Refining tomography with generative neural networks trained from geodynamics

Théo Santos, Thomas Bodin, F Soulez, Y Ricard, Yann Capdeville

Geophysical Journal International

Geophys. J. Int. (2024) **238,** 1676–1695
Advance Access publication 2024 July 19
GJI Seismology

https://doi.org/10.1093/gji/ggae240

Refining tomography with generative neural networks trained from geodynamics

T. Santos [1,2] T. Bodin [1] F. Soulez [2] Y. Ricard [1] and Y. Capdeville [3]

[1] *Univ Lyon 1, ENS de Lyon, CNRS, Laboratoire de Géologie de Lyon - Terre Planètes Environnement, UMR 5276, F-69622, Villeurbanne, France. E-mail: theo.santos@univ-lyon1.fr*
[2] *Univ Lyon 1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon, UMR 5574, F-69230, Saint-Genis-Laval, France*
[3] *Nantes Université, Univ Angers, Le Mans Université, CNRS, Laboratoire de Planétologie et Géosciences, LPG UMR 6112, 44000 Nantes, France*

Accepted 2024 July 2. Received 2024 June 21; in original form 2024 February 14

SUMMARY

Inverse problems occur in many fields of geophysics, wherein surface observations are used to infer the internal structure of the Earth. Given the non-linearity and non-uniqueness inherent in these problems, a standard strategy is to incorporate *a priori* information regarding the unknown model. Sometimes a solution is obtained by imposing that the inverted model remains close to a reference model and with smooth lateral variations (e.g. a correlation length or a minimal wavelength are imposed). This approach forbids the presence of strong gradients or discontinuities in the recovered model. Admittedly, discontinuities, such as interfaces between layers, or shapes of geological provinces or of geological objects such as slabs can be *a priori* imposed or even suggested by the data themselves. This is however limited to a small set of possible constraints. For example, it would be very challenging and computationally expensive to perform a tomographic inversion where the subducting slabs would have possible top discontinuities with unknown shapes. The problem seems formidable because one cannot even imagine how to sample the prior space: is each specific slab continuous or broken into different portions having their own interfaces? No continuous set of parameters seems to describe all the possible interfaces that we could consider. To circumvent these questions, we propose to train a Generative Adversarial neural Network (GAN) to generate models from a geologically plausible prior distribution obtained from geodynamic simulations. In a Bayesian framework, a Markov chain Monte Carlo algorithm is used to sample the low-dimensional model space depicting the ensemble of potential geological models. This enables the integration of intricate *a priori* information, parametrized within a low-dimensional model space conducive to efficient sampling. The application of this approach is demonstrated in the context of a downscaling problem, where the objective is to infer small-scale geological structures from a smooth seismic tomographic image.

Key words: Bayesian inference; Inverse theory; Machine learning; Statistical methods; Tomography.

1 INTRODUCTION

In seismic tomography, surface recordings of seismic waves are used to image the structure of the Earth (for comprehensive reviews, see Rawlinson *et al.* 2010; Liu & Gu 2012; Tromp 2020; Tsai 2023). Thanks to recent advances in the quality and quantity of seismic observations (Hetényi *et al.* 2018; Schmandt *et al.* 2019), to a number of theoretical breakthroughs in seismic waves modelling (Bozdağ *et al.* 2011; Fichtner & Trampert 2011; Lebedev *et al.* 2023), and the growing availability of computing power, seismologists are now obtaining increasingly detailed constraints on the 3-D structure inside the Earth (Bodin *et al.* 2015). In the mantle, several key features have been identified, including subducted plates penetrating deep into the mantle (Van der Hilst *et al.* 1997; Fukao & Obayashi 2013), or plumes connecting large low shear velocity provinces at the base of the mantle to hotspots at the surface (Bijwaard & Spakman 1999; Montelli *et al.* 2004; French & Romanowicz 2015). These discoveries are offering important insight into the mechanisms of plate tectonics and patterns of mantle convection.

1.1 Imaging a smooth effective Earth

Although the long-wavelength structure of the mantle ($\sim$ 1000 km) is well resolved by various global models (Dalton *et al.* 2008; Ritsema *et al.* 2011; Durand *et al.* 2017), there are still a number of debates in seismology and geodynamics about shorter wavelength structures and the associated processes at play, such as the fate of slabs (Grand *et al.* 1997; Kárason & Van Der Hilst 2000; Faccenda *et al.* 2009; Fukao & Obayashi 2013; Ballmer *et al.* 2015), the structure of continents (Aulbach 2012; Priestley *et al.* 2021; Boyce *et al.* 2024), the shape of plumes (French & Romanowicz 2015; Maguire *et al.* 2018; Ritsema *et al.* 2021) or the origin of large low shear velocity provinces (LLSVP) at the base of the mantle (Wenk *et al.* 2011; Koelemeijer *et al.* 2012; Cottaar & Romanowicz 2013; Thorne *et al.* 2013; Davies *et al.* 2015; Lau *et al.* 2017; Jones *et al.* 2020; Davaille & Romanowicz 2020). The Earth is heterogeneous at all scales, from crystals that make up rocks, to the tectonic blocks that make up continents, to the scale of entire tectonic plates (Schaeffer & Lebedev 2015). A well-known problem is that tomographic models obtained from long period seismograms do not depict this full spectrum of heterogeneities, but at most a smooth (without high spatial frequencies) effective (i.e. equivalent to the reality for the observed wavefield) medium, that is only accurate at long-wavelengths, and that is not suited for detailed interpretation (Ricard *et al.* 1996; Bodin *et al.* 2015; Capdeville & Métivier 2018; Tsai 2023). Structures at small scales cannot be resolved, are non-linearly mapped in the recovered long wavelength structures and subject to non-uniqueness. For example, it is well known that a stack of isotropic layers will be equivalent, at large scale, to a homogeneous anisotropic medium (Backus 1962) (which does not preclude the existence of intrinsic anisotropy, Mainprice *et al.* 2000). When the medium is observed at large wavelength, the ability to distinguish different layers disappears. Thus, part of the anisotropy observed at large scale may be artificial and simply the effect of unmapped fine layering (Fichtner *et al.* 2013; Alder *et al.* 2017; Magali *et al.* 2021). This may be the case in areas where a lot of fine geological layering is involved.

In this context, mantle models obtained from long period waves often lack the signature of compositional discontinuities and deformation mechanisms, and hence lack geological accuracy. They are difficult to interpret, and to compare to results from other disciplines such as geodynamics or mineralogy (Kárason & Van Der Hilst 2000; Garnero & McNamara 2008; McNamara *et al.* 2001; Thorne *et al.* 2013; Ballmer *et al.* 2015; Priestley *et al.* 2021; Wolf *et al.* 2024).

1.2 Adding information about small scales

Standard tomographic approaches use finite discrete basis (e.g. voxels, spherical harmonics), which imposes a limit on the scales that can be resolved. Thus traditional models are inevitably designed to capture a low-pass filtered version of the true earth. The most common way to include information about structures that cannot be resolved by the data is to use a reference model and to penalize solutions that are far from it (Ricard *et al.* 1996; Asnaashari *et al.* 2013). But a single reference model is often not adequate to fully represent our state of knowledge about small-scale geological structures.

Otherwise, there are a number of parametrization and regularization approaches to enhance discontinuities in reconstructed images. But their application to seismic tomography remains limited as they make the inverse problem highly non-linear and its resolution computationally challenging. In the case of full waveform inversion, some studies proposed a total variation regularization to preserve sharp interface of subsurface structures (e.g. Burstedde & Ghattas 2009; Guitton 2012; Lin & Huang 2014; Esser *et al.* 2018; Kalita *et al.* 2019). Many other strategies have also been proposed and tested such as wavelet parametrization with sparsity constraints (Simons *et al.* 2011; Charlety *et al.* 2013), compressed sensing (Li *et al.* 2012), Cauchy function regularization (Guitton 2012), model reduction (Barnier *et al.* 2019) and level set parametrizations (Muir & Tsai 2020). However, these approaches often have their limitations, using either structures that are too simple, or a number of parameters that is too large for an easy numerical implementation. Another common problem is that these methods solve an optimization problem where the solution is a single model that minimizes an objective function based on data derivatives and a regularization term. In this context, uncertainty analysis is difficult as the range of geologically plausible solutions that explain the data cannot be properly described and non-uniqueness can not be properly accounted for.

Some approaches have been used to produce seismic models that can be directly interpreted in terms of geological, mineralogical, or geodynamic processes. In this case, the produced Earth model is the direct result of a physical process and can therefore be interpreted geologically. Tsai *et al.* (2023) proposed to directly parametrize a crustal model in terms of blocks separated by faults. Munch *et al.* (2018) and Bissig *et al.* (2022) inverted body waves measurements for temperature and composition in the mantle. Magali *et al.* (2021) proposed to use geodynamic simulations for the parametrization, where the patterns of seismic anisotropy were solutions of Stoke's flow equations. In these approaches the relation between model parameters and seismic observations becomes highly non-linear, and sampling methods (e.g. Monte Carlo) may be used instead of standard linearized inversion schemes. Although this enables the quantification of uncertainties, sampling approaches can be computationally expensive if the number of parameters is large. This is aggravated by the well-known curse of dimensionality, which makes the sampling difficult at high dimensions (Curtis & Lomax 2001; Scheiter *et al.* 2022).

We also note that some approaches from geostatistics such as multiple point statistics (González *et al.* 2008; Bosch *et al.* 2010; Cordua *et al.* 2012; Lochbühler *et al.* 2014, 2015) or hidden Markov models (Feng *et al.* 2018) have been used to interpret seismic models in terms of geological facies, while including complex geological information. But here again, exploring the ensemble of possibilities that explain the data remains computationally expensive. Indeed, in these methods, each pixel is individually reconstructed with conditional probabilities, leading to multiple computations for only constructing one possible image (Mariethoz & Caers 2014).

As a summary, the multiscale parametrization of seismic models needs to address the following challenges:

(i) The inclusion of geologically accurate *a priori* information about small scales.
(ii) A low dimensionality of the model space.
(iii) A computationally cheap relationship between model parameters and the actual velocity model.
(iv) The reduction of non-linearities in the relationship between model parameters and observations.

1.3 Deep generative models

In order to address these parametrization challenges (small-scale geologically accurate *a priori*, low dimensionality, reduced computational cost and mitigated non-linearities), a number of recent studies have used deep generative models (Foster 2022): neural networks that are able to generate new images statistically similar to that of a chosen set of images, called training set. In geophysical inverse problems, to reconstruct geologically realistic earth models, Liu *et al.* (2022) used a Variational Auto-Encoder (VAE, a generative neural network that learns to encode input images into a low-dimensional space and from each point of this space, associate a new similar image, Kingma & Welling 2013). Others (Richardson 2018; Chan & Elsheikh 2019; Mosser *et al.* 2020; Lopez-Alvis *et al.* 2021; Bloem *et al.* 2023; Feng *et al.* 2024) preferred a Generative Adversarial Network (GAN, a generative neural network trained in an adversarial relationship with a discriminator which tries to identify the fake generated images, Goodfellow *et al.* 2014). Since the unknown model is described by a low number of parameters, it becomes possible to use a Markov chain Monte Carlo (McMC) algorithm to explore the range of possible solutions and embrace the non-uniqueness of the problem (Mosegaard & Tarantola 1995). In this case, the inverse problem can be defined in a Bayesian framework and the generative model can be seen as a tool to sample from the prior distribution (e.g. Patel & Oberai 2019; Bohra *et al.* 2022). Generative models can also be used in similar approaches to reduce the dimension of the data or the parameter spaces (Valentine & Trampert 2012; Chen & Saygin 2021; Scheiter *et al.* 2022).

In this study, we use a GAN trained on images from geodynamic simulations to efficiently explore the space of possible models. By generating random models from a low-dimensional space, our method aims to depict shapes representing chemical discontinuities in a convecting mantle, thus addressing the mentioned parametrization challenges. We follow the ideas of Capdeville & Métivier (2018) and Hedjazian *et al.* (2021), and view the tomographic problem as a two-step approach, consisting on first recovering a smooth effective model that gathers the information present in the data, before reconstructing geologically realistic small-scale structures from the smooth effective model. Our work focuses on the second step, called the downscaling problem. We cast this problem in a Bayesian framework, where the ensemble of potential fine-scale models sharing the same long wavelength effective properties are explored with a McMC algorithm, using the GAN for the parametrization.

To illustrate the potentiality of the method, we use a simple 2-D geodynamic toy problem, and train a generative neural network with a set of numerical simulations of 2-D sections of the upper mantle, the so-called 'marble cake models' (Allègre & Turcotte 1986; Alder *et al.* 2017; Magali *et al.* 2021), where the stirring of crustal and lithospheric components continuously introduced in the deep mantle by subductions leads to a veined structure reminiscent of a marble cake. Our goal is then to find the 'best' marble cake structure that would correspond to a given smooth tomographic image. In this way, this parametrization of marble cake models with a generative neural network allows us to downscale a synthetic smooth tomographic image.

In this paper, we introduce and illustrate the proposed method in five sections. In Section 2, we detail the geodynamic simulations leading to marble cake structures. We also discuss the elastic homogenization that enables us to compute a long wavelength effective medium, and thus to approximate the tomographic operator. Section 3 discusses the generation of artificial marble cake structures with Deep Learning. Section 4 details the methodology proposed for the inversion. Section 5 illustrates the benefits of the proposed method on the simple synthetic problem, and results are discussed in the last section.

2 FROM GEOLOGICAL STRUCTURES TO A TOMOGRAPHIC IMAGE

2.1 Simulating geological structures: the marble cake mantle model

Since the seminal work of Batchelor (1959), it is well accepted that mantle heterogeneities follow a $1/k$ power spectrum, where k is the wavenumber (Olson *et al.* 1984; Antonsen & Ott 1991). Allègre & Turcotte (1986) introduced a model for the upper mantle, resulting from the stirring of two components with basaltic and harzburgitic compositions and depicting marble cake structures. Mancinelli *et al.* (2016) showed that this structural configuration is responsible for the observed $1/k$ power spectrum behaviour, for wavelengths down to the kilometre. Therefore, for the purpose of this study, we approximate mantle structures, seen on a cross-section, as a binary geological system.

Here we adopt the numerical modelling framework described in Alder *et al.* (2017) to reproduce a 2-D representation of these marble cake structures, as illustrated in Fig. 1. This modelling approach involves the stirring of a circular anomaly by the chaotic convection mechanism into a confined spatial domain. For this study, the two components are in the same proportion, and the confined domain is a $L \times L$ square region. The convection field within the model is defined by a simple incompressible chaotic flow where the horizontal and vertical velocities are:

$$v_x = \frac{\partial \Psi}{\partial y}\,, \tag{1}$$

$$v_y = -\frac{\partial \Psi}{\partial x}\,, \tag{2}$$

derived from the stream function Ψ

$$\Psi(x, y, t) = \sin(\pi y)[\sin(2\pi x) + a(t)\sin(3\pi x) + b(t)\sin(4\pi x)]\,, \tag{3}$$

where $a(t)$ and $b(t)$ are sinusoidal functions of time:

$$a(t) = A\ \sin(f_a\, t + \phi_a)\,, \tag{4}$$

$$b(t) = B\ \cos(f_b\, t + \phi_b)\,. \tag{5}$$

These functions a and b, with their 6 parameters: amplitudes A, B, frequencies f_a, f_b and phases ϕ_a, ϕ_b introduce chaotic behaviour into the convection. The initial circular anomaly undergoes a sequence of deformations according to the prescribed velocities. The number of points defining the interface between the two components is increased over time, as the contour lengthens and becomes more convoluted. The resulting configuration gives rise to a binary image featuring a distinctive marble cake. This toy model is used as the marble cake geodynamic simulator $\mathcal{S}$. This simulator takes 9 input parameters gathered in the vector $\boldsymbol{p}$, including the initial position of a fixed-size circular anomaly, the six parameters defining the convection field and the number of time step. In response, it generates a binary image of N pixels, either 0 (black) or 1 (white). A visual representation of this convective construction process at various time intervals is provided in Fig. 1.

