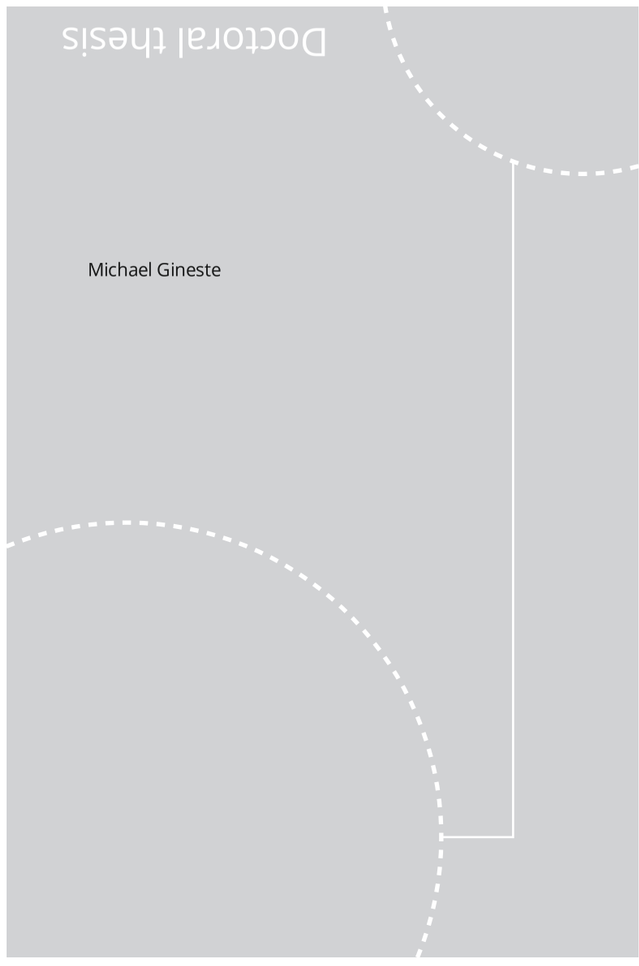
Doctoral theses at NTNU, 2020:323

Michael Gineste

Seismic waveform inversion using the iterative ensemble Kalman smoother

Doctoral thesis



Michael Gineste

ISBN 978-82-326-4994-5 (printed ver.)
ISBN 978-82-326-4995-2 (electronic ver.)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical
Engineering
Department of Mathematical Sciences

Doctoral theses at NTNU, 2020:323

NTNU
Norwegian University of
Science and Technology

Michael Gineste

Seismic waveform inversion using the iterative ensemble Kalman smoother

Thesis for the Degree of Philosophiae Doctor

Trondheim, October 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

ISBN 978-82-326-4994-5 (printed ver.)
ISBN 978-82-326-4995-2 (electronic ver.)
ISSN 1503-8181

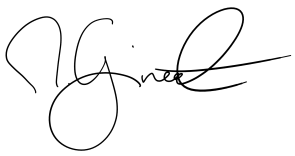
Doctoral theses at NTNU, 2020:323

Printed by NTNU Grafisk senter

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor (PhD) at the Norwegian University of Science and Technology (NTNU). The research is funded by the Uncertainty in Reservoir Evaluation (URE) consortium, and carried out at the Department of Mathematical Sciences (IMF). Supervisor on the project was Professor Jo Eidsvik, IMF and co-supervisor Børge Arntsen, Department of Geoscience and Petroleum, NTNU.

I would like to thank Jo Eidsvik for his excellent supervision, encouragement and for always having the time to discuss my work. Also, I would like to thank York Zheng, BP, for his involvement in my project, for the discussions and collaboration. I am grateful to the staff at Centre de Géosciences, MINES Paristech, Fontainebleau, for welcoming me during my research stay. Finally, I would like to thank Professor Henning Omre for always sharing his thoughts and good mood.



August 2020
Trondheim

Michael Gineste

Outline of thesis

Background

1 Introduction

The objective of this thesis is inference of subsurface rock properties from seismic reflection data. This constitutes an inverse problem where the unknown medium, the properties of the subsurface, are to be inferred from the observed seismic response of the system.

The motivation for this work is to provide a notion of uncertainty related to estimation of the subsurface properties from seismic data. Seismic data continues to be of uttermost importance for reliable petroleum reservoir characterization, being used in various stages of reservoir exploration and development. It serves as data for subsurface imaging and interpretation of geological features. The additional information of uncertainty can be utilized in the geological interpretation or in a decision-making process. Therefore the inverse problem is here considered in a Bayesian setting where the solution is a probability distribution and uncertainty an inherent part of the solution.

The field of full waveform inversion (FWI) was and is an increasing applied methodology in reservoir characterization, and the practice could benefit from having some uncertainty quantification associated with it. This was the initial research question of this thesis work; how to get estimation uncertainty into the full waveform application.

Within the geosciences, data assimilation describes state estimation using the Bayesian paradigm (Wikle et al. 2007; Carrassi et al. 2018). Data assimilation has received considerable attention in various communities and its use is becoming widespread. The aim of data assimilation is to integrate sources of information and to combine these in an optimal manner. Such sources of information are the mathematical/numerical model of a phenomena and the actual observations of this phenomena.

Envisioning full waveform application in 2- or 3D where the parameter dimension becomes very large, the method of choice must scale well. A data assimilation method that have shown robust in use for high-dimensional problems is the ensemble Kalman filter (EnKF). The decision to focus on the application of EnKF methodology was influenced by its successful application to the history matching (HM) problem. The HM problem shares features with the FWI problem, but also deviates fundamentally in the underlying physical system, as HM concerns a flow problem whereas FWI is

a wave problem.

The main ensemble methods used within HM applications are now iterative of nature. The nonlinear nature of the problem and the cost of numerical prediction has favored the iterative approach. Primary methods are the ensemble smoother-multiple data assimilation (ES-MDA, Emerick and Reynolds 2013) and the iterative ensemble smoother (IES, Chen et al. 2011), the method formerly known as ensemble randomized maximum likelihood (EnRML). Evensen (2018) provided a detailed comparison between the two. A third option might be the iterative ensemble Kalman smoother (IEnKS) used in this thesis.

Associating uncertainty to FWI is an ongoing effort. Thurin et al. (2019) used the ensemble Kalman framework and a frequency solver to get uncertainty estimation in a 2D, acoustic application. Others are seeking alternative paths to assess the uncertainty; Eikrem et al. (2019) used an iterative extended Kalman filter to seismic time-lapse data, whereas Zhang et al. (2020) considered a variational inference approach to assess the estimation uncertainty, the so-called Variational FWI, for a 2D case of elastic inversion. So quantifying uncertainty in large-scale seismic inversion is a hot topic.

Sampling based inversion requires repeated evaluations of a model for the observed phenomena and is inherently computationally demanding. This implies and demands parallel computations, and the work in this thesis has relied heavily on access to resources provided by the NTNU IDUN/-EPIC cluster facility (Själander et al. 2019).
The computational cost/time of evaluating the wave model is the primary reason for not considering an Markov-chain Monte Carlo (McMC) approach to the Bayesian inverse problem. The McMC approach would principally result in a correct Bayesian solution, but for high-dimensional problems the required time to evaluate (in non-parallel) the chain makes it unfeasible for this kind of seismic application. Another correct approach to the Bayesian inverse problem would be the application of a particle filter, but this approach has its issues with filter degeneracy/collapse for large-dimensional systems.

Finally, a note on terminology. The distinction between a filter and a smoother is more distinct when the state is dynamically evolving, and the state and observations shares a common time reference. For a static parameter state, this distinction is less clear. In the ensemble-based framework, the convention is that a smoother refers to assimilation of observa-

tions that spans several observation times. It is also known as 4D data assimilation (Hunt, Kalnay, et al. 2004) or asynchroneous data assimilation (Sakov, Evensen, et al. 2010). In this thesis, an assimilation step always covers several observation times and therefore always considers a smoother.

The remaining parts of this background will be brief walk-though, from the Kalman filter to an iterative ensemble smoother. This is to build up an introduction to the concepts and components used by the IEnKS. Section 2 will introduce the seismic forward problem, the general inverse problem, and the inverse problem from a Bayesian perspective is introduced in Section 3. Section 4 discusses the recursive estimation process of sequential filtering, and in Section 5 the Kalman filter and its extensions for nonlinear observations are presented. This leads to Section 6 that introduces the ensemble approximation to Kalman filters and square root filters, and finally the iterative ensemble Kalman smoother which is presented in Section 7. Section 8 discusses some aspects of elastic parameter estimation with ensemble-based methods. Some possible future work is discussed in Section 9, after which follows short summaries of presentations held and contributed papers in Section 10.

2 Seismic waveform inversion

Tomography refers to the technique of obtaining an image of a body or object's internal structure and doing so in a noninvasive manner. In general terms, tomography is done by passing some signal through the object, where the signal is modified in the interaction with medium changes (the internal structure), and observing the signal after the passage. Many examples of practical tomography uses a wave as signal that propagates through the medium. Examples of medical tomography are CT and ultrasound scanning. The former technique uses X-rays to pass through the body with observations made on the other side of the body. The latter uses high-frequency sound waves and reflections back towards source emitter. Seismic tomography deals with imaging properties of the Earth's subsurface, which is most often done with observations at the surface, thus in essence resembles ultrasound tomography.

The field of seismic tomography encompasses many techniques among which FWI is one. If the range of these techniques were ranked in terms of complexity and imaging resolution (generally complementary), FWI would be in the high-end of the spectrum.

The motivation for full waveform inversion is to use the best possible/-

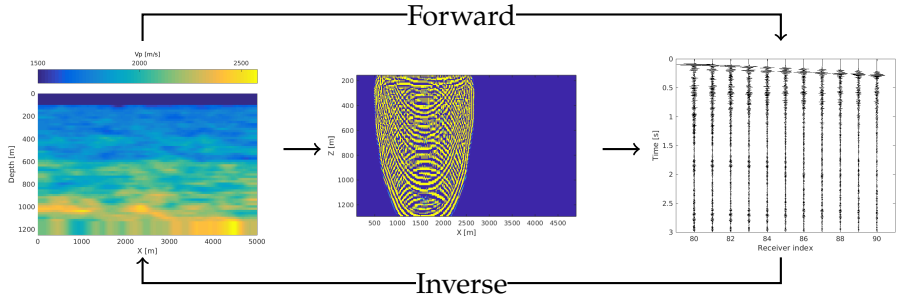


Figure 1: Depicting the forward and inverse relationship for seismic waveform. Left; acoustic velocity parameter throughout the domain. Middle; a source located at top of domain excites a propagating wavefield. Right; observed wavefield at receivers located at top of domain. Only a small fraction of the total wavefield is observed.

available physical model representation of the physics, and to use the observations with a minimum of processing as this tends to distort/alter the information content in measurements. Full waveform inversion operates directly on shot gather data and uses the residual between simulated and measured data, to iteratively update the subsurface model. The simulation setup used for FWI, is supposed to reflect the actual acquisition situation. Along with using a physical model that includes most of wave propagation phenomena, the expectation is that a simulated observation is quite close to the actual measurement. This was the vision in the first examples of what might be considered as FWI, as presented in Tarantola (1984). The field has evolved a lot since then, for introductory texts on FWI, see e.g. Virieux and Operto (2009), Fichtner (2010), and Virieux, Asnaashari, et al. (2017).

