration, *Oberwolfach Reports*, **3**(1), 805–882.
- Hansen, P., 2001. *The L-Curve and its Use in the Numerical Treatment of Inverse Problems*, pp. 119–142, WIT Press.
- Hawkins, R., Bodin, T., Sambridge, M., Choblet, G. & Husson, L., 2019. Trans-dimensional surface reconstruction with different classes of parameterization, *Geochem. Geophys. Geosyst.*, **20**(1), 505–529.
- Hoffman, M.D. & Gelman, A., 2014. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo, *J. Mach. Learn. Res.*, **15**(1), 1593–1623.
- Hopilaavuori, V. et al., 2025. Rayleigh wave focal spot imaging of the Kylylahti ore deposit, *EGU General Assembly Conference Abstracts*, pp. EGU25–6190.
- Ibragimov, I.A. & Rozanov, Y.A., 2012. *Gaussian Random Processes*, Springer-Verlag.
- Iliev, A., Kyurkchiev, N. & Markov, S., 2017. On the approximation of the step function by some sigmoid functions, *Math. Comput. Simul.*, **133**, 223–234.
- Jiang, W. & Zhang, J., 2017. First-arrival traveltimes tomography with modified total-variation regularization, *Geophys. Prospect.*, **65**(5), 1138–1154.
- Kaipio, J.P. & Somersalo, E., 2005. *Statistical and Computational Inverse Problems*, Springer-Verlag.
- Korostelev, F., Lu, Y., Magrini, F., Boschi, L., Leroy, S. & Vétel, W., 2022. Images of the East African Rift System by global adaptive-resolution surface-wave tomography, *J. Geophys. Res.: Solid Earth*, **127**(6), e2021JB023570.
- Liu, Y., Wang, B., Ying, D., Zhu, L., Wang, J. & Zhao, T., 2025. Uncertainty-quantified 3D ambient noise tomography using transdimensional Monte Carlo inversion, *Front. Earth Sci.*, **13**, 1660 737.
- Lu, Y., Stehly, L., Paul, A. & Group, AlpArray Working, 2018. High-resolution surface wave tomography of the European crust and uppermost mantle from ambient seismic noise, *Geophys. J. Int.*, **214**(2), 1136–1150.
- Lu, Y., Stehly, L., Brossier, R., Paul, A. & Group, AlpArray Working, 2020. Imaging Alpine crust using ambient noise wave-equation tomography, *Geophys. J. Int.*, **222**(1), 69–85.
- Maboudi Afkham, B., Riis, N.A.B., Dong, Y. & Hansen, P.C., 2024. Inferring object boundaries and their roughness with uncertainty quantification, *J. Math. Imaging Vis.*, **66**(6), 977–992.
- Maboudi Afkham, B., Soto, T., Karamehmedovic, M. & Roininen, L., 2026. Inhomogeneous priors for Bayesian inverse problems, preprint (arXiv:2602.07856).
- Muir, J.B. & Tsai, V.C., 2020. Geometric and level set tomography using ensemble Kalman inversion, *Geophys. J. Int.*, **220**(2), 967–980.
- Nolet, G., 2008. *A Breviary of Seismic Tomography*, Vol. **324**, Cambridge Press, UK.
- Olver, F.W., 2010. *NIST Handbook of Mathematical Functions Hardback and CD-ROM*, Cambridge University Press.
- OpenAI, 2026. ChatGPT (GPT-5.3) [Large language model], <https://chat.openai.com/>, Accessed: 2026-03-11.
- Owen, A.B., 2013. Monte Carlo Theory, *Methods and Examples*, chap. 11, section 11.11n, Art B. Owen (self-published online), <https://artowen.su.domains/mc/>, last accessed 23 September 2026.
- Paige, C.C. & Saunders, M.A., 1982. LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw. (TOMS)*, **8**(1), 43–71.
- Parker, R.L., 1994. *Geophysical Inverse Theory*, Vol. **1**, Princeton University Press.
- Pasztor, A. et al., 2019. PyTorch: an imperative style, high-performance deep learning library, *Advances in Neural Information Processing Systems*, Vol. **32**, pp. 8024–8035, Curran Associates, Inc.