Both phases of the material are assigned an isotropic S-waves velocity v_s, an isotropic P-waves velocity v_p and a density ρ. We choose $v_{s,1} = 4.13\,\mathrm{km\,s^{-1}}$ and $v_{s,2} = 3.38\,\mathrm{km\,s^{-1}}$, to reach

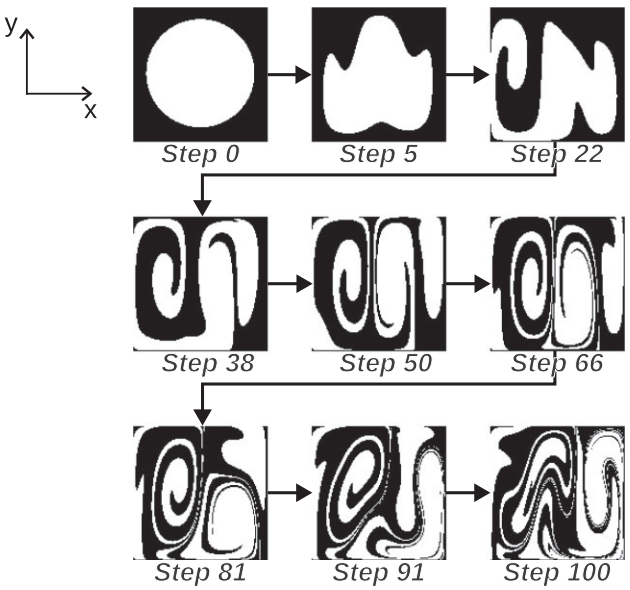

Figure 1. Illustration of the convective construction of a marble cake shape, at various time steps. It is obtained by stirring a circular anomaly with chaotic convection. The simulated marble cake is the final image. This particular example of marble cake was obtained with the following parameters : initial anomaly centred in (0.50696, 0.49770), $A = 1.04568$, $B = 0.88601$, $f_a = 0.72566$, $f_b = 0.55985$, $\phi_a = 0.12535$ and $\phi_b = 0.22108$.

10 per cent contrast of velocity which is in the upper range of what is expected in the upper mantle (Stixrude & Jeanloz 2007; Xu *et al.* 2008). The corresponding P velocities are given by $v_p/v_s = 1.8$. We finally fix $\rho = 3$ g cm^{-3}. In order to fully characterize the mechanical behaviour of the medium, we define a local elasticity tensor. In the isotropic case, the Lame's coefficients are:

$$\lambda = \rho \left(v_p^2 - 2v_s^2\right) , \tag{6}$$

$$\mu = \rho v_s^2 . \tag{7}$$

In the 2-D case, the isotropic elastic tensor expressed at each point of the medium, using Kelvin notation is

$$\boldsymbol{C} = \begin{pmatrix} \lambda + 2\mu & \lambda & 0 \\ \lambda & \lambda + 2\mu & 0 \\ 0 & 0 & 2\mu \end{pmatrix} . \tag{8}$$

2.2 Approximating the tomographic operator: the elastic homogenization

For each marble cake structure, we could in principle propagate seismic waves and obtain the synthetic seismic observables. However, we cannot directly use these synthetics to mimic a real experiment. Indeed, due to practical, theoretical and computational limitations, only a limited bandwidth of the seismic signal can be observed and inverted (this is true for both synthetic and real seismograms). In addition to these limitations, the imperfect data coverage is also an obvious problem with real observations (sources and observations are unevenly and too sparsely distributed). That is why currently, the images obtained with seismic tomography depict only a smooth effective representation of the true Earth. Even assuming a perfect data coverage, the limited bandwidth that can be used, suggests to use a two-steps procedure to compute the propagation of seismic waves in a marble cake structure:

(i) From the fine-scale structure, compute a smooth medium that is equivalent for the seismic wavelengths used in the tomographic inversion.

(ii) Use this effective medium for the wave propagation simulation. This makes the computations easier and leads to synthetics comparable with real observations.

The first step can be done with Fast Fourier homogenization (Capdeville *et al.* 2015). The input of this operation is the seismic medium (3×3 elastic tensor $\boldsymbol{C}$, defined at each point, see Section 2.1). The output is the homogenized anisotropic elasticity tensor $\boldsymbol{C}^*$, also defined in each point, that we call the effective medium. We illustrate an example of an effective medium in Fig. 2. We show the initial elastic tensor $\boldsymbol{C}$ and its homogenize counterpart $\boldsymbol{C}^*$. Both 3×3 tensors are symmetric (i.e. have 6 independent components) and spatial variations of the initial tensor are smoothed by the homogenization. Notice that the homogenization of a complex but isotropic medium leads to a fully anisotropic medium (Backus 1962; Magali *et al.* 2021). At any given location, the level of anisotropy of the tensor can be quantified. Indeed, the coefficients C^*_{11} are proportional to the squared velocity of P waves along the x-axis and C^*_{22} along the y-axis. $\xi = C^*_{22}/C^*_{11}$ corresponds to the anisotropy of the P waves, caused by small-scale layering resulting in a preferential direction. An example of the anisotropy for a homogenized tensor is provided in Fig. 3.

Assuming perfectly data-covered tomography, we follow Capdeville *et al.* (2013) and Capdeville & Métivier (2018) that show how elastic homogenization can provide the best image that one would get from a tomographic inversion; homogenization is equivalent to a 'tomographic operator'. Therefore, with this operator, it is possible, from a fine-scale structure, to compute the equivalent tomographic image, that is the image that would be obtained from a tomographic inversion.

The main objective of this study is to recover the underlying fine-scale marble cake mantle structure $\boldsymbol{m}$ from a low-resolution and homogenized tomographic image, that is the effective medium $\boldsymbol{t}$. Both are linked through the forward problem $\mathcal{H}$, illustrated in Fig. 2, and expressed as followed:

$$\boldsymbol{t} = \mathcal{H}(\boldsymbol{m}) + \boldsymbol{\epsilon} , \tag{9}$$

where $\boldsymbol{\epsilon}$ is an error term that will be detailed in Section 4.2. Although $\boldsymbol{m}$ is an array of isotropic velocities, $\boldsymbol{t}$ includes the six components of the homogenized velocity.

Note that the equation symbolizing the forward problem expressed by eq. (9) and the inverse problem that we want to deal with (somehow defining an inverse $\widetilde{\mathcal{H}}$ operator) look trivial but in fact are not. There is no safe way to sample all possible marble cake patterns, there is no obvious mathematical distance to define between two marble cake patterns and we cannot make sense of the differentiation of $\mathcal{H}(\boldsymbol{m})$ which is often necessary in inverse theory. However we will see that a deep generative network can perform the remarkable trick of describing $\boldsymbol{m}$ as a function of a vector $\boldsymbol{z}$ in a space of moderate dimensions, as discussed in the next section.

3 GENERATING ARTIFICIAL GEOLOGICAL STRUCTURES

3.1 A low-dimensional parametrization

Downscaling a tomographic image present numerous challenges. Indeed, as we described it in Section 2.2, there is information loss during the homogenization $\mathcal{H}$ that we try to inverse. Many different

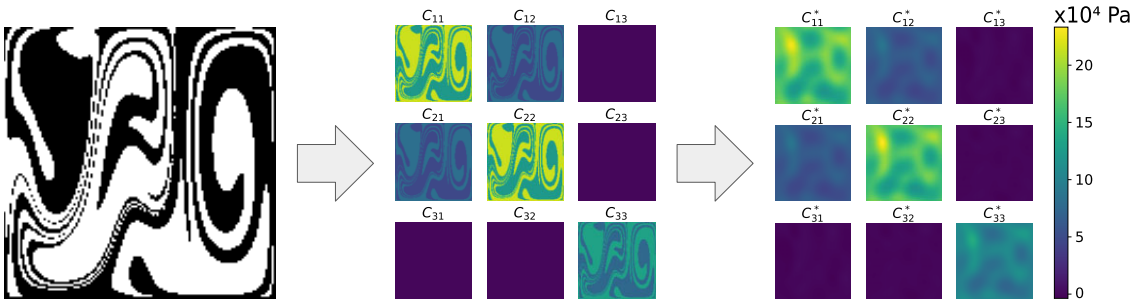

Figure 2. Summary of the physical forward problem. Left-hand panel: marble cake model. Middle panel: corresponding elastic medium, computed from v_s, v_p, ρ of each phase. Right-hand panel: effective medium, obtained with homogenization of the seismic medium, for $\lambda_0 = L/5$, where L is the side length of the box. For the wave equation, and down to a minimum period, this effective medium and the seismic medium (middle panel) are equivalent.

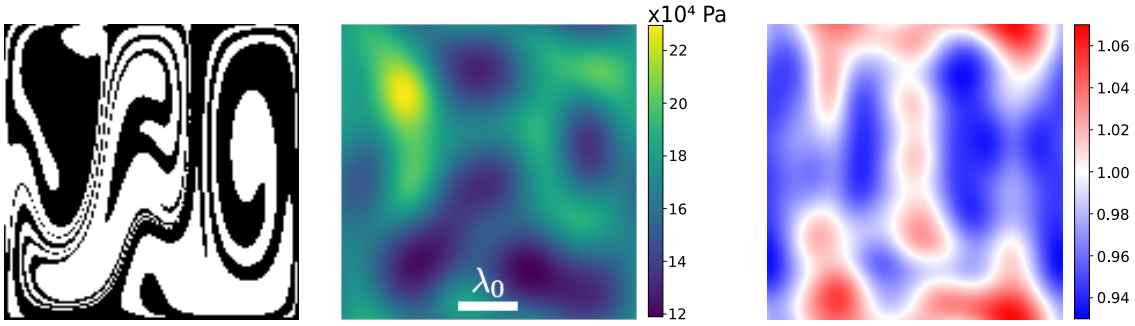

Figure 3. Effective medium and anisotropy of a marble cake image for homogenization at $\lambda_0 = L/5$, L being the side length of the box. Left-hand panel: a marble cake model, composed by two phases: an enriched component transformed into a dense assemblage (white) and a harburzgite phase (black). Middle panel: effective medium, obtained by homogenization of the marble cake at λ_0. No structures smaller than λ_0 are kept. Only the first coefficient C^*_{11} is represented on this figure. Right-hand panel: anisotropy of the effective medium, computed as the ratio $\xi = C^*_{22}/C^*_{11}$. The red and blue areas correspond to horizontal and vertical layering, respectively.

fine-scale models can homogenize to the same smooth tomographic image. To restrain the possible solutions to 'realistic' models, we want to restrict the solution to the manifold of images that look like convection models such as the one described in Section 2.1. To do this, we want to construct, by means of a deep generative network, a function $\mathcal{G}(\boldsymbol{z})$ that maps a low-dimensional space (called latent space) to the high-dimensional distribution of marble cake images. By sampling the latent vectors $\boldsymbol{z}$ from a distribution $p(\boldsymbol{z})$ (by construction a normal distribution with unit variance), the generator $\mathcal{G}(\boldsymbol{z})$ will sample the prior distribution $p(\boldsymbol{m})$ of all possible marble cake patterns. This generator is designed to possess several desirable attributes: (i) to be computationally efficient and ensure rapid image generation and (ii) to be defined in a latent space with dimensions as small as possible to facilitate efficient and effective sampling.

With this proposed parametrization, we reformulate the forward problem in a Bayesian framework: the aim is now to infer the posterior distribution $p(\boldsymbol{z} \mid \boldsymbol{t})$ of latent vectors $\boldsymbol{z}$. The complete forward problem (see Fig. 4) is the marble cake generation operator $\mathcal{G}$ followed by the homogenization operator $\mathcal{H}$ described in Section 2. Eq. (9) is thus reformulated as:

$$\boldsymbol{t} = \mathcal{H} \circ \mathcal{G}(\boldsymbol{z}) + \boldsymbol{\epsilon}\,. \tag{10}$$

This form of the forward problem is now suitable to an inverse problem. Initially, the difficulty with eq. (9) was to find the probability $p(\boldsymbol{m} \mid \boldsymbol{t})$ that the marble cake $\boldsymbol{m}$ corresponds to the observation $\boldsymbol{t}$. Now the reformulated problem (eq. 10) is to find $p(\boldsymbol{z} \mid \boldsymbol{t})$, the probability that $\boldsymbol{z}$ generates a marble cake whose tomographic image is $\boldsymbol{t}$.

3.2 Generative adversarial network

To build the generator $\mathcal{G}$, we use a Generative Adversarial neural Network (GAN) framework. This process involves training the GAN on a set of geodynamically simulated marble cake images, as described in Section 2.1. The geodynamic input parameters used to simulate the training images (see Section 2.1) were taken randomly on independent uniform distributions, with the exception of the stirring duration which was kept constant. This training set of 'true' marble cakes can be seen as a representative sample from the prior distribution. Once trained, the GAN is expected to generate samples $\boldsymbol{m}$ from the same distribution as the training set.

A Generative Adversarial Network, introduced by Goodfellow *et al.* (2014), is a machine learning framework designed for image generation. It consists of two primary components: a generator $\mathcal{G}$ and a discriminator $\mathcal{D}$. Throughout the training process, these two components engage in an adversarial and competitive relationship. The objective of the generator is to produce synthetic images that become progressively more challenging for the discriminator to distinguish from real images. On the other hand, the discriminator aim is to accurately differentiate between real and generated marble cake images. Both networks are successively trained, improving with each step. This framework is illustrated in Fig. 5.

In this study, the marble cake generator is made of a GAN of $Z = 30$ latent variable trained with gradient penalty on 19 456 marble cake models. These models are obtained with the simulator $\mathcal{S}$ from parameters $\boldsymbol{p}$ taken over a uniform distribution. More architecture and implementation details are given in Appendices A and B.

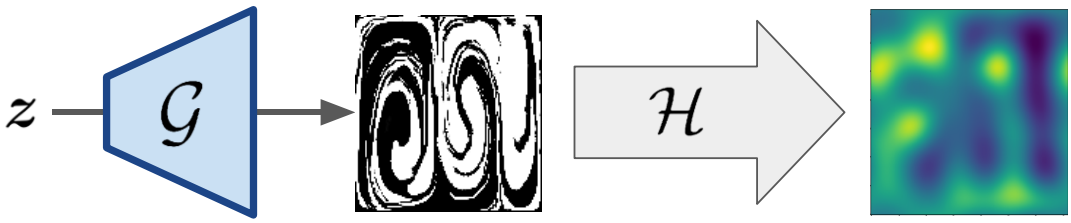

Figure 4. The complete forward problem. The neural network generator $\mathcal{G}$ generates a marble cake $\mathcal{G}(\boldsymbol{z})$ from a small dimensional latent vector $\boldsymbol{z}$. The marble cake is homogenized at a given cutoff wavelength, resulting in the computed data, that is the effective medium: $\mathcal{H} \circ \mathcal{G}(\boldsymbol{z})$. This effective medium can then be compared to the observations $\boldsymbol{t}$. The aim in our experiment is to sample the posterior probability $p(\boldsymbol{z} \mid \boldsymbol{t})$.