The prediction of an expected observation is dubbed the forward problem; given these causes, what is the effect? The inverse problem flips the question; given these effects, what are the causes? These arrows of causality are depicted in Figure 1.

Seismic waveform

Seismic data is in its raw form a recording of a wavefield, excited by a controlled source in the considered context. The recording might be pressure fluctuations in water, using hydrophones or it might be (particle) displacement using geophones. In either way it is a very complex signal resulting

from the propagating wave field.

When the source initiates its forcing, a wave front starts propagating from the source. Whenever this front meets a material inhomogeneity, the wave is scattered and the resulting wavefield is even more complex. The end result of all this interaction, observed near the surface relatively far from the medium area of interest, is what we have of measurement data and from which the medium properties must be inferred. That is to say, this inference is not a straight forward task.

Many processing techniques exist to reduce the amount of information in these recordings and keep only the essential information, tailored to a specific processing/analysis technique (Sheriff et al. 1995).

Full waveform often refers to the synthetic seismograms being simulated as realistically as possible in terms of physics. This involves numerical simulation of the wave equation in either time- or frequency domain. Either domain has its advantages, but here the focus is on time-domain.

The seismic wave propagation is described by the elastic wave equation (Fichtner 2010), in a displacement-stress formulation

$$\rho(\mathbf{s})\partial_t^2\mathbf{u}(\mathbf{s},t) - \nabla\cdot\boldsymbol{\sigma}(\mathbf{s},t) = \mathbf{f}(\mathbf{s},t)\,, \tag{1}$$

where $\mathbf{u}(\mathbf{s},t)$ is the particle displacement, $\boldsymbol{\sigma}(\mathbf{s},t)$ the stress tensor and $\rho(\mathbf{s})$ the density over the spatial domain $\mathbf{s}\in\mathcal{D}\subset\mathbb{R}^3$ and time $t\in[0,T]$. The wave field is excited by the force $\mathbf{f}(\mathbf{s},t)$ and the system is considered at rest for $t<0$. The latter is the initial condition and the system must also be supplied with appropriate boundary conditions. In the form (1), the subsurface properties of interest enters via the stress tensor through e.g. spatial fields of shear and bulk moduli, and the density field. The wave equation formulation can take other forms, but the details are irrelevant here.

The dynamic wavefield is to be integrated over the spatial domain of interest. The vertical extent, the subsurface depth, can be several kilometers and the lateral extent, supposed to reflect the acquisition situation with long receiver arrays, can extend many kilometers laterally. Hence the spatial domain is of considerable size, which must be discretized into a computational grid of fairly fine resolution in order to model the targeted spatial details and resolution of wavenumbers. The spatial derivatives can be approximates by finite-differences or using more flexible grid methods, such as finite element methods or even spectral finite element methods (Komatitsch et al. 1999).

The numerical simulation of seismic wave propagation is a large subject and different approaches and approximations can be taken. As such this is not a focus point of this thesis. Part of the purpose of applying ensemble-

based inversion is to have the details of the forward model out of the way, considering this as a black box.

Time-domain simulation of the elastic wave equation is computationally demanding as the wave field is the displacement vector field. An often applied practice is to consider only acoustic waves in the modeling. The measured waveform still contains recordings of shear waves, but as acoustic waves are dominating in the response, the absence of shear waves in predicting the observation can be considered an acceptable approximation. The simulation of an acoustic (scalar) wave field progation is considerably faster than its elastic counterpart (Fichtner 2010).

Another way of reducing computation time can be obtained by placing limitations on input, instead of output. By limiting the subsurface representation to horizontal layers wherein elastic attributes are constant, the elastic wave equation can be treated analytically and a synthetic seismogram computed significantly faster. This leads to the reflectivity method.

Reflectivity method

The reflectivity method has a long history and has been a workhorse method in reflection seismology. Kennett (2011) was among the key persons to develop the technique (Kennett and Kerry 1979) and is also the developer of the Erzsol3 (Kennett 2005) solver that has been used in this work.

The assumption of a stratified medium, where a 2D domain has variation only in the vertical direction whereas properties along the horizontal are constant, are of course a physical simplification but brings about computational efficiency. The elastic wave equation can in this case be transformed via cylindrical coordinates to the frequency-slowness domain and therein a solution can be derived analytically. This involves integration over wavenumbers and slowness, and the reflectivity method is also called a wavenumber integration method. The response of the stacked layers are calculated recursively and thus scales with the number of layers, but it nevertheless offers a significant speed-up compared to time-wise simulation of wave propagation. For computational aspects of the reflectivity method, see e.g. Müller (1985) and Mallick et al. (1987).

Forward model

Let the unknown state of interest be $\mathbf{x} \in \mathbb{R}^m$, representing a parameterization of the spatial field of elastic attributes in the subsurface. In case of a layered medium, the state dimension m would be number of layers, which

might be of order 10^1–10^3, times 3, the three elastic attributes. Whereas in a 2- or 3D wave simulation of Equation 1, the state dimension would reflect the computational grid in the region of interest and can be several orders of magnitude larger.

The state thus collects the elastic attributes $\mathbf{x} = [\log \mathbf{v}_p, \log \mathbf{v}_s, \log \boldsymbol{\rho}]^{\mathrm{T}}$, where $\mathbf{v}_p$ is the depth profile of acoustic velocity in the layers and similarly for the shear velocity $\mathbf{v}_s$ and density $\boldsymbol{\rho}$. The logarithmic transform is used to ensure that the physical properties remain positive.

The observed quantity is denoted $\mathbf{y} \in \mathbb{R}^p$ which holds the seismic waveforms over some measurement time for a series of receiver locations, i.e. a gather record. The operator relating the hidden state to observations is the forward model h. This operator implicitly holds all other parameters, such as source and receiver information, algorithmic parameters and possibly also some data processing. The forward model is considered perfect so that for the true state $\mathbf{x}^{\mathrm{t}}$, one would predict the true wavefield. Regrettably, the measurements are corrupted by errors and the observation model becomes

$$\mathbf{y} = h(\mathbf{x}) + \mathbf{e}\,. \tag{2}$$

The distribution of the measurement error $\mathbf{e}$ is assumed Gaussian with zero mean and covariance matrix $\mathbf{R}$.

For geophysical problems both model and observation space dimensions can be large. In seismic waveform applications, the data dimension p is on the order of: time sampling $\times$ length of recorded seismogram $\times$ number of receiver positions measuring seismograms $\times$ number of seismic shots. This quickly builds up to large dimensions, and seismic experiments generate vast amounts of data.

The number of data points, the sampled waveforms, is not the full picture. Figure 2 displays two synthetic seismograms at some arbitrary offset and time interval. They have been simulated using the same configuration in the reflectivity method, and the difference between the two is the frequency bandwidth. The trace with larger bandwidth obviously varies more than the other, which affects the possibility to represent variability though a sample covariance matrix.

Inversion

The inverse problem might simply be stated as $\mathbf{x} = h^{-1}(\mathbf{y})$ but contrary to $h(\cdot)$ which is defined via differential equations, $h^{-1}(\cdot)$ is unknown. Instead, the problem is reformulated into one of minimizing a functional of the misfit between prediction $\mathbf{y} = h(\mathbf{x})$ and observation $\mathbf{y}^{\mathrm{o}}$, assuming that if $\mathbf{y}$ is

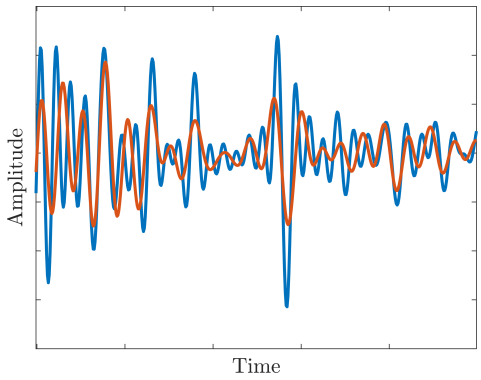


Figure 2: Examples of waveforms with different information level. Blue has larger frequency bandwidth (higher upper frequency limit) than red.

close to $\mathbf{y}^{\mathrm{o}}$, then $\mathbf{x}$ is close to the true state $\mathbf{x}^{\mathrm{t}}$. Such a data misfit functional can take various forms and recently FWI research is increasingly focused on this component. Still, the classical L_2-norm functional is much used and defines the optimization/variational problem as

$$\arg\min_{\mathbf{x}} \left\| \mathbf{y}^{\mathrm{o}} - h(\mathbf{x}) \right\|^2 .$$

If the error in observations are thought to have different uncertainty or are correlated, the least squares functional can be changed into a weighted or generalized least squares formulation:

$$\arg\min_{\mathbf{x}} \left\| \mathbf{y}^{\mathrm{o}} - h(\mathbf{x}) \right\|_{\mathbf{R}}^2 ,$$

where $\mathbf{R}$ is some weight matrix and the weighted norm squared means $\left\|\mathbf{a}\right\|_{\mathbf{B}}^2 = \mathbf{a}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{a}$.

Inverse problems are characterized by being ill-posed. A problem being well-posed means (i) a solution $\mathbf{x}$ exists, (ii) the solution is unique and (iii) that the solution depends continously on the data.

Seismic inverse problems are ill-posed. Seismic tomography is inherently non-unique in that different models for the subsurface can cause (almost) identical observations.

To address ill-posedness of a problem, one turns to regularizing it to make it behave a little more well-posed. Regularization can be seen as adding information to the problem in order to constrain it. Tikhonov regularization (Vogel 2002) is one common type of applied regularization that

adds a penalty term to the objective

$$\arg\min_{\mathbf{x}} \left\|\mathbf{y}^{\mathrm{o}} - h(\mathbf{x})\right\|^2_{\mathbf{R}} + \alpha\left\|\mathbf{L}(\mathbf{x} - \mathbf{x}^{\mathrm{f}})\right\|^2 ,$$

with α being a weight factor between the data misfit and the penalty and $\mathbf{L}$ is a structural matrix. The weight parameter is often referred to as the regularization parameter and $\mathbf{L}$ the Tikhnonov matrix. The reference solution $\mathbf{x}^{\mathrm{f}}$ is supposed to be a reasonable initial guess and constrains the solution $\mathbf{x}$ to be in its vicinity, and the regularization thus penalizes solutions of larger magnitude and deviations from this reference solution. If $\mathbf{L} = \mathbf{I}$, the regularization is called norm damping and the regularized solution remains close to $\mathbf{x}^{\mathrm{f}}$. If instead $\mathbf{L}$ is taken as a finite difference approximation to a derivative (first- or second-order), the regularizing is denoted derivative damping which will favorize smoother solutions.