- Quarteroni, A. & Valli, A., 1994. *Numerical Approximation of Partial Differential Equations*, Springer-Verlag.
- Rasmussen, C.E. & Williams, C.K.I., 2005. *Gaussian Processes for Machine Learning*, The MIT Press.
- Rawlinson, N. & Sambridge, M., 2003. Seismic traveltimes tomography of the crust and lithosphere, *Adv. Geophys.*, **46**, 81–199.
- Reynolds, J.M., 2011. *An Introduction to Applied and Environmental Geophysics*, John Wiley and Sons.
- Roininen, L., Huttunen, J.M. & Lasanen, S., 2014. Whittle-Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography, *Inverse Probl. Imaging*, **8**(2), 561–586.
- Ryberg, T., Kirsch, M., Haberland, C., Tolosana-Delgado, R., Viezzoli, A. & Gloaguen, R., 2022. Ambient seismic noise analysis of LARGE-N data for mineral exploration in the Central Erzgebirge, Germany, *Solid Earth*, **13**(3), 519–533.
- Sambridge, M. & Guðmundsson, Ó., 1998. Tomographic systems of equations with irregular cells, *J. Geophys. Res.: Solid Earth*, **103**(B1), 773–781.
- Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**(5715), 1615–1618.
- Stuart, A.M., 2010. Inverse problems: a Bayesian perspective, *Acta Numerica*, **19**, 451–559.
- Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.*, **20**(2), 219–232.
- Teodor, D. et al., 2022. High-resolution passive surface wave tomography using natural and anthropogenic noise sources in a mineral exploration context. *SEG International Exposition and Annual Meeting*, pp. D011S084R001, SEG.
- Tong, M.S. & Chew, W.C., 2020. *The Nyström Method in Electromagnetics*, John Wiley and Sons.
- Trampert, J. & Snieder, R., 1996. Model estimations biased by truncated expansions: possible artifacts in seismic tomography, *Science*, **271**(5253), 1257–1260.
- Tsai, V.C., 2023. The future of Earth imaging, *Seismol. Res. Lett.*, **94**(5), 2119–2128.
- Valentine, A.P. & Sambridge, M., 2018. Optimal regularization for a class of linear inverse problem, *Geophys. J. Int.*, **215**(2), 1003–1021.
- Wang, K., Yang, Y., Jiang, C., Wang, Y., Tong, P., Liu, T. & Liu, Q., 2021. Adjoint tomography of ambient noise data and teleseismic P waves: methodology and applications to central California, *J. Geophys. Res.: Solid Earth*, **126**(6), e2021JB021648.
- Wang, Y., Wang, X., Suwen, J., Deng, B., Chen, Q.F., Xu, T. & Wang, Z., 2025. Geologically guided ambient noise tomography inversion with 3-D interface structures: methodology and application to a gold mine region in China, *Geophys. J. Int.*, **243**(3), ggaf376.
- Wang, Z., Bovik, A.C., Sheikh, H.R. & Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.*, **13**(4), 600–612.
- Zhang, H. & Thurber, C., 2005. Adaptive mesh seismic tomography based on tetrahedral and Voronoi diagrams: application to Parkfield, California, *J. Geophys. Res.: Solid Earth*, **110**(B4).

- Zhang, X. & Curtis, A., 2022. Interrogating probabilistic inversion results for subsurface structural information, *Geophys. J. Int.*, **229**(2), 750–757.
- Zhang, X., Curtis, A., Galetti, E. & De Ridder, S., 2018. 3-D Monte Carlo surface wave tomography, *Geophys. J. Int.*, **215**(3), 1644–1658.
- Zhao, X., Irvin, L., Galetti, E. & Curtis, A., 2026. Direct 3-D variational Bayesian surface wave inversion and its application to ambient noise tomography beneath Great Britain, *Geophys. J. Int.* **245**(2), ggag079.
- Zunino, A., Gebraad, L., Ghirotto, A. & Fichtner, A., 2023. HMCLab: a framework for solving diverse geophysical inverse problems using the Hamiltonian Monte Carlo method, *Geophys. J. Int.*, **235**(3), 2979–2991.