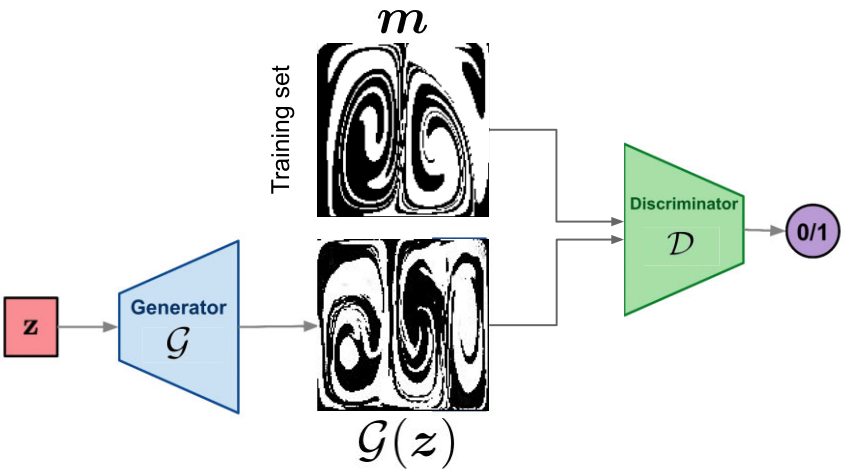

Figure 5. Training framework of a GAN. The discriminator learns to differentiate between training images and generated images, while the generator learns to generate images fooling the discriminator. They are trained successively various times, until the generator is able to produce images looking like the training images, that is to sample the prior $p(\boldsymbol{m})$ from which is taken the training set. In this way, the generator $\mathcal{G}$ is able to map a Gaussian distribution $p(\boldsymbol{z})$ on the latent space onto the prior distribution $p(\boldsymbol{m})$.

3.3 Generator assessment

Some examples of generated images are shown in Fig. 6 together with some images of the training set. The generator is able to mimic the shapes of the simulated marble cakes, providing 'realistic' geological structures. Note that, like the training images, the generated marble cakes have only local values of 0 or 1, although the generator is capable of producing a continuous range of values between 0 and 1. When comparing their respective power spectra on Fig. 7 the GAN reproduces the statistical properties of the training images, displaying a power spectral density that follows a $1/k$ curve fairly precisely, where k is the wavenumber. This is consistent with theoretical expectations (as discussed in Section 2.1 and in references such as Batchelor 1959; Olson *et al.* 1984; Antonsen & Ott 1991).

4 THE INVERSE PROBLEM: DOWNSCALING TOMOGRAPHIC IMAGES

The method proposed for downscaling tomographic images involves estimating the posterior distribution of marble cakes, given the available tomographic measurements. Since the solution to this problem is highly non-unique (as one effective medium can be explained by various fine-scale layered velocity models), it cannot be represented by a single best fitting model. Instead, we follow a Bayesian approach where the solution is represented by a collection of images, which are samples from the posterior probability density $p(\boldsymbol{z} \mid \boldsymbol{t})$ (Tarantola 2005). This approach facilitates a more comprehensive characterization of the uncertainties associated with the estimated marble cake images, embracing the inherent non-uniqueness of the solution.

4.1 Bayesian formulation

Given the tomographic measurements $\boldsymbol{t}$, the statistics of the upper mantle structure $\boldsymbol{m}$ is described by its posterior distribution $p(\boldsymbol{m} \mid \boldsymbol{t})$. As stated in Section, 3.1, this posterior distribution is difficult to estimate in practice, as the space of possible marble cakes $\boldsymbol{m}$ is hard to sample. Using the generator $\mathcal{G}$ defined in Section 3.2 to parametrize $\boldsymbol{m}$, we then define the posterior distribution of the latent variables $p(\boldsymbol{z} \mid \boldsymbol{t})$ where $\boldsymbol{m} = \mathcal{G}(\boldsymbol{z})$. According to Bayes' theorem, this posterior distribution writes:

$$p(\boldsymbol{z} \mid \boldsymbol{t}) \propto p(\boldsymbol{z})\, p(\boldsymbol{t} \mid \boldsymbol{z})\,, \tag{11}$$

where the prior probability distribution $p(\boldsymbol{z})$ quantifies our geological prior knowledge about the unknown model independently of the observations. By construction, and conveniently, it is a Z-dimensional normal distribution:

$$p(\boldsymbol{z}) \propto \exp\left(-\frac{\|\boldsymbol{z}\|^2}{2}\right), \tag{12}$$

where the generator $\mathcal{G}$ has been trained so that by exploring the latent space with probability $p(\boldsymbol{z})$, the generator $\mathcal{G}(\boldsymbol{z})$ samples all the possible marble cake patterns.

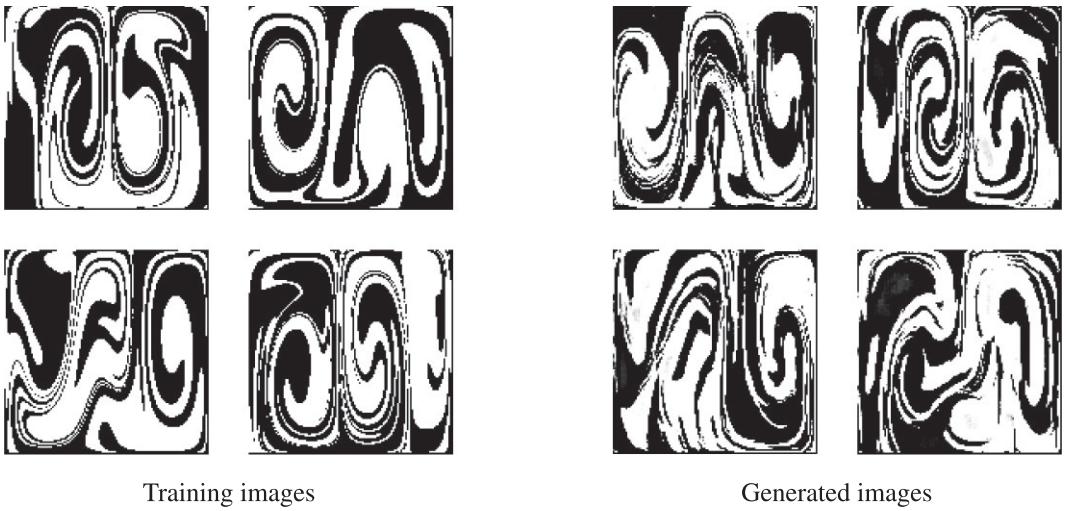

Figure 6. Left-hand panel: examples of training images, obtained from geodynamic simulations described in Section 2.1. These images represent realizations of the prior distribution $p(\boldsymbol{m})$. Right-hand panel: images generated by the generator $\mathcal{G}$ from latent vectors $\boldsymbol{z}$ taken on the distribution $p(\boldsymbol{z})$. These 'artificial marble cakes' depict similar convective structures to the training images.

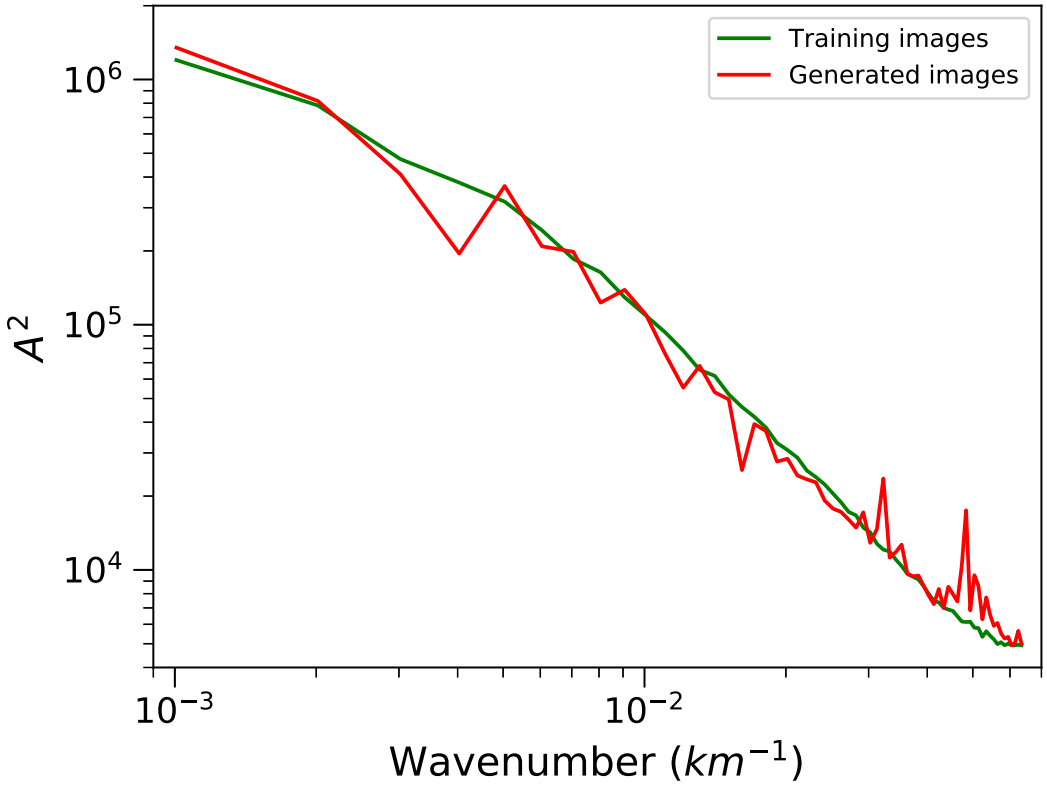

Figure 7. Mean power spectra of heterogeneities, for the set of training images (green) and the generated images (red).

The likelihood distribution $p(\boldsymbol{t} \mid \boldsymbol{z})$ quantifies how the model $\mathcal{G}(\boldsymbol{z})$ fits the geophysical observations $\boldsymbol{t}$. From eq. (10), it writes:

$$p(\boldsymbol{t} \mid \boldsymbol{z}) \propto \exp\left(-\frac{1}{2}\,\|\boldsymbol{t} - \mathcal{H} \circ \mathcal{G}(\boldsymbol{z})\|^2_{\mathbf{Cov}}\right), \quad (13)$$

with the norm :

$$\|\boldsymbol{t} - \mathcal{H} \circ \mathcal{G}(\boldsymbol{z})\|^2_{\mathbf{Cov}} = (\boldsymbol{t} - \mathcal{H} \circ \mathcal{G}(\boldsymbol{z}))^{\mathrm{T}}\mathbf{Cov}^{-1}(\boldsymbol{t} - \mathcal{H} \circ \mathcal{G}(\boldsymbol{z})), \quad (14)$$

where **Cov** is the covariance matrix of the data errors. We recall that the observed tomographic model $\boldsymbol{t}$ is described by 6 images of N pixels.

4.2 Covariance matrix and data errors

We consider that the errors in the six elastic coefficients obtained by homogenization are not correlated with each other. To simplify the numerical procedure, instead of computing the correlations on the N points and computing $\mathbf{Cov}^{-1}$ in eq. (14), we resample the data on a coarse grid such that the distance between two points is half the homogenization wavelength, $\lambda_0/2$. This ensures that the data errors are not spatially correlated and makes the covariance matrix diagonal. Hence, eq. (14) is

approximated by

$$\|\boldsymbol{t} - \mathcal{H} \circ \mathcal{G}(\boldsymbol{z})\|^2_{\mathbf{Cov}} \approx \sum_{i=1,6} \frac{1}{\sigma_i^2} \sum_j (\boldsymbol{t} - \mathcal{H} \circ \mathcal{G}(\boldsymbol{z}))^2_{i,j}, \quad (15)$$

where the sum over i applies to the 6 components of the images and the sum over j to the pixels of the coarse images. The coefficients σ_i can be interpreted as uncertainties with the form $\sigma_i^2 = \sigma_{oi}^2 + \sigma_{Gi}^2$, where σ_{oi} accounts for the observational errors and σ_{Gi} for the errors of the GAN process. We assume that σ_{oi} is typically 20 per cent [i.e. signal-to-noise ratio (SNR) = 14 dB] of the spatial variability of each elastic coefficient estimated for a set of homogenized marble cakes. The σ_{Gi} derives from the inability of the generator to produce exactly a chosen marble cake. We estimate the closest fit to a given marble cake that the generator can get, and compare them for each elastic coefficient after homogenization; allowing us to estimate $\sigma_{Gi} \approx 75$ per cent of the spatial variability. Combining the prior in eq. (12) and the likelihood in eq. (13), the posterior probability density function is expressed as follows:

$$p(\boldsymbol{z} \mid \boldsymbol{t}) \propto \exp\left(-\frac{\|\boldsymbol{z}\|^2}{2}\right) \times \exp\left(-\frac{1}{2}\|\boldsymbol{t} - \mathcal{H} \circ \mathcal{G}(\boldsymbol{z})\|^2_{\mathbf{Cov}}\right). \quad (16)$$

4.3 Sampling the posterior

To estimate the posterior probability density function $p(\boldsymbol{z} \mid \boldsymbol{t})$ in eq. (16), we use a Metropolis–Hastings McMC algorithm. Introduced by Hastings (1970), McMC has become a cornerstone in Bayesian statistical methods, offering an efficient means of approximating complex posterior distributions. It involves constructing a random walk to explore the model space (here the latent space), where each step is a perturbation of the previously visited model. At each step, the complete forward operator is computed for the proposed model and the data simulated for this model are compared to observations. The proposed model is then either accepted or rejected according to an acceptance law. The ensemble of sampled models asymptotically converges towards the posterior distribution. The solution is therefore a large ensemble of low-dimensional vectors in the latent space, from which different statistics (mean model, standard deviations...) can be taken for visualization and interpretation. More detailed explanations and considerations about McMC can be found for example in Brooks *et al.* (2011).

5 RESULTS

The performance of the presented method is assessed on synthetic data. We simulate a marble cake image with the geodynamic simulation presented in Section 2.1. This image $\boldsymbol{m}_t$ is considered to be the true structure to be recovered from its smooth homogenized version. It is homogenized with a cut-off wavelength of $\lambda_0 = L/5$ (where L is the side length of the box) and a SNR = 14 dB additive Gaussian noise. The simulated observation is then the resulting 6 component (3×3 symmetric) tensor $\boldsymbol{t} = \mathcal{H}(\boldsymbol{m}_t) + \boldsymbol{\epsilon}$. The target model $\boldsymbol{m}_t$ and the corresponding synthetic observations used in Sections 5.1, 5.2 and 5.3 are shown in Fig. 8.