Linearizing the observation operator

Let $\mathbf{P}_{\mathrm{f}}^{-1} = \mathbf{L}^{\mathrm{T}}\mathbf{L}$ and denote the variational problem $\arg\min_{\mathbf{x}} J(\mathbf{x})$ with the cost function given as

$$J(\mathbf{x}) = \tfrac{1}{2}\left\|\mathbf{y}^{\mathrm{o}} - h(\mathbf{x})\right\|^2_{\mathbf{R}} + \tfrac{1}{2}\left\|\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right\|^2_{\mathbf{P}_{\mathrm{f}}} .$$

The observational (first) part of the function $J(\mathbf{x})$ can be linearized in the vicinity of the state $\mathbf{x}^{\mathrm{f}}$, using a first order Taylor expansion

$$h(\mathbf{x}) = h(\mathbf{x}^{\mathrm{f}} + (\mathbf{x} - \mathbf{x}^{\mathrm{f}})) \approx h(\mathbf{x}^{\mathrm{f}}) + \mathbf{H}(\mathbf{x} - \mathbf{x}^{\mathrm{f}}) , \tag{3}$$

where the tangent linear (Jacobian) $\mathbf{H} = \frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}\big|_{\mathbf{x}=\mathbf{x}^{\mathrm{f}}}$ is introduced. Inserting this expansion into $J(\mathbf{x})$

$$\begin{aligned} 2J(\mathbf{x}) =& \left\|\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right\|^2_{\mathbf{P}_{\mathrm{f}}} + \left\|\mathbf{H}(\mathbf{x} - \mathbf{x}^{\mathrm{f}})\right\|^2_{\mathbf{R}} + \left\|\mathbf{y}^{\mathrm{o}} - h(\mathbf{x}^{\mathrm{f}})\right\|^2_{\mathbf{R}} \\ &- \left[\mathbf{y}^{\mathrm{o}} - h(\mathbf{x}^{\mathrm{f}})\right]^{\mathrm{T}} \mathbf{R}^{-1} \left[\mathbf{H}(\mathbf{x} - \mathbf{x}^{\mathrm{f}})\right] \\ &- \left[\mathbf{H}(\mathbf{x} - \mathbf{x}^{\mathrm{f}})\right]^{\mathrm{T}} \mathbf{R}^{-1} \left[\mathbf{y}^{\mathrm{o}} - h(\mathbf{x}^{\mathrm{f}})\right] \end{aligned}$$

and using the relation $\nabla(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}$ and $\nabla(\mathbf{d}^{\mathrm{T}}\mathbf{x}) = \nabla(\mathbf{x}^{\mathrm{T}}\mathbf{d}) = \mathbf{d}$ (Petersen et al. 2012), the resulting Jacobian and Hessian of $J(\mathbf{x})$ become

$$\begin{aligned} \nabla J(\mathbf{x}) &= \left[\mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right](\mathbf{x} - \mathbf{x}^{\mathrm{f}}) - \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y}^{\mathrm{o}} - h(\mathbf{x}^{\mathrm{f}})\right] \\ \nabla^2 J(\mathbf{x}) &= \mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} . \end{aligned}$$

Setting $\nabla J(\mathbf{x}) = \mathbf{0}$ and rearranging, the solution can written (Asch et al. 2016) as

$$\mathbf{x} = \mathbf{x}^{\mathrm{f}} + \left[\mathbf{P}_{\mathrm{f}}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y}^{\mathrm{o}} - h(\mathbf{x}^{\mathrm{f}})\right] . \tag{4}$$

This form of solution will come back when considering the stochastic filtering.

3 Bayesian inversion

The Bayesian viewpoint is that the unknown state vector is a random variable with an assigned prior probability distribution $p(\mathbf{x})$. Likewise, the observation vector is considerd random and the measurement is a realisation of this. The target of Bayesian inversion, or Bayesian inference, is the posterior probability distribution $p(\mathbf{x}|\mathbf{y})$, the conditional distribution of the state given an observation.

The foundation is Bayes' rule. Once the prior and likelihood function $p(\mathbf{y}|\mathbf{x})$ are specified, the posterior is:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x})}{p(\mathbf{y})} \propto p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x}) \ ,$$

where the expression to the right drops the marginal likelihood $p(\mathbf{y})$ – the probability of $\mathbf{y}$ prior to its observation – that acts as normalizing constant. This quantity is cumbersome to compute and not necessary as such, so it is conveniently dropped.

With an additive error structure in the observation model and an assumption of Gaussian distributed observation errors $p(\mathbf{e}) = \mathrm{N}(\mathbf{0}, \mathbf{R})$, the likelihood is Gaussian and

$$p(\mathbf{y}|\mathbf{x}) \propto \exp\left(-\tfrac{1}{2}\left\|\mathbf{y}^{\mathrm{o}} - h(\mathbf{x}^{\mathrm{f}})\right\|_{\mathbf{R}}^{2}\right) .$$

Assuming in addition a Gaussian prior $p(\mathbf{x}) = \mathrm{N}\big(\mathbf{x}^{\mathrm{f}}, \mathbf{P}_{\mathrm{f}}\big)$, the posterior becomes

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\tfrac{1}{2}\left(\left\|\mathbf{y}^{\mathrm{o}} - h(\mathbf{x}^{\mathrm{f}})\right\|_{\mathbf{R}}^{2} + \left\|\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right\|_{\mathbf{P}_{\mathrm{f}}}^{2}\right)\right) .$$

The state that maximizes the posterior density equals the solution that minimizes the (negative) log-posterior $-\log p(\mathbf{x}|\mathbf{y})$ which is equivalent to minimization of the previously introduced cost function

$$J(\mathbf{x}) = \tfrac{1}{2}\left\|\mathbf{y}^{\mathrm{o}} - h(\mathbf{x})\right\|_{\mathbf{R}}^{2} + \tfrac{1}{2}\left\|\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right\|_{\mathbf{P}_{\mathrm{f}}}^{2} . \tag{5}$$

As the most likely estimate of the true state, the choice of the state with maximum (posterior) probability,

$$\mathbf{x}^{\mathrm{a}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \arg\min_{\mathbf{x}} J(\mathbf{x})$$

is expected to be adequate.

The Bayesian interpretation of the additional regularization term is an imposed prior on the state. But framing the inverse problem in a Bayesian setting provides a more consistent approach, than the more ad-hoc addition of a regularization penalty term. With a prescribed prior the state is added information and regularization follows. The approach is not necessarily that much easier as the distributions need specification.

If the prior is Gaussian and the likelihood Gauss-linear, then closed form expressions for the first two moments of the Gaussian posterior is available. The second moment enables direct quantification of uncertainty through the variance-covariance information. When the observation model is nonlinear, uncertainty quantification is more difficult.

4 Sequential filtering

Stochastic filters most often deal with dynamic states, whereas in this parameter estimation context, the state is static. The subsurface properties might change over longer time, but do not change during the seismic experiment of some seconds recording time. Sequential here refers to the entire data set $\mathbf{y}$ being conditioned upon, is partitioned into disjoint sets $\mathbf{y}_k$ such that $\mathbf{y} = \cup_{k=1}^{K} \mathbf{y}_k$, which are processed in turn. For each data partition a prior distribution is used to predict observations, where the prior is the advancement of the current knowledge to predict these next observations. Taking into account the information provided by the actual observations, this knowledge is improved and is represented by the posterior distribution. This prediction followed by a correction constitutes an assimilation cycle. The updated knowledge then serves to predict the next set of observations. This is then a sequential process of estimation and uncertainty reduction.

With $\mathbf{x}$ given, observations $\mathbf{y}$ are independent since the measurement error sequence is assumed independent. Therefore

$$p(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K|\mathbf{x}) = \prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}) \ ,$$

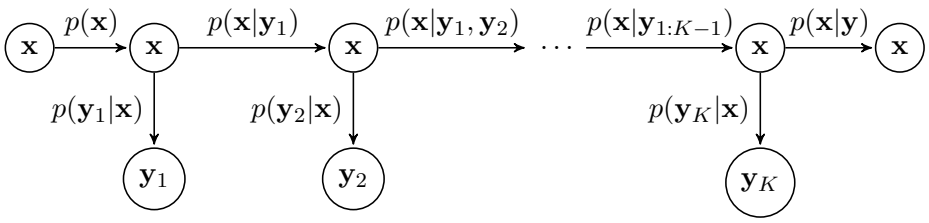


Figure 3: Diagram of the recursive structure. The unobserved variables $\mathbf{x}$ are inferred from observed variables $\mathbf{y}$. Each conditioning cycle provides the forecast distribution for the next observation.

and Bayes' rule can equivalently be written

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x}) \prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}) \ .$$

From the product form of likelihood, the order of assimilating or conditioning can principally be arbitrary, but the physical relation between state and observations does induce an order. Each data partition $\mathbf{y}_k$ is denoted a data assimilation window (DAW).

The sequential or recursive structure of conditioning is depicted in Figure 3, where the horizontal arrows entering a node indicates the prior, the vertical arrow the likelihood and the horizontal leaving a node is the (sequential) posterior.

In what follows, the (sequential) prior distribution is also referred to as the forecast distribution, as this forecasts or predicts the state distribution that causes the next observation. The sub-/superscript 'f' is used to denote this. Similarly, the posterior distribution resulting from conditioning on observation is dubbed the analysis state distribution, and is denoted with sub-/superscript 'a'.

State space model

The state vector describes static parameters which does not evolve dynamically between analysis and forecast state, and the persistence model (Bocquet and Sakov 2013) is a natural choice as forecast model for the state between assimilation cycles. Accordingly, the first two moments of the observation forecast distribution, using the observation model (2), are:

$$\mathrm{E}\left[\mathbf{y}_k|\mathbf{y}_{1:k-1}\right] = \mathbf{y}_k^{\mathrm{f}} = h_k(\mathbf{x}_k^{\mathrm{f}}) \ , \quad \mathrm{cov}\left[\mathbf{y}_k|\mathbf{y}_{1:k-1}\right] = \mathrm{cov}\left[h_k(\mathbf{x}|\mathbf{y}_{1:k-1})\right] + \mathbf{R}_k$$

where the forecast state estimate $\mathbf{x}_k^{\mathrm{f}} = \mathrm{E}\left[\mathbf{x}|\mathbf{y}_{1:k-1}\right]$, the forward model h_k forecasts the kth observation set and $\mathbf{R}_k$ is the observation error covariance matrix associated with this subset.

In Section 5-7 the cycle index k is omitted as this part concerns a single assimilation cycle.

5 Kalman filter, extended and iterative

In this section a quick description of the Kalman filter (KF), the extended Kalman filter (EKF) and the iterative extended Kalman filter (iEKF) is provided. This gives a basic picture of the methods and its components, before going into the ensemble-based approximation.

Kalman filter

Considering first the Gauss-linear case. With a Gaussian prior on $\mathbf{x} \sim \mathrm{N}\big(\mathbf{x}^{\mathrm{f}}\,,\,\boldsymbol{\Sigma}_{\mathrm{f}}\big)$, a linear observation model $\mathbf{y} = \mathbf{H}\mathbf{x}+\mathbf{e}$ so that $\mathbf{y}|\mathbf{x} \sim \mathrm{N}\big(\mathbf{H}\mathbf{x}^{\mathrm{f}}\,,\,\mathbf{R}\big)$. Then, the joint (Gaussian) distribution of $[\mathbf{x},\mathbf{y}]^{\mathrm{T}}$ is

$$\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix}\mathbf{x}^{\mathrm{f}}\\ \mathbf{H}\mathbf{x}^{\mathrm{f}}\end{bmatrix}\,,\,\begin{bmatrix}\boldsymbol{\Sigma}_{\mathrm{f}} & \boldsymbol{\Sigma}_{\mathrm{f}}\,\mathbf{H}^{\mathrm{T}}\\ \mathbf{H}\boldsymbol{\Sigma}_{\mathrm{f}} & \mathbf{H}\boldsymbol{\Sigma}_{\mathrm{f}}\,\mathbf{H}^{\mathrm{T}}+\mathbf{R}\end{bmatrix}\right)\,.$$

Given an observation $\mathbf{y}^{\mathrm{o}}$, the conditional distribution of $\mathbf{x}|\mathbf{y} = \mathbf{y}^{\mathrm{o}}$ is also Gaussian (Johnson et al. 2007) with mean and covariance

$$\begin{aligned}\mathbf{x}^{\mathrm{a}} &= \mathbf{x}^{\mathrm{f}} + \boldsymbol{\Sigma}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\boldsymbol{\Sigma}_{\mathrm{f}}\,\mathbf{H}^{\mathrm{T}}+\mathbf{R}\right)^{-1}\left(\mathbf{y}^{\mathrm{o}}-\mathbf{H}\mathbf{x}^{\mathrm{f}}\right)\\ \boldsymbol{\Sigma}_{\mathrm{a}} &= \boldsymbol{\Sigma}_{\mathrm{f}} - \boldsymbol{\Sigma}_{\mathrm{f}}\,\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\boldsymbol{\Sigma}_{\mathrm{f}}\,\mathbf{H}^{\mathrm{T}}+\mathbf{R}\right)^{-1}\mathbf{H}\boldsymbol{\Sigma}_{\mathrm{f}}\,.\end{aligned}$$

These equations for the conditional moments are essentially the Kalman analysis/update equation (Madsen 2007). Introducing the Kalman gain matrix

$$\mathbf{K} = \mathrm{cov}\left[\mathbf{x},\mathbf{y}\right]\mathrm{cov}\left[\mathbf{y}\right]^{-1} = \boldsymbol{\Sigma}_{\mathrm{f}}\,\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\boldsymbol{\Sigma}_{\mathrm{f}}\,\mathbf{H}^{\mathrm{T}}+\mathbf{R}\right)^{-1} \tag{6}$$

the analysis equations are written in a more common form

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{f}} + \mathbf{K}\left(\mathbf{y}^{\mathrm{o}}-\mathbf{H}\mathbf{x}^{\mathrm{f}}\right) \tag{7a}$$

$$\boldsymbol{\Sigma}_{\mathrm{a}} = \left(\mathbf{I}-\mathbf{K}\mathbf{H}\right)\boldsymbol{\Sigma}_{\mathrm{f}}\,. \tag{7b}$$

These Kalman filter formulations are valid in its strictest sense only for Gauss-linear systems. They describe how the information from an observation affects the state estimate through cross-covariances.

The gain matrix (6) can, using the Woodbury identity (Petersen et al. 2012);

$$\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1}+\mathbf{D}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}=\left(\mathbf{A}+\mathbf{B}\mathbf{C}\mathbf{D}\right)^{-1}\mathbf{B}\mathbf{C}\,,$$

equivalently be written as

$$\mathbf{K}=\left(\boldsymbol{\Sigma}_{\mathrm{f}}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\,.$$

Thus, the solution to the variational problem (4) corresponds for a Gauss-linear system to the conditional mode, which for Gaussian distributions equals the mean. This is the connection between the variational and the statistical approach to the estimation problem.

When the system is Gauss-linear, the Kalman filter is optimal in the sense that it fully describes the (Gaussian) conditional distribution by its mean and covariance. In case the observation operator is nonlinear, the joint distribution of $[\mathbf{x},\mathbf{y}]^{\mathrm{T}}$ is no longer Gaussian and neither is the distribution of $\mathbf{x}|\mathbf{y}$. The Kalman filter is then suboptimal as the true conditional mean and covariance will differ from the filter estimates. One way to circumvent the effect of nonlinearity is by invoking the linearization (3) and assume a Gaussian approximation (Särkkä 2013). This leads to the formulation of the extended Kalman filter.

Extended Kalman filter

The extended Kalman filter relies on the linearization (3) of the nonlinear forward model. In this filter formulation, the mean forecast uses the nonlinear forward model directly, while the error covariance is predicted using the gradient $\mathbf{H}$. The linearization approximation is not always valid and can make the filter performance behave badly, if the data is more than weakly nonlinear.

The measurement prediction

$$\begin{aligned}\mathrm{E}\left[\mathbf{y}\right]&=\mathrm{E}\left[h(\mathbf{x})\right]+\mathrm{E}\left[\mathbf{e}\right]=\mathrm{E}\left[h(\mathbf{x})\right]\\&\approx\mathrm{E}\left[h(\mathbf{x}^{\mathrm{f}})+\mathbf{H}(\mathbf{x}-\mathbf{x}^{\mathrm{f}})\right]=h(\mathbf{x}^{\mathrm{f}})+\mathbf{H}\mathrm{E}\left[\mathbf{x}-\mathbf{x}^{\mathrm{f}}\right]=h(\mathbf{x}^{\mathrm{f}})\end{aligned}\tag{8}$$

is hence the propagation of the state estimate through the nonlinear forward model. The predicted error covariance when inserting the linearization becomes

$$\begin{aligned}
\operatorname{cov}[\mathbf{y}] &= \mathrm{E}\left[(h(\mathbf{x}) - \mathrm{E}[h(\mathbf{x})])\,(h(\mathbf{x}) - \mathrm{E}[h(\mathbf{x})])^{\mathrm{T}}\right] + \operatorname{cov}[\mathbf{e}] \\
&\approx \mathrm{E}\left[\left(h(\mathbf{x}^{\mathrm{f}}) + \mathbf{H}(\mathbf{x} - \mathbf{x}^{\mathrm{f}}) - h(\mathbf{x}^{\mathrm{f}})\right)\left(h(\mathbf{x}^{\mathrm{f}}) + \mathbf{H}(\mathbf{x} - \mathbf{x}^{\mathrm{f}}) - h(\mathbf{x}^{\mathrm{f}})\right)^{\mathrm{T}}\right] \\
&\quad + \operatorname{cov}[\mathbf{e}] \\
&= \mathrm{E}\left[\mathbf{H}(\mathbf{x} - \mathbf{x}^{\mathrm{f}})(\mathbf{x} - \mathbf{x}^{\mathrm{f}})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right] + \operatorname{cov}[\mathbf{e}] \\
&= \mathbf{H}\mathrm{E}\left[(\mathbf{x} - \mathbf{x}^{\mathrm{f}})(\mathbf{x} - \mathbf{x}^{\mathrm{f}})^{\mathrm{T}}\right]\mathbf{H}^{\mathrm{T}} + \operatorname{cov}[\mathbf{e}] = \mathbf{H}\boldsymbol{\Sigma}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\,,
\end{aligned} \tag{9}$$

as the measurement error $\mathbf{e}$ and the predicted observation are assumed independent. Similarly for the cross-covariance

$$\begin{aligned}
\operatorname{cov}[\mathbf{x}, \mathbf{y}] &= \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])\,(h(\mathbf{x}) + \mathbf{e} - \mathrm{E}[\mathbf{y}])^{\mathrm{T}}\right] \\
&\approx \mathrm{E}\left[\left(\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right)\left(h(\mathbf{x}^{\mathrm{f}}) + \mathbf{H}(\mathbf{x} - \mathbf{x}^{\mathrm{f}}) + \mathbf{e} - h(\mathbf{x}^{\mathrm{f}})\right)^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[(\mathbf{x} - \mathbf{x}^{\mathrm{f}})(\mathbf{x} - \mathbf{x}^{\mathrm{f}})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right] = \boldsymbol{\Sigma}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\,,
\end{aligned} \tag{10}$$

with the assumption $\operatorname{cov}[\mathbf{x}, \mathbf{e}] = \mathbf{0}$, that state and error also are independent.

So the analysis equation for the state mean looks the same as for the linear case, except for the expected observation using the nonlinear forward model

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{f}} + \operatorname{cov}[\mathbf{x}, \mathbf{y}]\operatorname{cov}[\mathbf{y}]^{-1}\left(\mathbf{y}^{\mathrm{o}} - h(\mathbf{x}^{\mathrm{f}})\right)\,, \tag{11}$$

whereas the analysis covariance is identifical to Equation 7b apart from it using the gradient (or tangent linear model).

Iterative extended Kalman filter

When the forward model is more than weakly nonlinear, the linearization (3) around the forecast state $\mathbf{x}^{\mathrm{f}}$ is questionable. The analysis state can then benefit from being found iteratively. Jazwinski (1970) called this approach the iterated extended Kalman filter.

The benefit of iterating comes from a re-linearization around an iteratively improved reference estimate $\mathbf{x}_j$,

$$h(\mathbf{x}^{\mathrm{f}}) \approx h(\mathbf{x}_j) + \mathbf{H}_j(\mathbf{x}^{\mathrm{f}} - \mathbf{x}_j)\,,$$

with the gradient $\mathbf{H}_j = \mathbf{H}(\mathbf{x}_j) = \partial_{\mathbf{x}} h(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_j}$. The update form follows the EKF analysis, which is then iteratively repeated as

$$\mathbf{x}_{j+1} = \mathbf{x}^{\mathrm{f}} + \mathbf{K}(\mathbf{x}_j) \left(\mathbf{y}^{\mathrm{o}} - h(\mathbf{x}_j) - \mathbf{H}_j(\mathbf{x}^{\mathrm{f}} - \mathbf{x}_j) \right) ,$$

where the iterate state vector $\mathbf{x}_j$ initializes as $\mathbf{x}_0 = \mathbf{x}^{\mathrm{f}}$, and the converged solution $\mathbf{x}^*$ defines the analysis state $\mathbf{x}^{\mathrm{a}} = \mathbf{x}^*$. The first iteration $\mathbf{x}_1$ thus equals the EKF analysis (11). The iterate Kalman gain equals

$$\mathbf{K}(\mathbf{x}_j) = \mathbf{\Sigma}_{\mathrm{f}} \mathbf{H}_j^{\mathrm{T}} \left(\mathbf{H}_j \mathbf{\Sigma}_{\mathrm{f}} \mathbf{H}_j^{\mathrm{T}} + \mathbf{R} \right)^{-1} .$$

The forecast covariance $\mathbf{\Sigma}_{\mathrm{f}}$ is fixed, and the analysis covariance uses the final re-evaluation as $\mathbf{\Sigma}_{\mathrm{a}} = \big(\mathbf{I} - \mathbf{K}(\mathbf{x}^{\mathrm{a}}) \mathbf{H}(\mathbf{x}^{\mathrm{a}}) \big) \mathbf{\Sigma}_{\mathrm{f}}$.