5.1 Appraising the posterior solution

The data shown in Fig. 8 is inverted by running in parallel 16 McMCs of 100 000 iterations each to produce an empirical estimate of the posterior distribution $p(\boldsymbol{z} \mid \boldsymbol{t})$. The number of chains and iterations was chosen as a compromise between ensuring the McMC convergence (see more details about the convergence assessment in Appendix C) and reasonable computing time (mostly controlled by the homogenization process).The first 4 of the Z dimensions of the posterior distribution are represented on Fig. 9. On the presented histograms, the sampled posterior distribution has a smaller standard deviation than the prior (that has a unit variance). This is expected from the information gain brought by the observations. In addition, this posterior distribution is bimodal meaning there are at least two classes of marble cake images that explain well the data.

Each sampled latent vector $\boldsymbol{z}$ corresponds to a binary marble cake model $\boldsymbol{m} = \mathcal{G}(\boldsymbol{z})$. Therefore, the complete posterior distribution can be appraised in different ways in the image space. Examples of images randomly drawn from the ensemble solution are shown in Fig. 10. They all share the same overall shapes that match the target structures, and only differ on small scales.

We can also visualize our results by showing different statistical estimates of the distribution in the marble cake space. In Fig. 11, we show the mode and mean of the distribution taken either in the latent space or the image space: $\mathcal{G}(\boldsymbol{z}_m)$, $\mathcal{G}(\overline{\boldsymbol{z}})$, $\mathcal{G}(\boldsymbol{z})_m$, $\overline{\mathcal{G}(\boldsymbol{z})}$. Only the last panel showing $\overline{\mathcal{G}(\boldsymbol{z})}$ has a continuous scale between 0 and 1, that can be interpreted as the probability of the presence of the white phase at each location. Note that the standard deviation of the ensemble solution at each pixel can be used as a map of uncertainties (see Figs 12 and 14).

5.2 Testing different homogenization wavelengths

Here we present results for different homogenization cut-off wavelengths λ_0 (see Fig. 12). As expected, the smaller the cut-off wavelength λ_0, the sharper the mean model and the smaller the uncertainties. Indeed, for short cut-off wavelength, smaller scales are kept, and the solution is better constrained by the measurements.

Conversely, the observations with $\lambda_0 = L/2$ only contain a low level of information, resulting in a low-contrast mean model and large standard deviations. On the histogram of sampled values in the latent space, the posterior distribution is close to the prior, highlighting the fact that solutions are poorly constrained by measurements. Despite these poor constraints, the global structure of the target is recovered, thanks to the high level of constraints given by the *a priori* distribution.

When the cut-off wavelength is equal to the length of the box ($\lambda_0 = L$), there is no information left after the homogenization. As expected, the McMC only samples the prior distribution. The mean image shows no particular structure.

5.3 Effect of anisotropy

As stated in Section 1.1, after homogenization, fine layering in the underlying structures is non-linearly mapped into effective anisotropy. In this way, it is clear that long wavelength anisotropy carries information on the small scales, such as the direction and the amplitude of layering. In order to investigate the impact of anisotropy on the estimated posterior, we conduct a comparative analysis between the anisotropic case, where the complete homogenized elastic tensor is inverted, and the isotropic case where solely the $\boldsymbol{t}_{11}$ coefficient is used. This comparative assessment is performed with a cut-off wavelength of $\lambda_0 = L/2$. The synthetic data are presented in Fig. 13, while the characteristic images of the estimated posterior distribution are depicted in Fig. 14.

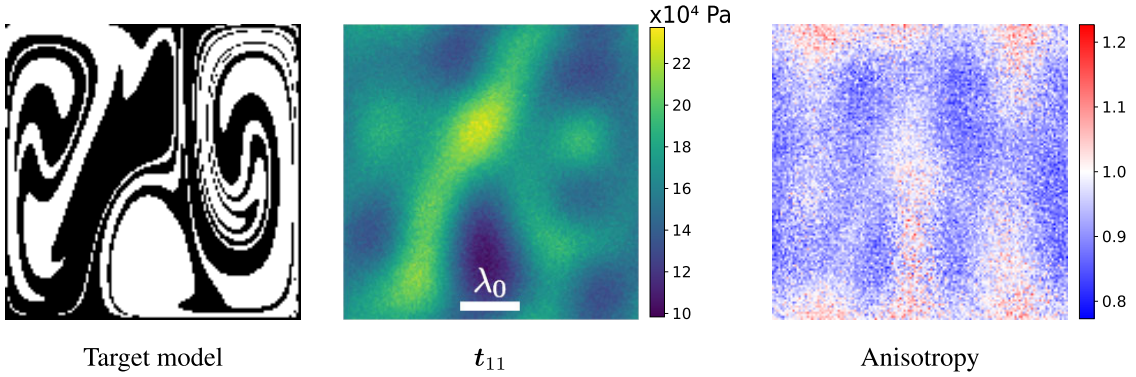

Figure 8. Left-hand panel: target model $\boldsymbol{m}_\mathrm{t}$ created with the geodynamic simulation described in Section 2.1. Middle panel: first coefficient of the observations $\boldsymbol{t} = \mathcal{H}(\boldsymbol{m}_\mathrm{t}) + \boldsymbol{\epsilon}$, obtained by homogenization of $\boldsymbol{m}_\mathrm{t}$ at $\lambda_0 = L/5$ (where L is the side length of the box) with Gaussian noise with SNR = 14 dB. Right-hand panel: effective anisotropy obtained after homogenization.

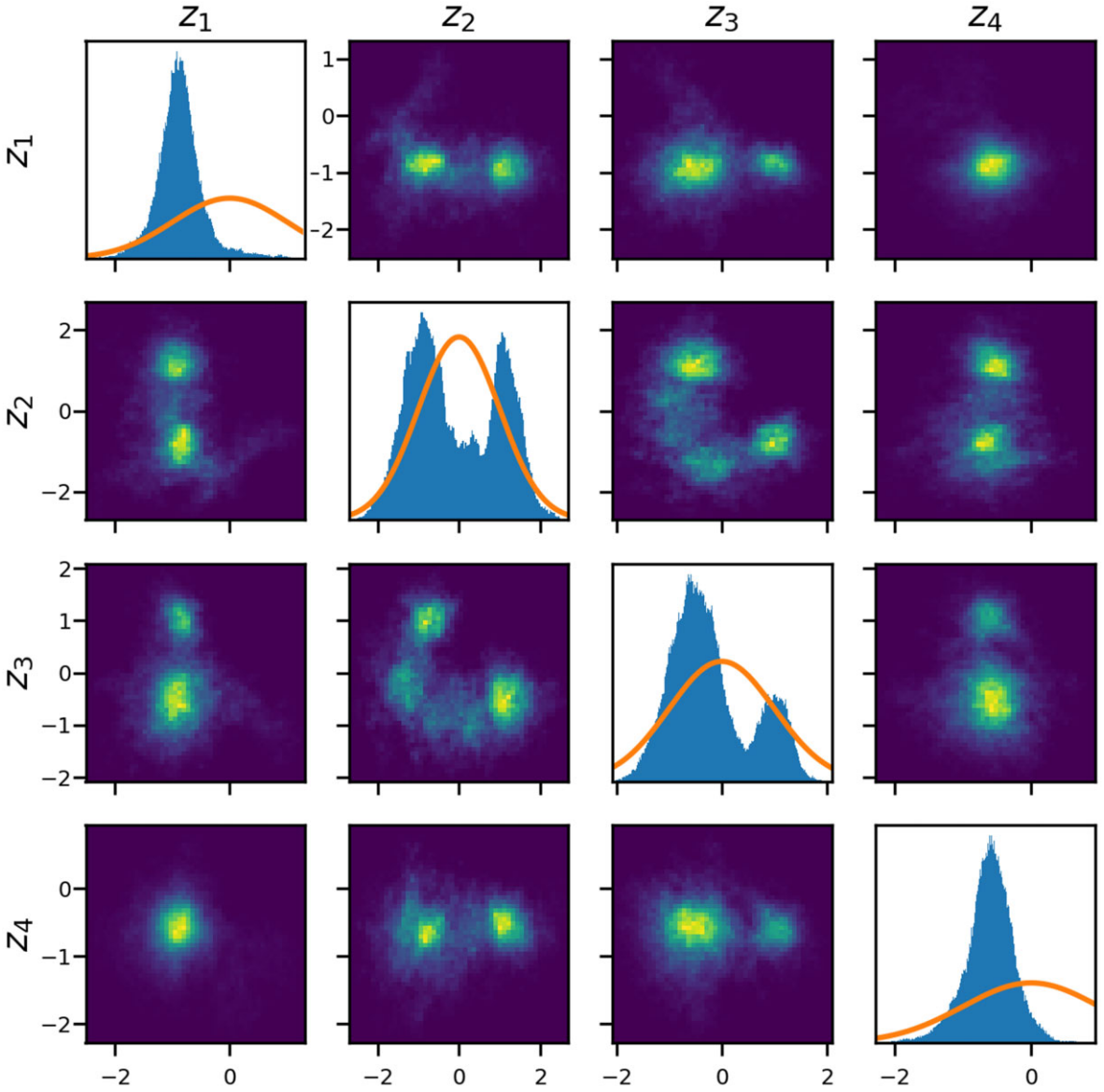

Figure 9. McMC sampling (2-D and 1-D marginals) of the posterior distribution $p(\boldsymbol{z} \mid \boldsymbol{t})$ for the 4 first dimensions. For more details, see Appendix D. The prior $p(\boldsymbol{z})$ (a standard normal distribution) is plotted in orange on the 1-D marginals for comparison. The posterior is tighter than the prior, thus showing the gain of information brought by the data. Two modes can be identified on the posterior distribution.

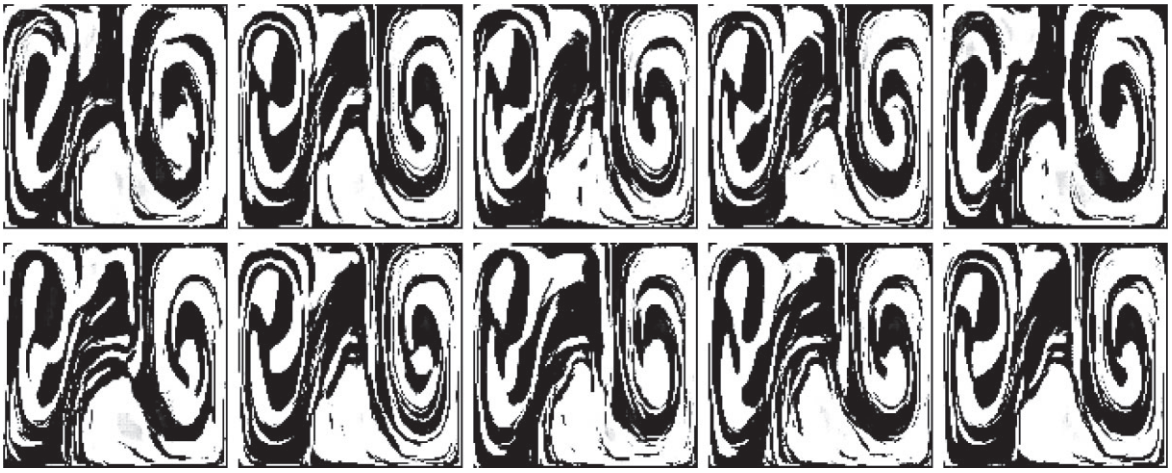

Figure 10. 10 images randomly drawn from the ensemble solution describing the posterior distribution.

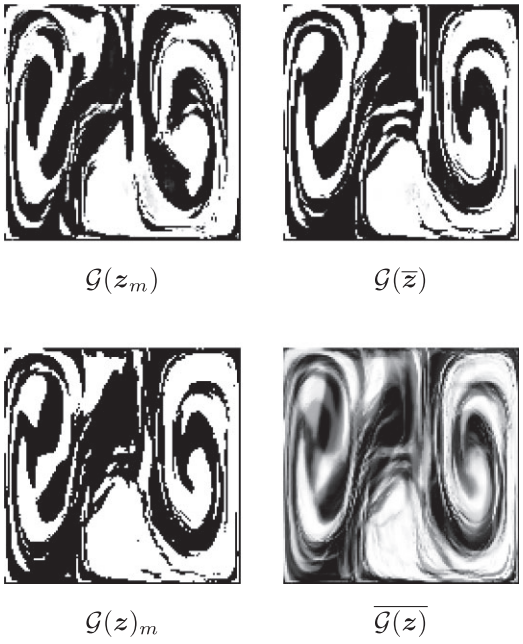

Figure 11. Characteristic images of the distribution. Mode (first column) and mean (second column), computed on the latent space (first row) and on the pixels space (second row).

Compared to the anisotropic case, the posterior mean image $\overline{\mathcal{G}(z)}$ in the isotropic scenario shows increased blurring, accompanied by a higher standard deviation, indicative of a less informative and broader posterior distribution. It is difficult to recognize the target image after inversion. Specifically, it fails to disclose specific localizations of layering, unlike in the anisotropic scenario where distinct regions exhibiting pronounced vertical layering can be identified and correspond to the targeted model layering. This emphasizes the crucial role of anisotropy (and the ability of homogenization to account for this extrinsic anisotropy) when constraining small-scale structures, thus providing valuable information to the downscaling procedure.

5.4 Methodology assessment

To verify the general applicability of the proposed methodology, we test it on a large number of different target models along with their corresponding simulated observations. For each model, we run the McMC sampling for one chain of 100 000 iterations. Fig. 15 presents the results for a randomly selected subset of 10 models. For each synthetic test, we show the target model, the observed data, the mean solution model and its associated estimated data. To verify that each tested target model $\boldsymbol{m}_t$ (a marble cake image consisting of 0 and 1) is significantly different from the training set, we also compute its distance to the training set as:

$$D_{\min}(\boldsymbol{m}_t) = \min_{\boldsymbol{m} \in \text{training set}} D(\boldsymbol{m}, \boldsymbol{m}_t)\,, \tag{17}$$

where $D(\boldsymbol{m}, \boldsymbol{m}_t)$ is the Mean Squared Error between $\boldsymbol{m}$ and $\boldsymbol{m}_t$, defined as:

$$D(\boldsymbol{m}, \boldsymbol{m}_\mathrm{t}) = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{m} - \boldsymbol{m}_\mathrm{t})_i^2\,, \tag{18}$$

where i runs through the number of pixels. In our context, pixel values are either 0 or 1. Therefore the distance D is simply the proportion of pixels that have different values. $D_{\min}$ is the minimal value of this distance between the target and all the training images. The lowest value for $D_{\min}$ is about 0.23 (7th value of the bottom row in Fig. 15), indicating that the target models are significantly different from any of the 19 456 training images. For all the tested models, we get results similar to those shown in the previous sections. The proposed methodology consistently demonstrates its effectiveness by successfully recovering the structural properties of the target models, thereby identifying probable small-scale features that are in accordance with the observed data.

6 DISCUSSION

The aim of this study is to present and assess our methodology, which involves encoding prior geological information into a generative neural network and subsequently integrating it into a Bayesian inversion framework with McMC. We evaluate this methodology through a synthetic inversion problem, the downscaling of a smooth tomographic image.

As shown in Section 5, the obtained results are promising. Despite the loss of small scales in the forward problem, due to the homogenization step, the prior information allows us to successfully reconstruct the structure of the target model, even capturing probable small-scale features. Additionally, the Bayesian approach provides access to a complete distribution of possible models, enabling the quantification of uncertainties in the solution.