6 Ensemble Kalman filter

For problems with a large state dimension m, the storage of the $m \times m$ covariance matrix becomes impractical, even impossible. Instead of propagating a state estimate and the state error covariance, the ensemble approach uses a set of states to propagate this information.

Another issue of practical/realistic problems is that they often are nonlinear of nature and that linearizing them are cumbersome, if possible at all. The forward model is most often the numerical solution from a complex implementation of a mathematical model, which is far from straightforward to linearize.

The frameworks inherent inclination towards use of black box models and the parallel evaluation of this, makes it very suitable to modern computational resources.

Ensemble and Monte Carlo

The Monto Carlo principle is often expressed as an approximation to the expectation operator:

$$\mathrm{E}\left[f(\mathbf{x}) \right] \approx \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_{[i]}) , \tag{12}$$

where the approximation improves as the number of Monte Carlo samples $n \to \infty$. For a multivariate random variable $\mathbf{x}$, the sample mean

$$\mathrm{E}\left[\mathbf{x} \right] \approx \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_{[i]} = \bar{\mathbf{x}} , \tag{13}$$

and the sample covariance

$$\begin{aligned} \operatorname{cov}[x_j, x_k] &= \operatorname{E}[(x_j - \operatorname{E}[x_j])(x_k - \operatorname{E}[x_k])] \\ &\approx \frac{1}{n-1} \sum_{i=1}^{n} \left(x_{[i],j} - \bar{\mathbf{x}}_j\right)\left(x_{[i],k} - \bar{\mathbf{x}}_k\right) , \end{aligned} \tag{14}$$

where $x_{[i],j}$ is the ith sample of the jth variable.

An ensemble is a set of state samples from its underlying distribution. Whereas in the previous section, the Gausian distribution was fully described by its first two moments. Now the distribution is represented by this sample set and the first two moments estimated from these. These samples are stored in an ensemble matrix $\mathbf{E}^{\mathrm{f}} = \left[\mathbf{x}^{\mathrm{f}}_{[1]}\ \mathbf{x}^{\mathrm{f}}_{[2]}\ \cdots\ \mathbf{x}^{\mathrm{f}}_{[n]}\right]$ where each sample (column) is called an ensemble member.

The sample mean and covariance of the forecast distribution are

$$\bar{\mathbf{x}}^{\mathrm{f}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}^{\mathrm{f}}_{[i]} = \mathbf{E}^{\mathrm{f}}\, \mathbf{1}/n \tag{15a}$$

$$\mathbf{P}_{\mathrm{f}} = \mathbf{X}_{\mathrm{f}}\, \mathbf{X}_{\mathrm{f}}^{\mathrm{T}} , \tag{15b}$$

where the $m \times n$ forecast anomaly (or perturbation) matrix $\mathbf{X}_{\mathrm{f}}$ is

$$\mathbf{X}_{\mathrm{f}} = \left(\mathbf{E}^{\mathrm{f}} - \bar{\mathbf{x}}^{\mathrm{f}}\, \mathbf{1}^{\mathrm{T}}\right) / (n-1)^{1/2} . \tag{16}$$

Similarly, a $p \times n$ observation anomaly matrix $\mathbf{Y}_{\mathrm{f}}$ is defined by

$$\mathbf{Y}_{\mathrm{f}} = \left(h(\mathbf{E}^{\mathrm{f}}) - \bar{\mathbf{y}}^{\mathrm{f}}\, \mathbf{1}^{\mathrm{T}}\right) / (n-1)^{1/2} ,$$

using the expected observation $\bar{\mathbf{y}}^{\mathrm{f}}$. The expected observation can be taken as the ensemble average $\bar{\mathbf{y}}^{\mathrm{f}} = h(\mathbf{E}^{\mathrm{f}})\mathbf{1}/n$ or as the forward evaluation of the (state) ensemble mean $\bar{\mathbf{y}}^{\mathrm{f}} := h(\bar{\mathbf{x}}^{\mathrm{f}})$. Both have been used and they do not show significant differences for the application considered.

By viewing this observation anomaly matrix as $\mathbf{Y}_{\mathrm{f}} \equiv \mathbf{H}\mathbf{X}_{\mathrm{f}}$, with $\mathbf{H}$ the forward model gradient at the forecast estimate $\mathbf{H} = \partial_{\mathbf{x}} h(\bar{\mathbf{x}}^{\mathrm{f}})$, a major obstacle is bypassed by the operator $\mathbf{H}$ not being explicitly necessary. Instead, this is approximated by an ensemble linearization. It relies on the assumption that linearizing around the mean state is acceptable/passable, thus $h(\mathbf{x}^{\mathrm{f}}_{[i]}) \approx h(\bar{\mathbf{x}}^{\mathrm{f}}) + \mathbf{H}(\mathbf{x}^{\mathrm{f}}_{[i]} - \bar{\mathbf{x}}^{\mathrm{f}})$ and $h(\bar{\mathbf{x}}^{\mathrm{f}}) \approx \bar{\mathbf{y}}^{\mathrm{f}}$, hence $h(\mathbf{x}^{\mathrm{f}}_{[i]}) - \bar{\mathbf{y}}^{\mathrm{f}} \approx$

$\mathbf{H}(\mathbf{x}^{\mathrm{f}}_{[i]} - \bar{\mathbf{x}}^{\mathrm{f}})$ and therefore $\mathbf{Y}_{\mathrm{f}} = \mathbf{H}\mathbf{X}_{\mathrm{f}}$. If the forward model is linear, this replacement is of course exact.

Inserting the ensemble approximation (15b) into the cross-covariance (10) and the covariance (9), the ensemble versions of these become

$$\begin{aligned} \mathrm{cov}\left[\mathbf{x}, \mathbf{y}\right] &= \mathbf{X}_{\mathrm{f}}\,\mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \\ \mathrm{cov}\left[\mathbf{y}\right] &= \mathbf{Y}_{\mathrm{f}}\,\mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} + \mathbf{R}\,, \end{aligned}$$

and the ensemble version of the Kalman gain (6) is written as

$$\mathbf{K} = \mathbf{X}_{\mathrm{f}}\,\mathbf{Y}_{\mathrm{f}}^{\mathrm{T}}\left(\mathbf{Y}_{\mathrm{f}}\,\mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} + \mathbf{R}\right)^{-1}. \tag{17}$$

The method referred to as EnKF (Evensen 2009) uses this ensemble gain (17), along with perturbed observations $\mathbf{y}^{\mathrm{o}}_{[i]} = \mathbf{y}^{\mathrm{o}} + \mathbf{e}_{[i]}$, $\mathbf{e}_{[i]} \sim \mathrm{N}(\mathbf{0}, \mathbf{R})$, to update each ensemble as

$$\mathbf{x}^{\mathrm{a}}_{[i]} = \mathbf{x}^{\mathrm{f}}_{[i]} + \mathbf{K}(\mathbf{y}^{\mathrm{o}}_{[i]} - \mathbf{y}^{\mathrm{f}}_{[i]})\,.$$

The addition of errors to the observations is to ensure the correct posterior covariance, in the limit $n \to \infty$. This perturbation of observations adds a source of sampling error in addition to the inherent sampling error caused by having a finite sized ensemble. As an alternative, square root filters have been developed that seeks to update the ensemble, still to have correct posterior covariance (in the linear case) but without adding sampling.

Ensemble covariances

A fundamental challenge common to ensemble Kalman filters in its various forms, is the ensemble based estimation of error covariance matrices. The finite size of the ensemble makes the sample estimators used in the Kalman gain (17), prone to what is referred to as spurious correlations, an umbrella term that covers the effect of sampling errors due to the limited sample size.

Spurious correlations between certain parameters and observations that principally are uncorrelated, causes these parameters to be updated in the analysis, along with a reduction in variance as well. Over sequential assimilation cycles, this can cause significant underestimation of the ensemble variance.

Two techniques to alleviate the impact of sampling errors are often applied in practical applications, namely localization and inflation.

Covariance inflation artificially increases the ensemble covariance with an inflation factor larger than one. This can be applied to either forecast or

analysis ensemble and in a multiplicative or additive fashion. The inflation factor can be constant or adaptive, global or local, and there is no general rule on how to do ensemble inflation optimally.

The localization technique is divided in two classes; domain and covariance localization, each with its strengths and drawbacks. Sakov and Bertino (2011) showed that their effect are comparable when the analysis update is relatively small. Domain localization, or local analysis, refers to local state variables (single variable or spatially close group) are updated by assimilating only observations in its spatial vicinity. The Local Ensemble Transform Kalman Filter is an example of a local analysis formulation. Covariance localization on the other hand, seeks to modify/weight the forecast error covariance matrix such that distant (in space and time) observations are forced to have zero correlation. This is most often achieved through the element-wise multiplication of a tapering matrix, but not all filter formulations support this.

Square root filter

The concept of square root filters (SRFs) was introduced shortly after the Kalman paper in the early 1960's, as practitioners saw numerical issues (in precision and stability) in the direct implementation of the Kalman filter. Back then computers were more limited than today, and square root formulations alleviated some of the problems. While with modern ensemble methods, the concern that SRFs is supposed to address is the sampling error introduced when using the stochastic observation perturbation of the standard/stochastic EnKF. Therefore, the ensemble SRFs are often refered to as deterministic ensemble Kalman filters.