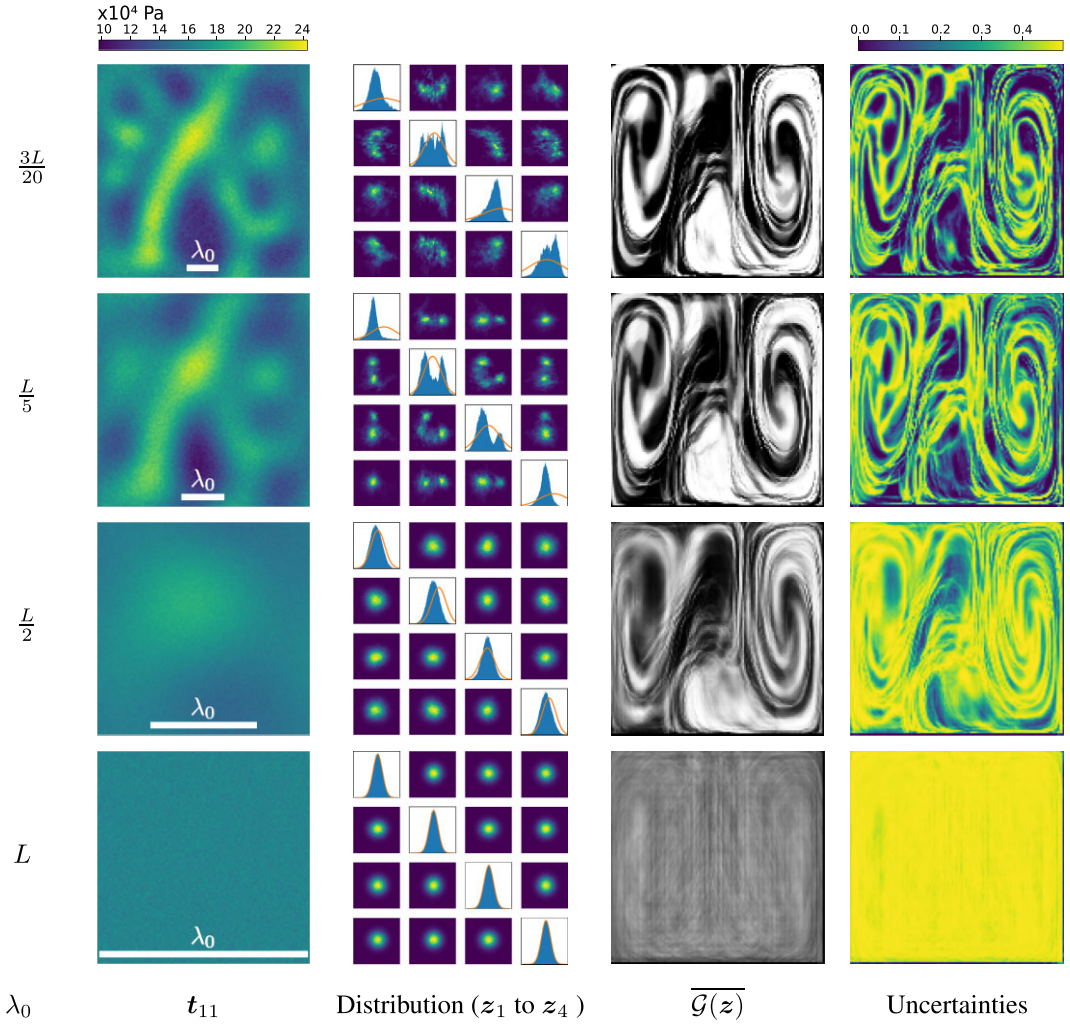

Figure 12. Results for different homogenization wavelengths λ_0. First column: Observations (first coefficient of the effective medium). They show different levels of details and thus different levels of difficulty for downscaling, as no structures lower than λ_0 are kept. Second column: sampled posterior distribution (first 4 latent dimensions, same setup as Fig. 9) with McMC. Smoother observations lead to wider, and closer to the prior, posterior distributions. Third column: mean model over the pixels space. Smoother observations lead to less accurate mean models. Forth column: Uncertainties on the image space, given by the standard deviation of the ensemble solution at each pixel.

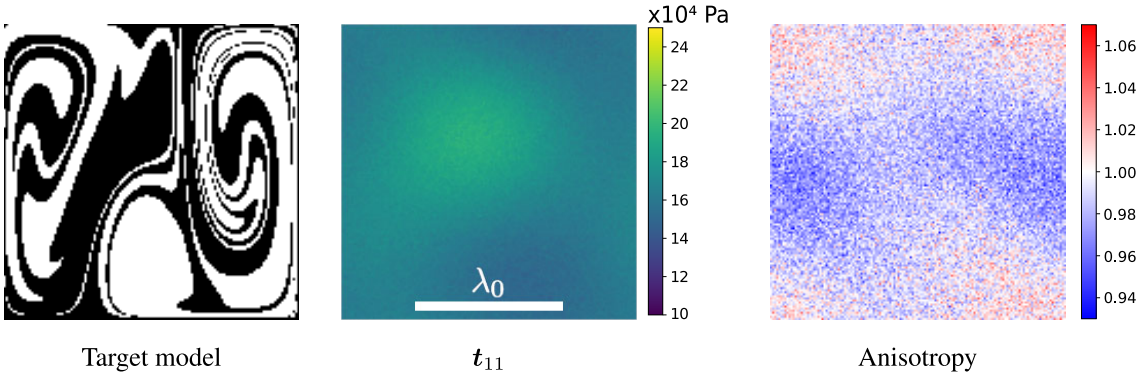

Figure 13. Target model (reminder), first component of the observations and resulting anisotropy $\xi = \boldsymbol{t}_{11}/\boldsymbol{t}_{22}$, for $\lambda_0 = L/2$. The homogenization leaves few information in the data, as λ_0 is equal to half the box length.

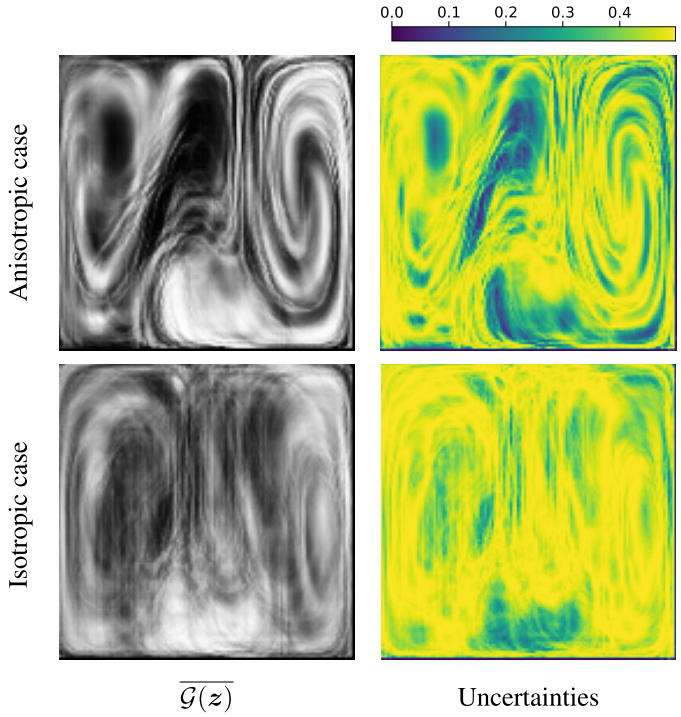

Figure 14. Comparison of results with and without anisotropy in the observed tomographic image. First column: mean image $\overline{\mathcal{G}(\boldsymbol{z})}$. Second column: uncertainties in the image space.

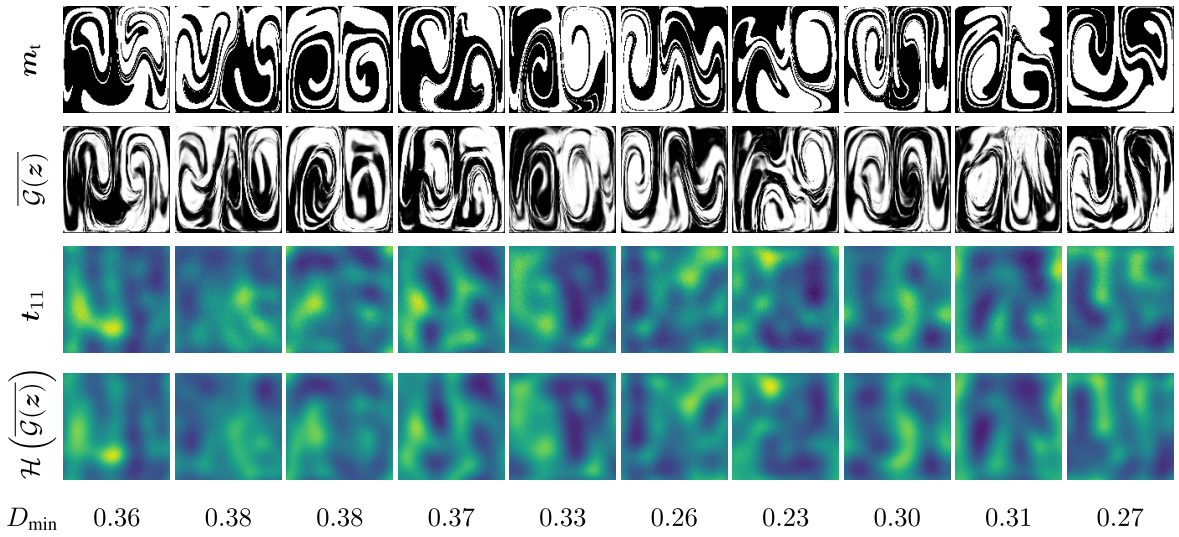

Figure 15. Inversions for different target models. First row: target models $\boldsymbol{m}_{\mathrm{t}}$. Second row: mean image $\overline{\mathcal{G}(\boldsymbol{z})}$ of the inversed sampled posterior distribution. Third row: observations $\boldsymbol{t}$. Fourth row: data estimated for the mean solution image. Fifth row: minimum distance between the target model and the training set images (quantifies the proximity between the target image and the training prior).

6.1 Advantage of sampling a highly informative prior

With very poorly informative observations (homogenization with $\lambda_0 = L/2$, see Fig. 12), we successfully capture the global structure of the target model. This underscores the advantages of using a generator able to sample a highly informative prior. By adopting this approach, the exploration of the model space becomes tightly constrained to a set of 'realistic' models, in a sense that it follows the multi-scale structures corresponding to the prior, which is learned with examples. Consequently, the proposed models not only explain the observations but also suggest structures similar to the educated guesses, thereby significantly narrowing down the range of plausible models. A common strategy for incorporating *a priori* information in inverse problems involves the application of a regularization or prior function to the proposed model. Formulating such a function, especially for an informative and complex prior, presents challenges. For instance, in our synthetic problem, quantifying the degree to which an image exhibits a marble cake shape is non-trivial. Our approach, which involves sampling the prior directly during the generation process, circumvents the need for defining this intricate function.

6.2 Advantage of the learning approach

In the toy model that we have used, the marble cake simulator described in Section 2.1 only needs a few input parameters. It may seem possible to invert directly for these parameters rather than using a neural network. However, this is impracticable for at least two reasons.

First, the marble cake simulator is several orders of magnitude slower ($\approx 10^3$ for our problem) than the GAN-based generation (typically executed in a few milliseconds, making homogenization the time-limiting factor). Only $\sim$ 20 000 images ($\sim$ 6 hr of simulation) are needed for the training set, and the training lasted approximately for 40 hr. This is much more efficient than the 19 d that would be needed for running 16 McMCs with 100 000 marble cakes. Furthermore, the training is done only one time : once trained, the GAN can be directly used for new inversions without another training.

Second, the chaotic behaviour of stirring makes the search of an optimum and the exploration impossible as two infinitely close sets of parameters lead to totally different marble cakes. In contrast, the GAN-generated marble cakes exhibit smooth variations with respect to the latent parameters.To illustrate this behaviour, we arbitrarily choose a point $\boldsymbol{z}_0$ in the latent space and a set of physical parameters $\boldsymbol{p}_0$ [these parameters are the amplitudes, frequencies and phases in eqs (4) and (5)]. We then define two perturbations $\boldsymbol{\delta z}$ and $\boldsymbol{\delta p}$ with amplitudes equal to the standard deviations of the prior distribution in the latent and physical space, respectively. Top panels in Fig. 16 show the evolution of resulting marble cakes when $\boldsymbol{z}$ and $\boldsymbol{p}$ vary in the ranges $[\boldsymbol{z}_0 - \boldsymbol{\delta z}; \boldsymbol{z}_0 + \boldsymbol{\delta z}]$ and $[\boldsymbol{p}_0 - \boldsymbol{\delta p}; \boldsymbol{p}_0 + \boldsymbol{\delta p}]$. Clearly, the marble cakes vary erratically in the physical space and continuously in the latent space. The bottom panel in Fig. 16 shows the behaviour of the distances of each image to the central image $D(\mathcal{G}(\boldsymbol{z}_0 + \alpha\boldsymbol{\delta z}), \mathcal{G}(\boldsymbol{z}_0))$ and $D(\mathcal{S}(\boldsymbol{p}_0 + \alpha\boldsymbol{\delta p}) - \mathcal{S}(\boldsymbol{p}_0))$ when α varies between -1 and 1. Analysis of these curves reveals a larger attraction basin for GAN-generated marble cakes, displaying greater continuity around the central model and avoiding multiple local minima.

With this property, the learning approach provides a smoother distribution to sample. As a result, fewer samples will be needed to represent the solution. Coupled with the accelerated generation of each sample, this renders the proposed method appealing.

6.3 Critical points for real data application

In the context of further work, our aim is to apply the methodology to real datasets spanning various observational domains like tomographic images of the mantle, seismic and gravimetric (Panet 2018; Scarponi *et al.* 2020; Panet *et al.* 2022) data, among others. In particular, the challenges we address here present significant similarities with those encountered in exploration geophysics (Kalla *et al.* 2009; Dupuy *et al.* 2016; Bayuk & Tikhotskiy 2018; Hedjazian *et al.* 2021). Furthermore, this versatile approach can be used with different, potentially more complex and realistic, prior models (e.g. models with temperature fields or rheological fields) with continuous variables or categorical ones, like in our study, and adapted to variations in scale and depth (global or regional models).

Transitioning from synthetic to real data introduces several challenges to be addressed. The application on real data necessitates a careful consideration of issues related to data heterogeneity, varied data shapes, the presence of noise in the training set, potential limitations arising from small datasets, and the precise definition of the likelihood function. One particular challenge will be to switch to 3-D models: geodynamic simulations can have different properties and structures between 2-D and 3-D models (e.g. Ferrachat & Ricard 1998). Further work includes adapting the methodology to 3-D models. Addressing these challenges will be crucial for ensuring the robustness and practical utility of the methodology in real-world inversions.

The key point to ensure the success of our approach lies in accurately defining the prior distribution. Unlike much less informative priors such as Tikhonov regularizations, our approach relies on a highly informative prior to effectively constrain the posterior distribution. Defining such a prior necessitates a profound understanding of the underlying geophysical model. For example, in our experiment, we model the underlying structures as marble cakes, akin to samples from the prior distribution. If the actual model deviates from this representation—that is if the prior does not align with the true model—any sample of the estimated posterior would still conform to the marble cake model. Hence, our method is doomed to produce models exclusively within the prior, making its choice extremely critical. In other words, even if the target image to be recovered were the Mona Lisa painting, the code would still be trying to reconstruct a marble cake! In that case, we expect a poor level of data fit from the sampled models.