A (matrix) square root $\mathbf{S}$ of a covariance matrix $\mathbf{P}$ is a matrix such that $\mathbf{P} = \mathbf{S}\mathbf{S}^{\mathrm{T}}$ and SRFs concerns how to update the square root matrix. The Potter method was introduced in early 1960s and its presentation here closely follows its explanation in Bierman (1977) and Tippett et al. (2003). The Kalman equation for the covariance update $\mathbf{P}_{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}_{\mathrm{f}} = \mathbf{P}_{\mathrm{f}} - \mathbf{K}\mathbf{D}\mathbf{K}^{\mathrm{T}}$ with $\mathbf{D} = \mathbf{H}\mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}$ and $\mathbf{K} = \mathbf{P}_{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{-1}$. Then introducing the square root form for both forecast and analysis covariances as $\mathbf{S}_{\mathrm{f}}$ and $\mathbf{S}_{\mathrm{a}}$ respectively, the update can be written as

$$\begin{aligned}
\mathbf{P}_{\mathrm{a}} &= \mathbf{S}_{\mathrm{a}}\mathbf{S}_{\mathrm{a}}^{\mathrm{T}} = \mathbf{S}_{\mathrm{f}}\mathbf{S}_{\mathrm{f}}^{\mathrm{T}} - \mathbf{S}_{\mathrm{f}}\mathbf{S}_{\mathrm{f}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{H}\mathbf{S}_{\mathrm{f}}\mathbf{S}_{\mathrm{f}}^{\mathrm{T}} \\
&= \mathbf{S}_{\mathrm{f}}\left[\mathbf{I} - \mathbf{S}_{\mathrm{f}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{H}\mathbf{S}_{\mathrm{f}}\right]\mathbf{S}_{\mathrm{f}}^{\mathrm{T}} \\
&= \mathbf{S}_{\mathrm{f}}\left[\mathbf{I} - \mathbf{V}\mathbf{D}^{-1}\mathbf{V}^{\mathrm{T}}\right]\mathbf{S}_{\mathrm{f}}^{\mathrm{T}} \text{ with } \mathbf{V} = \left(\mathbf{H}\mathbf{S}_{\mathrm{f}}\right)^{\mathrm{T}} \text{ and } \mathbf{D} = \mathbf{V}^{\mathrm{T}}\mathbf{V} + \mathbf{R}\,,
\end{aligned} \tag{18}$$

from which the choice of $\mathbf{S}_\mathrm{a} = \mathbf{S}_\mathrm{f} \left(\mathbf{I} - \mathbf{V}\mathbf{D}^{-1}\mathbf{V}^\mathrm{T}\right)^{1/2}$ is intuitive. In the above, one sets $\mathbf{D}^{-1} = \left(\mathbf{D}^{-1}\right)^\mathrm{T}$ as $\mathbf{D}^{-1}$ is symmetric since $\mathbf{D}$ is symmetric.

The square root matrix is not unique so a great deal of flexibility is present in chosing a square root form, and various schemes use different approaches to this. Also, other matrix factorizations are possible: when the $m \times m$ covariance matrix $\mathbf{P}$ has at most rank r, there exists a $m \times r$ square root matrix $\mathbf{S}$ that fulfills the factorization. Some of the earlier application of the square-root approach to the (ensemble) analysis step was the ensemble adjustment Kalman filter (EAKF, Anderson 2001) and the ensemble transform Kalman filter (ETKF, Bishop et al. 2001). Tippett et al. (2003) analysed these in greater detail.

The square root of the forecast and analysis error covariance matrices replaces the full covariance matrices in the Kalman covariance update (7b), and becomes an update of the square root only. This is the common concept of SRFs.

Of other ensemble SRF schemes can be mentioned the singular evolutive interpolated Kalman filter (SEIK, Pham 1996; Pham 2001), which Nerger et al. (2012) showed that was indeed an ensemble SRF and that SEIK was more or less equivalent in performance to ETKF. During the last decade there has been a continued and ongoing effort to explore ensemble based SRFs and to make these increasing robust, especially towards non-linearity.

Ensemble transform Kalman filter

The foundation update scheme for the IEnKS is the ETKF, originally introduced by Bishop et al. (2001). Hunt, Kostelich, et al. (2007) revised and further developed it into the Local ETKF (LETKF) which is widely used within the numerical weather prediction (NWP) community. Also, in Hunt, Kostelich, et al. (2007) the ensemble subspace aspect is emphasized.

Ensemble subspace representation

Using the ensemble Kalman gain (17) in the analysis state equation (7a), one can see that the correction to the forecast state will always be found as a linear combination within the column space of $\mathbf{X}_\mathrm{f}$. This column space is referred to as the ensemble subspace (Hunt, Kostelich, et al. 2007). Making this parametrization of the analysis state explicit through a $n \times 1$ control variable $\mathbf{w}$, a state $\mathbf{x} = \mathbf{x}(\mathbf{w})$ is given as

$$\mathbf{x} = \bar{\mathbf{x}}^\mathrm{f} + \mathbf{X}_\mathrm{f}\,\mathbf{w}\,, \tag{19}$$

such that $\mathbf{w}$ acts as a coefficient vector within the ensemble subspace.

The analysis control $\mathbf{w}^{\mathrm{a}}$ that gives the analysis state $\mathbf{x}^{\mathrm{a}}$, can be found by inserting the parameterization into the Kalman equation (7a) along with the ensemble Kalman gain (17), resulting in

$$\mathbf{w}^{\mathrm{a}} = \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \left(\mathbf{Y}_{\mathrm{f}} \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} + \mathbf{R} \right)^{-1} (\mathbf{y}^{\mathrm{o}} - \bar{\mathbf{y}}^{\mathrm{f}}) .$$

If the control variable is standard Gaussian $\mathbf{w} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$, then the state $\mathbf{x} = \bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}} \mathbf{w}$ is also Gaussian $\mathbf{x} \sim \mathrm{N}\big(\bar{\mathbf{x}}^{\mathrm{f}}, \mathbf{X}_{\mathrm{f}} \mathbf{X}_{\mathrm{f}}^{\mathrm{T}}\big)$.

The ETKF is a SRF, so its characteristic is the update of the covariance square root, such that $\mathbf{X}_{\mathrm{a}} = \mathbf{X}_{\mathrm{f}} \mathbf{T}$ where $\mathbf{T}$ is the ensemble transform matrix. The analysis ensemble is then obtained by shifting the center of the analysis perturbations to the analysis mean

$$\mathbf{E}^{\mathrm{a}} = \big(\bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}} \mathbf{w}^{\mathrm{a}}\big) \mathbf{1}^{\mathrm{T}} + (n-1)^{1/2} \mathbf{X}_{\mathrm{f}} \mathbf{T} . \tag{20}$$

Ensemble transform matrix

Inserting the covariance estimator (15b) into the Kalman covariance update (7b), and using the matrix inversion lemma,

$$\begin{aligned}
\mathbf{X}_{\mathrm{a}} \mathbf{X}_{\mathrm{a}}^{\mathrm{T}} &= (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{X}_{\mathrm{f}} \mathbf{X}_{\mathrm{f}}^{\mathrm{T}} \\
&= \left(\mathbf{I} - \mathbf{X}_{\mathrm{f}} \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \left(\mathbf{Y}_{\mathrm{f}} \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} + \mathbf{R} \right)^{-1} \mathbf{H} \right) \mathbf{X}_{\mathrm{f}} \mathbf{X}_{\mathrm{f}}^{\mathrm{T}} \\
&= \mathbf{X}_{\mathrm{f}} \left(\mathbf{I} - \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \left(\mathbf{Y}_{\mathrm{f}} \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} + \mathbf{R} \right)^{-1} \mathbf{H} \mathbf{X}_{\mathrm{f}} \right) \mathbf{X}_{\mathrm{f}}^{\mathrm{T}} \\
&= \mathbf{X}_{\mathrm{f}} \left(\mathbf{I} - \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \left(\mathbf{Y}_{\mathrm{f}} \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} + \mathbf{R} \right)^{-1} \mathbf{Y}_{\mathrm{f}} \right) \mathbf{X}_{\mathrm{f}}^{\mathrm{T}}
\end{aligned}$$

from which it is seen, compared to Equation 18, that if $\mathbf{T}$ is chosen such $\mathbf{T}\mathbf{T}^{\mathrm{T}} = \mathbf{I} - \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \left(\mathbf{Y}_{\mathrm{f}} \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} + \mathbf{R} \right)^{-1} \mathbf{Y}_{\mathrm{f}}$, the covariance is updated according to the Kalman equation. This can be reduced even further to avoid the $p \times p$

matrix inversion:

$$\begin{aligned}
\mathbf{T}\mathbf{T}^{\mathrm{T}} &= \mathbf{I} - \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \left(\mathbf{Y}_{\mathrm{f}} \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} + \mathbf{R} \right)^{-1} \mathbf{Y}_{\mathrm{f}} \\
&= \mathbf{I} - \left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \right)^{-1} \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \\
&= \left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \right)^{-1} \left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \right) \\
&\quad - \left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \right)^{-1} \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \\
&= \left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \right)^{-1} \left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} - \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \right) \\
&= \left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \right)^{-1}
\end{aligned}$$

where the Woodbury formula

$$\mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \left(\mathbf{Y}_{\mathrm{f}} \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} + \mathbf{R} \right)^{-1} = \left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \right) \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} ,$$

is used in second step. Ergo the choice

$$\mathbf{T} = \left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_{\mathrm{f}} \right)^{-1/2} \tag{21}$$

fulfills the Kalman covariance equation.

The square root matrix in (21) is taken as the symmetric square root. The symmetric square root of a matrix $\mathbf{A}$ invokes the spectral decomposition $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$, where $\mathbf{V}$ contains the normalized eigenvectors in columns and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues. Then the symmetric square root is $\mathbf{A}^{1/2} = \mathbf{A}\mathbf{\Lambda}^{1/2}\mathbf{V}^{\mathrm{T}} = (\mathbf{A}^{1/2})^{\mathrm{T}}$ (e.g. Johnson et al. 2007). This form of square root preserves the mean of the anomalies (Wang et al. 2004; Sakov and Oke 2008), so that these remain centered around zero.

The transform used to update the ensemble anomalies is not unique and right multiplying the transform matrix with a (random) orthonormal matrix $\mathbf{U}$ with $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$, will still satisfy the analysis error covariance equation. The rotation matrix $\mathbf{U}$ must satisfy $\mathbf{U}\mathbf{1} = \mathbf{1}$ in order to be unbiased (Livings et al. 2008) but is otherwise free to choose. If its construction involves randomness, the determinism of the filter is gone.

7 Iterative ensemble Kalman smoother

The iterative ensemble Kalman smoother is a hybrid approach, combining a variational formulation to solve for the best guess, the ensemble mean,

and the use of an ensemble to approximate the linearization and cost function sensitivities (partial derivatives with respect to control variable). The method was initially derived as a filter (IEnKF) in (Sakov, Oliver, et al. 2012), aimed at NWP application. In this application, the main objective is to obtain the initial condition for simulating atmospheric systems over a time horizon. The idea is to use the information at the next observation time to improve the filter state at the current time. Shortly afterwards the method was extended to a smoother (Bocquet and Sakov 2014) with the main difference being that the horizon of observations is extended, enlarging the amount of information to estimate the current state. While the formulation was with a dynamic state in mind, the method can easily be adapted to static state estimation. It can be thought of as a deterministic version of the iterative ensemble smoother which is clearly apparent in its recent reformulation (Raanes et al. 2019).