Beyond these considerations of the domain spanned by the prior distribution, attention must be given to its shape within that domain. Through our GAN-based sampler, the prior is constructed from the distribution of images in the training set, assigning higher probabilities to models well-represented in the training data. Building an optimal GAN-based sampler requires a training set showcasing a substantial variability across all conceivable structures with a distribution that matches the expected distribution of the actual underlying structures.

In the future, given the versatility of this approach, it may help to understand the processes at play in the Earth's subsurface. Indeed, we believe it can provide a useful perspective in geophysical debates and contribute to, or reinforce, existing analyses combining geodynamics, seismology or other fields. This includes for example studies about dislocation versus diffusion creep in the lower mantle (McNamara *et al.* 2001), lower mantle dynamics (Lay *et al.* 1998; Garnero & McNamara 2008), deep slab hydration and interactions with the transition zone (Faccenda *et al.* 2009), partial melt in the upper mantle (Panet *et al.* 2024) or, as mentioned in the introduction, other analyses about slabs, plumes, cratons or LLSVP (e.g. Garnero *et al.* 2016; Davaille & Romanowicz 2020).

7 CONCLUSION

This paper presents a novel approach for downscaling tomographic images (i.e. recovering fine-scale geological structures that cannot be resolved by long period seismic waves). This is done by introducing a learned parametrization of geophysical models that accounts for sharp chemical discontinuities. In a Bayesian framework, the posterior solution significantly relies on the definition of the prior distribution. Given the impracticality of explicitly defining such a prior distribution for geophysical models, we propose a GAN-based sampler to implicitly generate geophysical models under this prior.

This method offers a solution to challenges that some inverse methods struggle to address. It provides a straightforward, computationally efficient, and low-dimensional parametrization for potentially complex models. Additionally, it offers a straightforward way to include *a priori* information about the models implicitly defined by a training set.

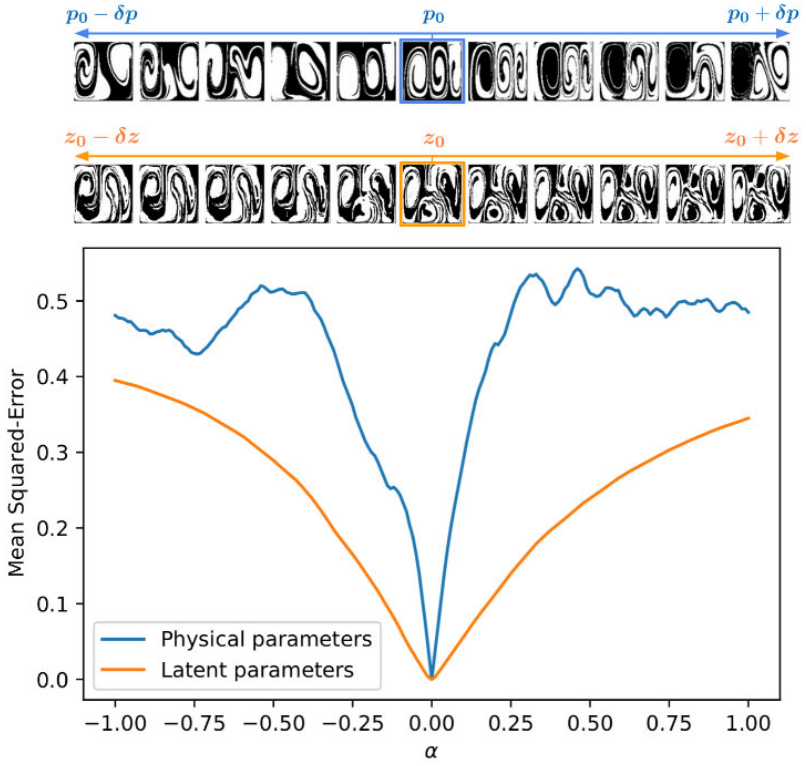

Figure 16. Top panel: marble cakes generated by varying parameters in the physical and latent space. Bottom panel: distances to the central marble cake $D(\mathcal{S}(\boldsymbol{p}_0), \mathcal{S}(\boldsymbol{p}_0 + \alpha\,\boldsymbol{\delta p}))$ and $D(\mathcal{G}(\boldsymbol{z}_0), \mathcal{G}(\boldsymbol{z}_0 + \alpha\,\boldsymbol{\delta z}))$, where α varies from -1 to 1; the components amplitudes of $\boldsymbol{\delta p}$ and $\boldsymbol{\delta z}$ are equal to the standard deviations of the corresponding prior distributions.

We illustrate the potential of such a method on synthetic data generated from binary images of marble cakes mimicking the oceanic lithosphere undergoing chaotic stirring in the mantle (Allègre & Turcotte 1986). We train a GAN to generate such structures, thus learning a reparametrization of these images with a small number (30 in the presented results) of latent variables. On the presented results, the structure of the target model is generally well recovered despite dealing with highly complex and fine-scale models, as well as poorly resolved observations.

We hope that this work will lead to similar approaches with more realistic *a priori* models, in order to apply it on real data. This will help to reconcile geodynamics and seismology and will result in more accurate and efficient interpretations of geological structures from seismic data, significantly enhancing our understanding of Earth subsurface.

AUTHOR CONTRIBUTIONS

T. Santos (Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing), T. Bodin (Conceptualization, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – review & editing), F. Soulez (Conceptualization, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – review & editing), Y. Ricard (Methodology, Resources, Writing – review & editing) and Y. Capdeville (Methodology, Resources, Writing – review & editing).

ACKNOWLEDGMENTS

The authors would like to thank the Editor Andrew Valentine, the Assistant Editor Louise Alexander, the anonymous reviewer and the reviewer Matthias Scheiter for the constructive comments and suggestions. The authors are grateful to the LABEX Lyon Institute of Origins (ANR-10-LABX-0066) Lyon for its financial support within the Plan France 2030 of the French government operated by the French National Research Agency (ANR).

DATA AVAILABILITY

The data underlying this paper will be shared upon reasonable request to the corresponding author.

REFERENCES

Alder, C., Bodin, T., Ricard, Y., Capdeville, Y., Debayle, E. & Montagner, J.-P., 2017. Quantifying seismic anisotropy induced by small-scale chemical heterogeneities, *Geophys. J. Int.,* **211**(3), 1585–1600.

Allègre, C.J. & Turcotte, D.L., 1986. Implications of a two-component marble-cake mantle, *Nature,* **323**(6084), 123–127.

Antonsen, T., jr & Ott, E., 1991. Multifractal power spectra of passive scalars convected by chaotic fluid flows, *Phys. Rev. A,* **44**(2), doi:10.1103/PhysRevA.44.851.

Arjovsky, M., Chintala, S. & Bottou, L., 2017. Wasserstein generative adversarial networks, in *International Conference on Machine Learning,* pp. 214–223, PMLR.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P. & Virieux, J., 2013. Regularized seismic full waveform inversion with prior model information, *Geophysics,* **78**(2), R25–R36.

Aulbach, S., 2012. Craton nucleation and formation of thick lithospheric roots, *Lithos,* **149,** 16–30.

Backus, G., 1962. Long-wave elastic anisotropy produced by horizontal layering, *J. geophys. Res.,* **67,** 4427–4440.

Ballmer, M.D., Schmerr, N.C., Nakagawa, T. & Ritsema, J., 2015. Compositional mantle layering revealed by slab stagnation at˜ 1000-km depth, *Sci. Adv.,* **1**(11), e1500815.

Barnier, G., Biondi, E. & Clapp, R., 2019. Waveform inversion by model reduction using spline interpolation, in *Proceedings of the SEG International Exposition and 89th Annual Meeting,* 1400–1404, SEG.

Batchelor, G., 1959. Small-scale variation of convected quantities like temperature in turbulent fluid. Part 1 – general discussion and the case of small conductivity, *J. Fluid Mech.,* **5,** 113–133.

Bayuk, I. & Tikhotskiy, S., 2018. Upscaling and downscaling of reservoir rock elastic properties-rock physics approach, in *Proceedings of the SEG International Exposition and Annual Meeting,* pp. 3653–3657, SEG.

Bijwaard, H. & Spakman, W., 1999. Tomographic evidence for a narrow whole mantle plume below Iceland, *Earth planet. Sci. Lett.,* **166**(3–4), 121–126.

Bissig, F., Khan, A. & Giardini, D., 2022. Evidence for basalt enrichment in the mantle transition zone from inversion of triplicated p-and s-waveforms, *Earth planet. Sci. Lett.,* **580,** doi:10.1016/j.epsl.2022.117387.

Bloem, H., Curtis, A. & Tetzlaff, D., 2023. Introducing conceptual geological information into Bayesian tomographic imaging, *Basin Res.,* **36**(1), doi:10.1111/bre.12811.

Bodin, T., Capdeville, Y., Romanowicz, B. & Montagner, J.-P., 2015. Interpreting radial anisotropy in global and regional tomographic models, *The Earth's Heterogeneous Mantle: A Geophysical, Geodynamical, and Geochemical Perspective,* pp. 105–144, Khan,, A. & Deschamps, F., Springer.

Bohra, P., Pham, T.-A., Dong, J. & Unser, M., 2022. Bayesian inversion for nonlinear imaging models using deep generative priors, *IEEE Trans. Comput. Imag.,* **8,** 1237–1249.

Bosch, M., Mukerji, T. & Gonzalez, E.F., 2010. Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: a review, *Geophysics,* **75**(5), 75A165–75A176.

Boyce, A., Bodin, T., Durand, S., Soergel, D. & Debayle, E., 2024. Seismic evidence for craton formation by underplating and development of the mld, *Geophys. Res. Lett.,* **51**(4), e2023GL106170.

Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.,* **185**(2), 845–870.

Brooks, S., Gelman, A., Jones, G. & Meng, X.-L., 2011. *Handbook of Markov Chain Monte Carlo,* CRC Press.

Burstedde, C. & Ghattas, O., 2009. Algorithmic strategies for full waveform inversion: 1D experiments, *Geophysics,* **74**(6), WCC37–WCC46.

Capdeville, Y. & Métivier, L., 2018. Elastic full waveform inversion based on the homogenization method: theoretical framework and 2-D numerical illustrations, *Geophys. J. Int.,* **213,** 1093–1112.

Capdeville, Y., Stutzmann, E., Wang, N. & Montagner, J., 2013. Residual homogenization for seismic forward and inverse problems in layered media, *Geophys. J. Int.,* **194,** 470–487.

Capdeville, Y., Zhao, M. & Cupillard, P., 2015. Fast fourier homogenization for elastic wave propagation in complex media, *Wave Motion,* **54,** 170–186.

Chan, S. & Elsheikh, A.H., 2019. Parametric generation of conditional geological realizations using generative neural networks, *Comput. Geosci.,* **23,** 925–952.

Charlety, J., Voronin, S., Nolet, G., Loris, I., Simons, F.J., Sigloch, K. & Daubechies, I.C., 2013. Global seismic tomography with sparsity constraints: comparison with smoothing and damping regularization, *J. geophys. Res.,* **118**(9), 4887–4899.

Chen, Y. & Saygin, E., 2021. Seismic inversion by hybrid machine learning, *J. geophys. Res.,* **126**(9), e2020JB021589.

Cordua, K.S., Hansen, T.M. & Mosegaard, K., 2012. Monte Carlo full-waveform inversion of crosshole gpr data using multiple-point geostatistical a priori information, *Geophysics,* **77**(2), H19–H31.

Cottaar, S. & Romanowicz, B., 2013. Observations of changing anisotropy across the southern margin of the african llsvp, *Geophys. J. Int.,* **195**(2), 1184–1195.

Curtis, A. & Lomax, A., 2001. Prior information, sampling distributions, and the curse of dimensionality, *Geophysics,* **66**(2), 372–378.

Dalton, C., Ekström, G. & Dziewoński, A., 2008. The global attenuation structure of the upper mantle, *J. geophys. Res.,* **113**(B9), doi:10.1029/2007JB005429.

Davaille, A. & Romanowicz, B., 2020. Deflating the llsvps: bundles of mantle thermochemical plumes rather than thick stagnant "piles", *Tectonics,* **39**(10), e2020TC006265.

Davies, D., Goes, S. & Lau, H., 2015. Thermally dominated deep mantle LLSVPS: a review, in *The Earth's Heterogeneous Mantle: A Geophysical, Geodynamical, and Geochemical Perspective,* pp. 441–477, Khan, A. & Deschamps, F., Springer.

Dupuy, B., Asnaashari, A., Brossier, R., Garambois, S., Métivier, L., Ribodetti, A. & Virieux, J., 2016. A downscaling strategy from fwi to microscale reservoir properties from high-resolution images, *Leading Edge,* **35**(2), 146–150.

Durand, S., Debayle, E., Ricard, Y., Zaroli, C. & Lambotte, S., 2017. Confirmation of a change in the global shear velocity pattern at around 1000 km depth, *Geophys. J Int.,* **211,** 1628–1639.

Esser, E., Guasch, L., van Leeuwen, T., Aravkin, A.Y. & Herrmann, F.J., 2018. Total variation regularization strategies in full-waveform inversion, *SIAM J. Imag. Sci.,* **11**(1), 376–406.

Faccenda, M., Gerya, T.V. & Burlini, L., 2009. Deep slab hydration induced by bending-related variations in tectonic pressure, *Nat. Geosci.,* **2**(11), 790–793.

Feng, R., Luthi, S.M., Gisolf, D. & Angerer, E., 2018. Reservoir lithology classification based on seismic inversion results by hidden markov models: applying prior geological information, *Mar. Petrol. Geol.,* **93,** 218–229.

Feng, R., Mosegaard, K., Grana, D., Mukerji, T. & Hansen, T.M., 2024. Stochastic facies inversion with prior sampling by conditional generative adversarial networks based on training image, *Math. Geosci.,* **56,** 665–690.

Ferrachat, S. & Ricard, Y., 1998. Regular vs. chaotic mantle mixing, *Earth planet. Sci. Lett.,* **155**(1-2), 75–86.

Fichtner, A. & Trampert, J., 2011. Resolution analysis in full waveform inversion, *Geophys. J. Int.,* **187**(3), 1604–1624.

Fichtner, A., Kennett, B.L. & Trampert, J., 2013. Separating intrinsic and apparent anisotropy, *Phys. Earth planet. Inter.,,* **219,** 11–20.

Foster, D., 2022. *Generative Deep Learning,* O'Reilly Media, Inc.

French, S.W. & Romanowicz, B., 2015. Broad plumes rooted at the base of the earth's mantle beneath major hotspots, *Nature,* **525**(7567), 95–99.

Fukao, Y. & Obayashi, M., 2013. Subducted slabs stagnant above, penetrating through, and trapped below the 660 km discontinuity, *J. geophys. Res.,* **118**(11), 5920–5938.

Garnero, E.J. & McNamara, A.K., 2008. Structure and dynamics of Earth's lower mantle, *Science,* **320**(5876), 626–628.

Garnero, E.J., McNamara, A.K. & Shim, S.-H., 2016. Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nat. Geosci.,* **9**(7), 481–489.