The IEnKS is based on the subspace parameterization (19), and a cost function similar to Equation 5 is expressed in terms of the control variable $\mathbf{w}$. Considering first the log-prior term of (5). Inserting the parameterization and the ensemble state error covariance, the log-prior term takes the form

$$\begin{aligned}(\mathbf{x}-\bar{\mathbf{x}}^{\mathrm{f}})\mathbf{P}_{\mathrm{f}}^{-1}(\mathbf{x}-\bar{\mathbf{x}}^{\mathrm{f}}) &= (\bar{\mathbf{x}}^{\mathrm{f}}+\mathbf{X}_{\mathrm{f}}\mathbf{w}-\bar{\mathbf{x}}^{\mathrm{f}})^{\mathrm{T}}\left(\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\mathrm{T}}\right)^{-1}(\bar{\mathbf{x}}^{\mathrm{f}}+\mathbf{X}_{\mathrm{f}}\mathbf{w}-\bar{\mathbf{x}}^{\mathrm{f}}) \\ &= \mathbf{w}^{\mathrm{T}}\mathbf{X}_{\mathrm{f}}^{\mathrm{T}}\left(\mathbf{X}_{\mathrm{f}}^{-\mathrm{T}}\mathbf{X}_{\mathrm{f}}^{-1}\right)\mathbf{X}_{\mathrm{f}}\mathbf{w} \\ &= \mathbf{w}^{\mathrm{T}}\mathbf{w}\,.\end{aligned}$$

The inverses here must be understood in a pseudo-inverse fashion.

The corresponding cost function when the solution is found within the ensemble subspace thus becomes

$$J(\mathbf{w}) = \tfrac{1}{2}\left\|\mathbf{y}^{\mathrm{o}}-h(\bar{\mathbf{x}}^{\mathrm{f}}+\mathbf{X}_{\mathrm{f}}\mathbf{w})\right\|_{\mathbf{R}}^{2}+\tfrac{1}{2}\left\|\mathbf{w}\right\|^{2}\,,$$

whose minimization is the variational aspect of IEnKS and whose solution provides the analysis state. This non-linear least squares problem is solved iteratively via a Gauss-Newton (Baldick 2006) scheme

$$\mathbf{w}_{j+1} = \mathbf{w}_j - \mathbb{H}_j^{-1}\nabla J_j\,,$$

where j is the iteration index. The Jacobian ∇J and (approximative, ignoring second order derivatives) Hessian $\mathbb{H}$ are evaluated at the iterate mean $\mathbf{x}_j = \bar{\mathbf{x}}^{\mathrm{f}}+\mathbf{X}\mathbf{w}_j$ as

$$\begin{aligned}\nabla J_j &= \mathbf{w}_j + \mathbf{Y}_j^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}^{\mathrm{o}}-\bar{\mathbf{y}}_j\right) \\ \mathbb{H}_j &= \mathbf{I}+\mathbf{Y}_j^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_j\,,\end{aligned}$$

where the observation anomalies $\mathbf{Y}_j$ replace the explicit use of the tangent linear model and the sensitivities rely on the ensemble linearization.

With the solution $\mathbf{w}^{\mathrm{a}}$ to the variational problem and a transform matrix $\mathbf{T} = \mathbb{H}^{-1/2}\big|_{\mathbf{w}^{\mathrm{a}}}$, the analysis ensemble is available as in Equation 20.

Ensemble linearization

The observation anomalies $\mathbf{Y}_j$ are supposed to be the mapping of the prior state anomalies $\mathbf{X}_{\mathrm{f}}$ though the re-evaluation of the gradient around the iterate mean $\mathbf{H}_j = \mathbf{H}(\mathbf{x}_j)$, as in the iterative EKF. This can be achieved in two alternative approaches, called the transform and the bundle variants (Sakov, Oliver, et al. 2012; Asch et al. 2016).

The chosen approach in this work has been the transform variant. In this approach, the observation anomalies are formed from the forward evaluation of a preconditioned ensemble. This is an ensemble that is centered at the iterate mean $\mathbf{x}_j$ and whose anomalies have been transformed using the current estimate of the transform matrix $\mathbf{T}_j = \mathbb{H}_j^{-1/2}$, i.e.

$$\mathbf{E}_j = \mathbf{x}_j \mathbf{1}^{\mathrm{T}} + (n-1)^{1/2} \mathbf{X}_{\mathrm{f}} \mathbf{T}_j \, .$$

Evaluating the observation anomalies would relate these to the preconditioned ensemble $\mathbf{Y} = \mathbf{H}_j \mathbf{X}_j$ where $\mathbf{X}_j = \mathbf{X}_{\mathrm{f}} \mathbf{T}_j$. Therefore the anomalies are right multiplied by the inverse transform so that $\mathbf{Y}_j = \mathbf{Y}\mathbf{T}_j^{-1} = \mathbf{H}_j \mathbf{X}_j \mathbf{T}_j^{-1} = \mathbf{H}_j \mathbf{X}_{\mathrm{f}}$ (Asch et al. 2016; Evensen et al. 2019), hence

$$\mathbf{Y}_j = \left(h(\mathbf{E}_j) - \bar{\mathbf{y}} \mathbf{1}^{\mathrm{T}} \right) \mathbf{T}_j^{-1} / (n-1)^{1/2} \, .$$

A pseudocode description of the method is presented in Algorithm 1. It is a quite compact algorithm that does not add significant complexity to e.g. an ETKF method.

The alternative to the preconditioned iterative ensemble is the so-called bundle variant of IEnKS. This uses a down-scaling of the iterative ensemble by a constant factor $\varepsilon \ll 1$ (that incorporates $\sqrt{n-1}$) as

$$\mathbf{E}_j = \mathbf{x}_j \mathbf{1}^{\mathrm{T}} + \varepsilon \mathbf{X}_{\mathrm{f}} \, ,$$

followed by an up-scaling of the observation anomalies

$$\mathbf{Y}_j = \left(h(\mathbf{E}_j) - \bar{\mathbf{y}} \mathbf{1}^{\mathrm{T}} \right) / \varepsilon \, .$$

Algorithm 1 Iterative Ensemble Kalman Smoother for kth DAW.

Require: Prior ensemble $\mathbf{E}^{\mathrm{f}} = \mathbf{E}_k^{\mathrm{f}}$
$j = 0$, $\mathbf{w}_j = \mathbf{0}$, $\mathbf{T}_j = \mathbf{I}_n$
$\bar{\mathbf{x}}^{\mathrm{f}} = \mathbf{E}^{\mathrm{f}}\mathbf{1}/n$
$\mathbf{X}_{\mathrm{f}} = \left(\mathbf{E}^{\mathrm{f}} - \bar{\mathbf{x}}^{\mathrm{f}}\mathbf{1}^{\mathrm{T}}\right)/\sqrt{n-1}$
repeat
$\mathbf{x}_j = \bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w}_j$
$\mathbf{E}_j = \mathbf{x}_j\mathbf{1}^{\mathrm{T}} + \sqrt{n-1}\mathbf{X}_{\mathrm{f}}\mathbf{T}_j$
$\bar{\mathbf{y}} = h_k(\mathbf{E}_j)\mathbf{1}/n$
$\mathbf{Y} = \left[h_k(\mathbf{E}_j) - \bar{\mathbf{y}}\mathbf{1}^{\mathrm{T}}\right]\mathbf{T}_j^{-1}/\sqrt{n-1}$
$\nabla J = \mathbf{w}_j - \mathbf{Y}^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathbf{y}_k^{\mathrm{o}} - \bar{\mathbf{y}}\right)$
$\mathbb{H} = \mathbf{I} + \mathbf{Y}^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{Y}$
$\mathbf{w}_{j+1} = \mathbf{w}_j - \mathbb{H}^{-1}\nabla J$
$\mathbf{T}_{j+1} = \mathbb{H}^{-1/2}$
$j = j + 1$
until termination criteria met
$\mathbf{E}_k^{\mathrm{a}} = \bar{\mathbf{x}}^{\mathrm{f}}\mathbf{1}^{\mathrm{T}} + \mathbf{X}_{\mathrm{f}}\left(\mathbf{w}_{j-1}\mathbf{1}^{\mathrm{T}} + \sqrt{n-1}\mathbf{T}_j\right)$

The reason for not using this variant, besides the question of which value to choose for this scaling factor, was that this approach does not reduce the variance of the parameters already estimated during iterations. So within an observed region, the upper part would have its mean estimated but the observation ensemble would include variability from parameter in this upper part. This makes the cross-covariances between parameters further down and the waveform data less distinct, and counterproductive for the gradient estimation.

8 Elastic parameter estimation using seismic waveform data

Inverting for elastic properties using the ensemble Kalman framework poses some interesting challenges. It is the finite sample approximations of covariance $\mathrm{cov}\left[\mathbf{y}_k\right]$ and cross-covariance $\mathrm{cov}\left[\mathbf{x}, \mathbf{y}_k\right]$ matrices that forms the Kalman gain, and these must be fairly representative for the estimation to succeed. Kalman filtering is known to be sensitive to the degree of suboptimality, which can lead to filter divergence. Sources of suboptimality are essentially any deviation from the full rank Gauss-linear case, it be nonlin-

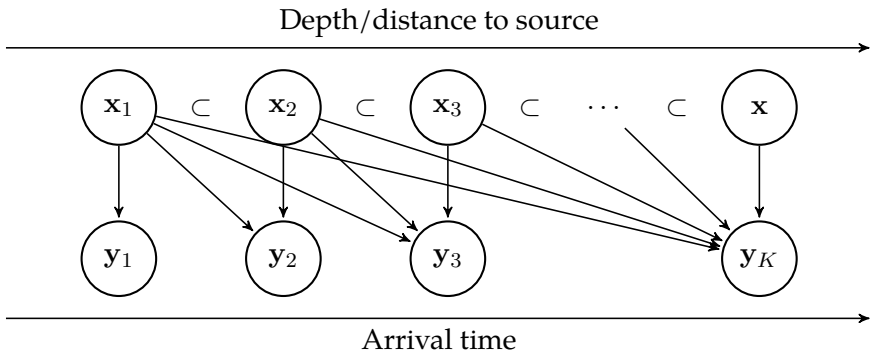


Figure 4: Schematic of relation between depth of parameters and arrival time of reflected seismic wave. Earlier reflections only relates to shallower parameters while later reflections have traveled longer.

earity, non-Gaussianity, sampling approximations, misspecification of error statistics or a biased forward model.

A central strategy in this thesis work has been the data partitioning, where gather data is blocked in time windows as $\mathbf{y}_k$, and sequentially processed in order of increasing arrival time. Figure 4 sketches the relationship between the depths of elements in the parameter state vector and arrival times of observations. The $\mathbf{x}_1, \mathbf{x}_2$ etc are depth subsets of $\mathbf{x}$ from the top and downwards and $\mathbf{x}_2$ spans deeper and $\mathbf{x}_1$. Observations in a time window, e.g. $\mathbf{y}_1$ are reflected waves from a particular depth range $\mathbf{x}_1$, but the subset $\mathbf{x}_1$ is not known precisely, just that it is a subset which makes local analysis diffult. As the correction to the state relies on cross-covariances $\mathrm{cov}\left[\mathbf{x}, \mathbf{y}_k\right]$, parameters at shallower depth must be accounted for before cross-covariances between parameter further down can be well estimated. Hence, to work the inversion from top to bottom is essential.