González, E.F., Mukerji, T. & Mavko, G., 2008. Seismic inversion combining rock physics and multiple-point geostatistics, *Geophysics,* **73**(1), R11–R21.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y., 2014. Generative adversarial nets, in *Proceedings of the Advances in Neural Information Processing Systems 27,* 8–13 December 2014, Montreal, Quebec, Canada..

Grand, S.P., Van der Hilst, R.D. & Widiyantoro, S., 1997. Global seismic tomography: a snapshot of convection in the Earth, *GSA Today,* **7**(4), 1–7.

Guitton, A., 2012. Blocky regularization schemes for full-waveform inversion, *Geophys. Prospect.,* **60**(5), 870–884.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A., 2017. Improved training of Wasserstein GANS, in *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems,* December 2017, Long Beach, California, USA, 5769–5779.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika,* **57**(1), 97–109.
Hedjazian, N., Capdeville, Y. & Bodin, T., 2021. Multiscale seismic imaging with inverse homogenization, *Geophys. J. Int.,* **226**(1), 676–691.
Hetényi, G. *et al.*, 2018. The AlpArray seismic network: a large-scale European experiment to image the Alpine Orogen, *Surv. Geophys.,* **39,** 1009–1033.
Jones, T.D., Maguire, R.R., van Keken, P.E., Ritsema, J. & Koelemeijer, P., 2020. Subducted oceanic crust as the origin of seismically slow lower-mantle structures, *Prog. Earth planet. Sci.,* **7,** 1–16.
Kalita, M., Kazei, V., Choi, Y. & Alkhalifah, T., 2019. Regularized full-waveform inversion with automated salt flooding, *Geophysics,* **84**(4), R569–R582.
Kalla, S., White, C.D., Gunning, J. & Glinsky, M.E., 2009. Downscaling multiple seismic inversion constraints to fine-scale flow models, *SPE J.,* **14**(04), 746–758.
Kárason, H. & Van Der Hilst, R.D., 2000. Constraints on mantle convection from seismic tomography, *Geophys. Monogr. Ser.,* **121,** 277–288.
Kingma, D. & Welling, M., 2013. Auto-encoding variational bayes, preprint (arXiv:1312.6114).
Koelemeijer, P., Deuss, A. & Trampert, J., 2012. Normal mode sensitivity to Earth's D layer and topography on the core-mantle boundary: what we can and cannot see, *Geophys. J. Int.,* **190**(1), 553–568.
Lau, H.C., Mitrovica, J.X., Davis, J.L., Tromp, J., Yang, H.-Y. & Al-Attar, D., 2017. Tidal tomography constrains Earth's deep-mantle buoyancy, *Nature,* **551**(7680), 321–326.
Lay, T., Williams, Q. & Garnero, E.J., 1998. The core–mantle boundary layer and deep earth dynamics, *Nature,* **392**(6675), 461–468.
Lebedev, S., Fullea, J., Xu, Y. & Bonadio, R., 2023. *Seismic Thermography,* AGU.
LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P., 1998. Gradient-based learning applied to document recognition, *Proc. IEEE,* **86,** 2278–2324.
Li, X., Aravkin, A.Y., van Leeuwen, T. & Herrmann, F.J., 2012. Fast randomized full-waveform inversion with compressive sensing, *Geophysics,* **77**(3), A13–A17.
Lin, Y. & Huang, L., 2014. Acoustic-and elastic-waveform inversion using a modified total-variation regularization scheme, *Geophys. J. Int.,* **200**(1), 489–502.
Liu, M., Grana, D. & de Figueiredo, L.P., 2022. Uncertainty quantification in stochastic inversion with dimensionality reduction using variational autoencoder, *Geophysics,* **87**(2), M43–M58.
Liu, Q. & Gu, Y., 2012. Seismic imaging: from classical to adjoint tomography, *Tectonophysics,* **566,** 31–66.
Lochbühler, T., Pirot, G., Straubhaar, J. & Linde, N., 2014. Conditioning of multiple-point statistics facies simulations to tomographic images, *Math. Geosci.,* **46,** 625–645.
Lochbühler, T., Vrugt, J.A., Sadegh, M. & Linde, N., 2015. Summary statistics from training images as prior information in probabilistic inversion, *Geophys. J. Int.,* **201**(1), 157–171.
Lopez-Alvis, J., Laloy, E., Nguyen, F. & Hermans, T., 2021. Deep generative models in inversion: the impact of the generator's nonlinearity and development of a new approach based on a variational autoencoder, *Comput. Geosci.,* **152,** doi:10.1016/j.cageo.2021.104762.
Magali, J.K., Bodin, T., Hedjazian, N., Samuel, H. & Atkins, S., 2021. Geodynamic tomography: constraining upper-mantle deformation patterns from Bayesian inversion of surface waves, *Geophys. J Int.,* **224**(3), 2077–2099.
Maguire, R., Ritsema, J., Bonnin, M., van Keken, P.E. & Goes, S., 2018. Evaluating the resolution of deep mantle plumes in teleseismic traveltime tomography, *J. geophys. Res.,* **123**(1), 384–400.
Mainprice, D., Barruol, G. & Ismaïl, W.B., 2000. The seismic anisotropy of the Earth's mantle: from single crystal to polycrystal, in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic Scale to the Global Scale,* Vol. **117,** pp. 237–264, eds Karato, S.-I., Forte, A., Liebermann, R., Masters, G. & Stixrude, L., AGU.
Mancinelli, N., Shearer, P. & Liu, Q., 2016. Constraints on the heterogeneity spectrum of Earth's upper mantle, *J. geophys. Res.,* **121,** 3703–3721.
Mariethoz, G. & Caers, J., 2014. *Multiple-Point Geostatistics: Stochastic Modeling With Training Images,* John Wiley & Sons.
McNamara, A.K., Karato, S.-I. & Van Keken, P.E., 2001. Localization of dislocation creep in the lower mantle: implications for the origin of seismic anisotropy, *Earth planet. Sci. Lett.,* **191**(1–2), 85–99.
Montelli, R., Nolet, G., Dahlen, F., Masters, G., Engdahl, E.R. & Hung, S.-H., 2004. Finite-frequency tomography reveals a variety of plumes in the mantle, *Science,* **303**(5656), 338–343.
Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100**(B7), 12431–12447.
Mosser, L., Dubrule, O. & Blunt, M.J., 2020. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior, *Math. Geosci.,* **52,** 53–79.
Muir, J.B. & Tsai, V.C., 2020. Geometric and level set tomography using ensemble Kalman inversion, *Geophys. J. Int.,* **220**(2), 967–980.
Munch, F.D., Khan, A., Tauzin, B., Zunino, A. & Giardini, D., 2018. Stochastic inversion of P-to-S converted waves for mantle composition and thermal structure: methodology and application, *J. geophys. Res.,* **123**(12), 10–706.
Olson, P., Yuen, D. & Balsiger, D., 1984. Mixing of passive heterogeneities by mantle convection, *J. geophys. Res.,* **89,** 425–436.
Panet, I., 2018. An analysis of gravitational gradients in rotated frames and their relation to oriented mass sources, *J. geophys. Res.,* **123**(12), 11–062.
Panet, I., Narteau, C., Lemoine, J.-M., Bonvalot, S. & Remy, D., 2022. Detecting preseismic signals in GRACE gravity solutions: application to the 2011 Tohoku Mw 9.0 earthquake, *J. geophys. Res.,* **127**(8), e2022JB024542, 10.1029/2022JB024542.
Panet, I., Greff-Lefftz, M. & Romanowicz, B., 2024. Partial melt in mesoscale upper mantle upwellings beneath ocean basins, *Earth planet. Sci. Lett.,* **639,** doi:10.1016/j.epsl.2024.118763.
Patel, D. & Oberai, A.A., 2019. Bayesian inference with generative adversarial network priors, preprint (arXiv:1907.09987).
Priestley, K., Ho, T. & McKenzie, D., 2021. The formation of continental roots, *Geology,* **49**(2), 190–194.
Rawlinson, N., Pozgay, S. & Fishwick, S., 2010. Seismic tomography: a window into deep earth, *Phys. Earth planet. Inter.,,* **178**(3–4), 101–135.
Ricard, Y., Nataf, H.-C. & Montagner, J.-P., 1996. The three-dimensional seismological model a priori constrained: confrontation with seismic data, *J. geophys. Res.,* **101**(B4), 8457–8472.
Richardson, A., 2018. Generative adversarial networks for model order reduction in seismic full-waveform inversion, preprint (arXiv:1806.00828).
Ritsema, J., Deuss, A., Van Heijst, H. & Woodhouse, J., 2011. S40rts: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophys. J. Int.,* **184,** 1223–1236.
Ritsema, J., Maguire, R., Cobden, L. & Goes, S., 2021. Seismic imaging of deep mantle plumes, in *Mantle Convection and Surface Expressions,* pp. 353–369, eds Marquardt, Hauke, Ballmer, Maxim, Cottaar, Sanne & Konter, Jasper, AGU.
Scarponi, M., Hetényi, G., Berthet, T., Baron, L., Manzotti, P., Petri, B., Pistone, M. & Müntener, O., 2020. New gravity data and 3-D density model constraints on the Ivrea Geophysical Body (Western Alps), *Geophys. J. Int.,* **222**(3), 1977–1991.
Schaeffer, A. & Lebedev, S., 2015. Global heterogeneity of the lithosphere and underlying mantle: a seismological appraisal based on multimode surface-wave dispersion analysis, shear-velocity tomography, and tectonic regionalization, in *The Earth's Heterogeneous Mantle: A Geophysical, Geodynamical, and Geochemical Perspective,* pp. 3–46, eds Khan, A. & Deschamps, F., Springer.
Scheiter, M., Valentine, A. & Sambridge, M., 2022. Upscaling and downscaling Monte Carlo ensembles with generative models, *Geophys. J. Int.,* **230**(2), 916–931.
Schmandt, B., Jiang, C. & Farrell, J., 2019. Seismic perspectives from the western us on magma reservoirs underlying large Silicic Calderas, *J. Volc. Geotherm. Res.,* **384,** 158–178.
Simons, F.J., Loris, I., Brevdo, E. & Daubechies, I.C., 2011. Wavelets and wavelet-like transforms on the sphere and their application to geophysical

data inversion, in *Wavelets and Sparsity XIV,* Vol. **8138,** pp. 224–238, SPIE.

Stixrude, L. & Jeanloz, R., 2007. 1.22 - Constraints on seismic models from other disciplines – constraints from mineral physics on seismological models, in *Treatise on Geophysics,* pp. 775–803, ed. Schubert, G., Elsevier.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* SIAM.

Thorne, M.S., Garnero, E.J., Jahnke, G., Igel, H. & McNamara, A.K., 2013. Mega ultra low velocity zone and mantle flow, *Earth planet. Sci. Lett.,* **364,** 59–67.

Tromp, J., 2020. Seismic wavefield imaging of Earth's interior across scales, *Nat. Rev. Earth Environ.,* **1**(1), 40–53.

Tsai, V.C., 2023. The future of Earth imaging, *Seismol. Res. Lett.,* **94**(5), 2119–2128.

Tsai, V.C., Huber, C. & Dalton, C.A., 2023. Towards the geological parametrization of seismic tomography, *Geophys. J. Int.,* **234**(2), 1447–1462.

Valentine, A.P. & Trampert, J., 2012. Data space reduction, quality assessment and searching of seismograms: autoencoder networks for waveform data, *Geophys. J. Int.,* **189**(2), 1183–1202.

Van der Hilst, R.D., Widiyantoro, S. & Engdahl, E., 1997. Evidence for deep mantle circulation from global tomography, *Nature,* **386**(6625), 578–584.

Wenk, H.-R., Cottaar, S., Tomé, C.N., McNamara, A. & Romanowicz, B., 2011. Deformation in the lowermost mantle: from polycrystal plasticity to seismic anisotropy, *Earth planet. Sci. Lett.,* **306**(1–2), 33–45.

Wolf, J., Li, M., Long, M.D. & Garnero, E., 2024. Advances in mapping lowermost mantle convective flow with seismic anisotropy observations, *Rev. Geophys.,* **62**(2), e2023RG000833, 10.1029/2023RG000833.

Xu, W., Lithgow-Bertelloni, C., Stixrude, L. & Ritsema, J., 2008. The effect of bulk composition and temperature on mantle seismic structure, *Earth planet. Sci. Lett.,* **275**(1–2), 70–79.

APPENDIX A: WASSERSTEIN GENERATIVE ADVERSARIAL NETWORK (WGAN)

To train a GAN, the generator $\mathcal{G}$ and discriminator $\mathcal{D}$ are trained alternatively. Both trainings have to progress at the same speed, otherwise the slower-trained network will fail its training. Indeed, in this case, it would have reached an area where the gradient is too low because the dominating network won't provide useful feedback (either the discriminator distinguishes too easily between real and fake, or the generator produces samples too hard to distinguish), preventing any training by gradient descent. This constraint can harden the training, or even make it impossible. One other problem is the so-called 'mode collapse' (Goodfellow *et al.* 2014): the generator ends up generating a small set of outputs, lacking of diversity. The mapping between the latent space and the pixels space is redundant and only done on a very restricted subset of the distribution $p(\boldsymbol{m})$.

One way to avoid those problems is to use a particular GAN: the Wassertein-GAN (WGAN). Arjovsky *et al.* (2017) proposes this type of GAN, by using the Wasserstein distance measure into the loss equation. The loss equation is

$$\mathcal{L}_{minimax} = \min_{\mathcal{G}} \max_{\mathcal{D}} \; \mathbb{E}_{\boldsymbol{m}\sim p(\boldsymbol{m})}(\mathcal{D}(\boldsymbol{m})) - \mathbb{E}_{\boldsymbol{z}\sim p(\boldsymbol{z})}(\mathcal{D}\circ\mathcal{G}(\boldsymbol{z}))\,, \quad \text{(A1)}$$

that is the generator and the discriminator alternatively learn to maximize, respectively, $\mathbb{E}_{\boldsymbol{z}\sim p(\boldsymbol{z})}(\mathcal{D}\circ\mathcal{G}(\boldsymbol{z}))$ and $\mathbb{E}_{\boldsymbol{m}\sim p(\boldsymbol{m})}(\mathcal{D}(\boldsymbol{m})) - \mathbb{E}_{\boldsymbol{z}\sim p(\boldsymbol{z})}(\mathcal{D}\circ\mathcal{G}(\boldsymbol{z}))$. This loss function is proved to improve training stability and reduce mode-collapse (Arjovsky *et al.* 2017). In addition, we use a gradient penalty, that is we add a term on the loss that penalizes the gradient norm of the discriminator output with respect to its input images. This smoothly enforces the discriminator to be a 1-Lipschitz function, thus enhancing even more the stability and performance of the WGAN (Gulrajani *et al.* 2017).

APPENDIX B: IMPLEMENTATION DETAILS

To construct the generator $\mathcal{G}$, we train a WGAN with gradient penalty (using coefficient $\lambda_{GP} = 10$). The network architecture is based on convolutional layers, inspired by LeCun *et al.* (1998), and is slightly adapted from the WGAN implementation available at github.com/Zeleni9/pytorch-wgan.git. We use a latent space dimension of $Z = 30$, as a compromise between a low dimension, for easier McMC sampling, and a high dimension for better representation of the training set. The generator produces output images with dimensions of 128×128. The architecture details (layers, output sizes, stride values and padding) used for the discriminator and the generator are given in Tables A1 and A2, respectively. The GAN training process spanned approximately 130 epochs, with each epoch consisting of 304 batches of 64 images, for a total training set of 19 456 samples. During training, the generator is updated once per step, while the discriminator is updated five times, with each update using a different batch of data. The trainings of $\mathcal{D}$ and $\mathcal{G}$ are done with Adam optimizer, with a learning rate value of $\lambda_{LR} = 10^{-4}$, $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The negative slope of the LeakyReLU in $\mathcal{D}$ is 0.2. This training regime allows the GAN to learn and refine its ability to generate realistic marble cake images.

APPENDIX C: McMC CONVERGENCE ASSESSMENT

Although the McMC algorithm has a theoretical convergence, it is only asymptotic and not guaranteed for a finite number of samples. It is then necessary to assess the convergence. In our work, we verified each time the convergence by looking at the posterior value of sampled models along each chain after the burn-in period (the initial phase of the sampling where early samples are discarded to allow the chain to converge to its stationary distribution). For each chain, the curve oscillate around a stable value with no obvious trend. We show some of these curves in Fig. A1. Another verification can be done by looking at the mean latent vector in the beginning of the sampling (first 1/3) compared to the ending (last 1/3). As Fig. A2 shows, both mean vectors are similar. Finally, we also plotted, for each chain and for each latent dimension, the autocorrelation curve. Some of them, along with the mean of all of them, are shown in Fig. A3. The autocorrelation curves are tight enough to show that most of the samples are not correlated between themselves.

From these three verifications (posterior density evolutions, comparison of the mean vector at the beginning versus ending of the sampling, and autocorrelation curves), we are confident that the McMC converged. The figures of this appendix were computed for the principal experiment of Section 5.1.

APPENDIX D: COMPLEMENT ON THE POSTERIOR DISTRIBUTION

In Fig. 9, the posterior distribution is only represented on the first four dimensions. The other 2-D marginals present similar bimodal structures. Although showing the complete plot is not possible for the sake of legibility, Fig. A5 presents all the 1-D marginals and Fig. A4 presents the covariance matrix of the 30 sampled latent variables.

Table A1. Structure of the discriminator with an input space of $128 \times 128 \times 1$.

Discriminator $\mathcal{D}$			
Layers	Output size	Stride	Padding
Conv 4×4 + LeakyReLU	$64 \times 64 \times 32$	2	1
Instance normalization	$64 \times 64 \times 32$		
Conv 4×4 + LeakyReLU	$32 \times 32 \times 64$	2	1
Instance normalization	$32 \times 32 \times 64$		
Conv 4×4 + LeakyReLU	$16 \times 16 \times 128$	2	1
Instance normalization	$16 \times 16 \times 128$		
Conv 4×4 + LeakyReLU	$4 \times 4 \times 512$	2	1
Instance normalization	$4 \times 4 \times 512$		
Conv 4×4	$1 \times 1 \times 1$	1	0

Table A2. Structure of the generator with an input space of $1 \times 1 \times 30$.

Generator $\mathcal{G}$			
Layers	Output Size	Stride	Padding
ConvTranspose 4×4 + ReLU	$4 \times 4 \times 2048$	1	0
Batch normalization	$4 \times 4 \times 2048$		
ConvTranspose 4×4 + ReLU	$8 \times 8 \times 1024$	2	1
Batch normalization	$8 \times 8 \times 1024$		
ConvTranspose 4×4 + ReLU	$16 \times 16 \times 512$	2	1
Batch normalization	$16 \times 16 \times 512$		
ConvTranspose 4×4 + ReLU	$32 \times 32 \times 256$	2	1
Batch normalization	$32 \times 32 \times 256$		
ConvTranspose 4×4 + ReLU	$64 \times 64 \times 128$	2	1
Batch normalization	$64 \times 64 \times 128$		
ConvTranspose 4×4 + tanh	$128 \times 128 \times 1$	2	1

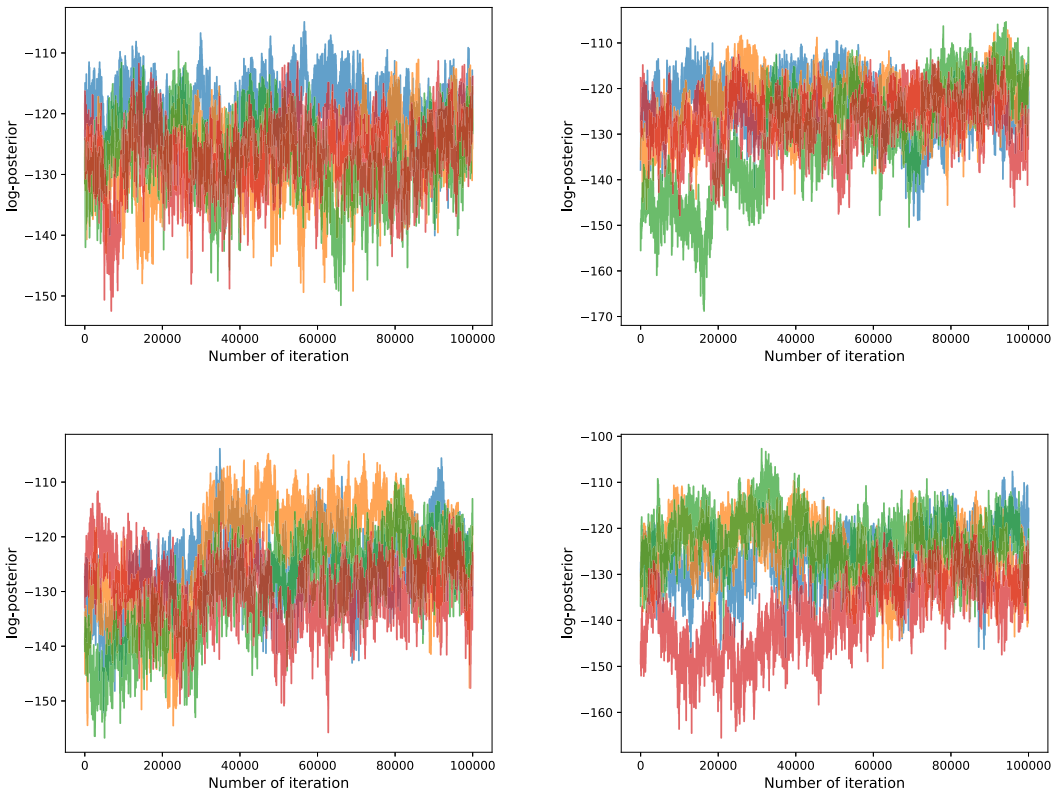

Figure A1. Evolution of the log-posterior of each chains in the principal experiment, after the burn-in period. The 16 chains are represented by groups of 4 for better legibility. The curves oscillates around the same values.

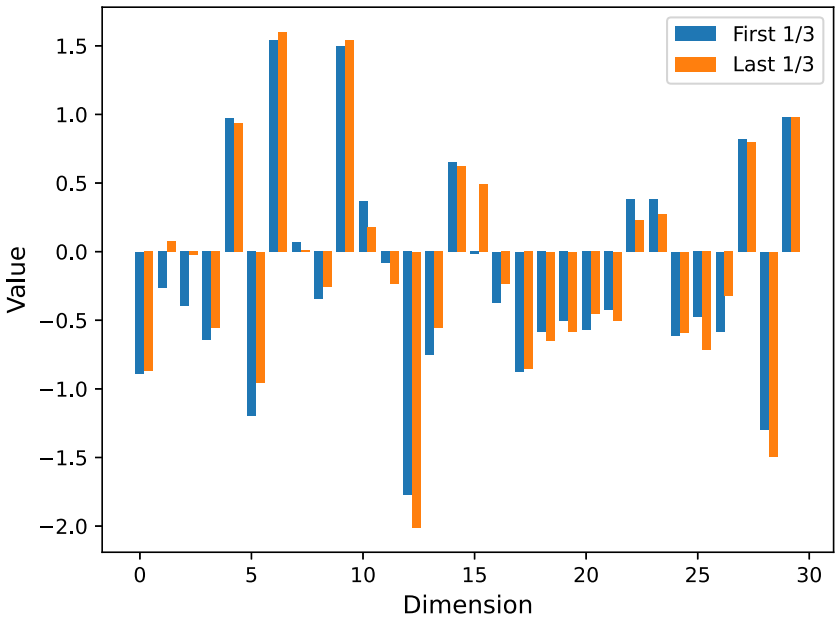

Figure A2. Comparison between the means (on all chains) of the latent vectors on each dimension, for the first 1/3 (blue) and last 1/3 (orange) samples, for the principal experiment.

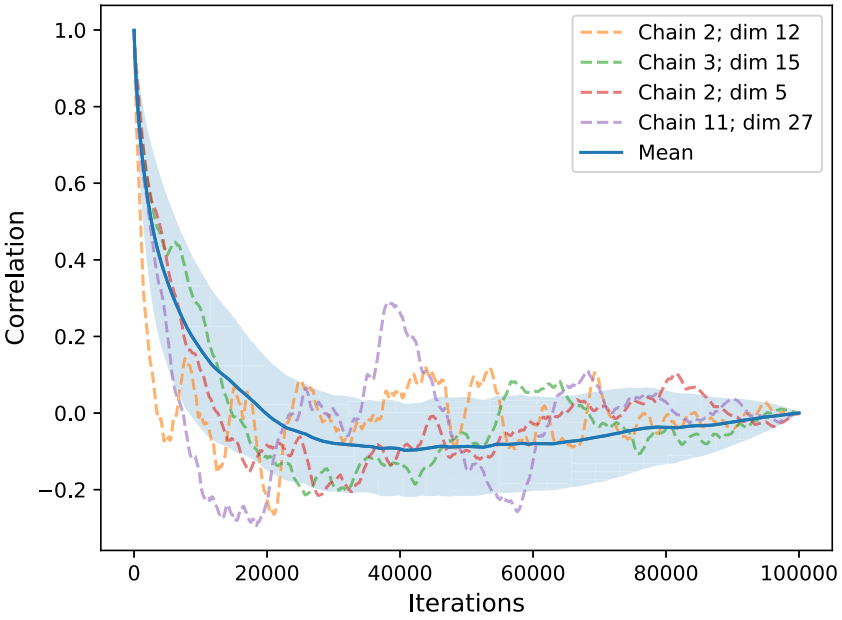

Figure A3. Autocorrelations curves of McMC sampling for the principal experiment. We represent the curves for 4 randomly chosen pairs of (latent dimension; chain), alongside with the mean curve on all dimensions and chains.

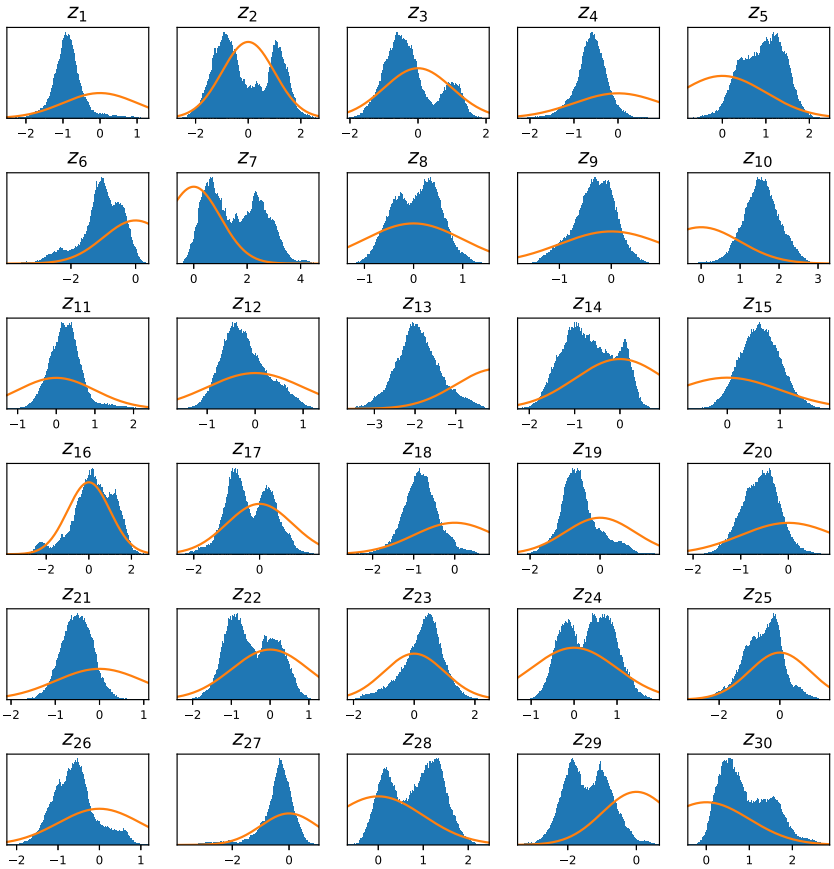

Figure A4. 1-D marginals of the posterior distribution sampled in the principal experiment shown in In Fig. 9, for all 30 dimensions. The prior, a standard normal distribution, is plotted in orange for comparison.

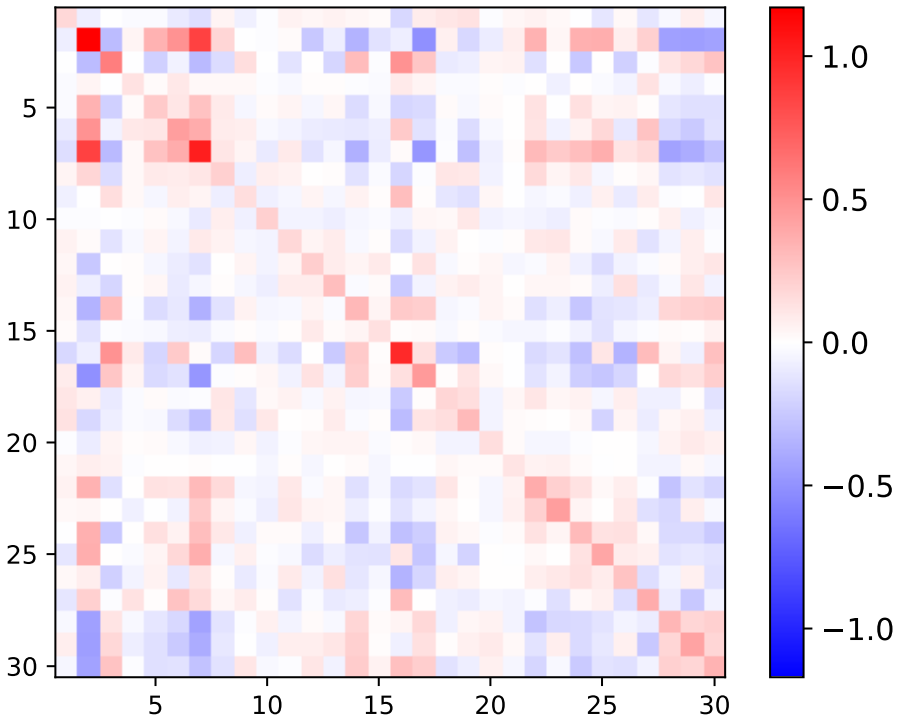

Figure A5. Covariance matrix of all the vectors sampled in the principal experiment (shown in Fig. 9).