With respect to the predicted observation error covariance $\mathrm{cov}\left[\mathbf{y}\right] = \mathbf{Y}\mathbf{Y}^{\mathrm{T}} + \mathbf{R}$, the sample estimate $\mathbf{Y}\mathbf{Y}^{\mathrm{T}}$ will have rank $\min(n-1, p_k)$ with p_k number of data points in kth DAW. When the observation dimension p_k is larger than ensemble size $p_k \gg n$, this constitutes a reduction in total variance the sample covariance matrix is capable of representing. The total variance of the true, full covariance would equal the sum of its p_k eigenvalues. If these were indexed in descending magnitude, we can say that the sample approximation truncates this sum after $n-1$ terms. Hence, if the eigenvalue magnitude around the $n-1$ term is small, the remainder of the total variance truncation might be less significant and the misspecification not too bad. Vice versa, if the eigenvalue magnitude around this truncation

point is significant, the remainder can be a considerable contribution to the total variance, and the misspecification more severe.

Another aspect of the cross-covariance is that all three elastic properties are to be inferred from the same observations, the same wave amplitudes. The effect of the different elastic properties on the waveform data is not of equal scale. Where the acoustic and shear velocities are much more directly affecting the reflected waveforms, the density effect is much more subtle. Estimation of cross-covariance between density and waveforms is more prone to spurious correlations than are the velocities. By iterating, this sampling error effect is reduced as the iterations offer the possibility to re-evaluate the cross-covariances, but the ensemble still needs to be of substantial size. If there is a potential multimodal solution, the misestimation of the density gradient can cause the iterate solution to move towards the wrong local mode/minimum.

9 Future work

There are still many interesting challenges to the practical/industrial application of elastic inversion using ensemble filtering. The two subjects presented next are but a subset, albeit of the more difficult kind.

Uncertainty in prior

The spatial covariance structure that enters the prior specification, was always fixed to a specific structure in this work where the true state was known. As the analysis state is found as a linear combination within this prior ensemble, the ability to correctly represent a true subsurface profile heavily depends on this ensemble. Roughly speaking, it will not be possible to represent a highly varying profile from a basis of almost linear ensemble members. Secondly, the spatial structure might be different at different depths, i.e. nonstationary throughout the domain, which again is difficult to known anything about beforehand.

This sort of flexibility in prior structure leads towards considering hyperpriors on the spatial covariance structure, maybe even different spatial structures for elastic properties taken as independent.

While the usefulness of including multivariate non-stationary covariance in the prior model (as hyperprior) is attractive, the way to do so is less straight forward. Just augmenting the ensemble with some (several) hyperparameters would most likely not be the way to go. For adequate sampling of the hyperprior space and the (conditional to this) elastic parameter

space, the necessary ensemble size grows very quickly. Emerick (2016) considered assembling the prior ensemble with sets of samples from different, pre-chosen multivariate Gaussian distributions each with different parameterization, and got some promising results. So the question of uncertainty in prior specification is in the wind and would be valuable future work.

Model error

Model error, or model misspecification, is a serious concern. The data are waveforms and if the forward model does not simulate the data well, the error between prediction and observation will contain systematic features. This bias, that is not represented by the predicted observation error, is likely to make the filter diverge quickly. How to include this is an interesting question. Perhaps the answer is to not consider data as time domain waveform amplitudes, but instead transform these into another domain where the effect of a model bias is reduced. How to transfer the likelihood function is then the challenge.

10 Summary

First are summaries of conference presentations given in chronological order. This is to give a historical perspective on the development through the project. Then follows summaries and the background for the contributed papers.

Petroleum Geostatistics 2015, Biarritz, France (Gineste and Eidsvik 2015)

The first presentation was on acoustic velocity inversion only. While using a 2D spectral element solver (SpecFem2D) for the (acoustic) wave simulation, the computational grid elements were blocked as to form a 1.5D model (horizontally constant). The full data set was said to be a collection of shot gathers and each of these were assimilated sequentially by an EnKS method with perturbed observation. A fairly small ensemble was used, and collapse was immediate due the large observation dimension. So post-analysis inflation was used.

EAGE 2017, Paris, France
(Gineste and Eidsvik 2017)

At this point the aspect of data dimension and ensemble collapse was in focus. The presentation continued the case of acoustic velocity inversion and the use of the spectral element solver for simulating observations. The assumption of layered subsurface model was still applied although with a finer resolution than previously. The assimilation method was still EnKS with perturbed observations. The data covered 9 shots and now data partitioning started to block in time windows as well.

An approach for reducing the data dimensionality, inspired by an application in history matching (Sætrom et al. 2010), was investigated. This made use of Partial Least Squares (PLS) modeling to form the Kalman gain matrix. The PLS basis was truncated to suppress ensemble collapse and the truncation level was found using a cross-validation technique.

EAGE 2018, Copenhagen, Denmark
(Gineste, Eidsvik, and Zheng 2018)

Now, the forward model was changed to the reflectivity method, while inversion still only considered acoustic velocity. The data partitioning made used of very small windows in time- and offsets intervals that were set manually. These were processed in an order of far-to-close offsets, early-to-late arrival times. An adaptive strategy for efficiency was applied. This consisted of evaluating the analysis of each data window. If the analysis seemed negligible, the window was skipped and the ensemble evaluation re-used for next window. If the analysis showed sign of overfitting, the data window was processed using an MDA approach thereby regularizing each update of the MDA iterations.

EAGE/PESGB Workshop on Velocities 2019, London, England
(Gineste, Eidsvik, and Zheng 2019)

This presentation was on the preliminary results of Paper I. The assimilation method was the IEnKS which was applied over partitions of time windows of increasing arrival time. The time intervals was found by using two-way traveltime at zero offset, such that each time window would cover layers in a $500\,\mathrm{m}$ depth range. This time division was generated prior to any assimilation and used the prior mean as reference velocity profile.

Petroleum Geostatistics 2019, Florence, Italy
(Gineste and Eidsvik 2019)

This presentation was on the results of Paper I, combined with a preliminary version of the adaptive data assimilation window selection from Paper II.

Paper I
(Gineste, Eidsvik, and Zheng 2020)

This paper sets the framework for doing elastic inversion of seismic data with the IEnKS. The true elastic profile was contributed by BP and based on this, a synthetic gather was used as observations. The gather data was partitioned in time windows and the IEnKS used to obtain analysis ensemble. The partitioning was made previous to any assimilation and used a two-way traveltime at zero offset to partition the full depth range in blocks.

This paper also made use of the adaptive inflation provided by the finite-size EnKF-N (Bocquet 2011; Bocquet and Sakov 2012; Bocquet, Raanes, et al. 2015) formulation. The inflation was included to counteract reduction in ensemble spread over successive assimilation cycles. In retrospect, this successive deflation was more related to premature termination of the iterations, than to sampling error. The paper ends in a discussion on the DAW size and how choosing it too large leads to instability.

The main contribution of this paper is, what might be the first example of a such, the application of an deterministic, iterative ensemble smoother to the inversion of elastic properties using reflected waveform data.

Paper II

In this paper stability and efficiency are keywords, and continues the discussion of an suitable DAW size of Paper I. The problem configuration of Paper I is repeated, whereas the previous manually and preset partitioning was to be selected automatically, prior to each assimilation cycle. Imaging a departure from the stacked layer assumption (i.e. use in 2D application), the two-way traveltime approach to blocking had to be abandoned. The research question of this article is thus how the assimilation window can be chosen as large as possible while minimizing the risk of overfitting.

The paper has as focus a spectral decomposition of the initial Gauss-Newton update direction. This update direction is analyzed in the light of being composed of contributions from prior and likelihood. Arguing for balance between these contributions, two criteria to stop extending the

DAW was presented. In a repeated experiment, one approach was found to work as intended and the other not so.

This paper also introduces the use of an information theoretic measure, the mutual information measure, as a possible stopping criterion for iterating.

The main contribution of this paper is the increased usability and improved robustness of elastic inversion using IEnKS, resulting from the adaptive determination of an assimilation window length. The establishment of an evaluation criterion, determining whether a proposed data partition is appropriate, might be useful in other applications that share the same particular problem.

Paper III

This report is a tutorial on using IEnKS for seismic invesion using different types of seismic data. Three examples are used, ranging from linear traveltime over AVO to elastic waveform data. The examples span varying degree of nonlinearity thus building up a justification for the additional computational effort of iterating. This report is to be offered to industrial consortium partners, as well as other PhD students that might continue along the research lines of this thesis. It might also be used as material for course projects at NTNU. The intention is thus transfer of knowledge and exemplification of how to do iterative, ensemble-based parameter estimation. Its distribution is accompanied by a MATLAB code basis that shortens the time from modeling to implementation.

The main contribution of this report is the availability of a code base for seismic inversion examples in a common formulation.

References

Anderson, J. L. (2001). "An Ensemble Adjustment Kalman Filter for Data Assimilation". In: *Monthly Weather Review* 129.12, pp. 2884–2903.

Asch, M., M. Bocquet, and M. Nodet (2016). *Data Assimilation: Methods, Algorithms, and Applications*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Baldick, R. (2006). *Applied optimization. Formulation and algorithms for engineering systems*. Cambridge Univ. Press. 768 pp.

Bierman, G. J. (1977). *Factorization methods for discrete sequential estimation.* Vol. 128. Elsevier, Amsterdam.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar (2001). "Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects". In: *Monthly Weather Review* 129.3, pp. 420–436.

Bocquet, M. (2011). "Ensemble Kalman filtering without the intrinsic need for inflation". In: *Nonlinear Processes in Geophysics* 18.5, pp. 735–750.

Bocquet, M., P. N. Raanes, and A. Hannart (2015). "Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation". In: *Nonlinear Processes in Geophysics* 22.6, pp. 645–662.

Bocquet, M. and P. Sakov (2012). "Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems". In: *Nonlinear Processes in Geophysics* 19.3, pp. 383–399.

Bocquet, M. and P. Sakov (2013). "Joint state and parameter estimation with an iterative ensemble Kalman smoother". In: *Nonlinear Processes in Geophysics* 20.5, pp. 803–818.

Bocquet, M. and P. Sakov (2014). "An iterative ensemble Kalman smoother". In: *Quarterly Journal of the Royal Meteorological Society* 140.682, pp. 1521–1535.

Carrassi, A. et al. (2018). "Data assimilation in the geosciences: An overview of methods, issues, and perspectives". In: *WIREs Climate Change* 9.5, e535.

Chen, Y. and D. S. Oliver (2011). "Ensemble Randomized Maximum Likelihood Method as an Iterative Ensemble Smoother". In: *Mathematical Geosciences* 44.1, pp. 1–26.

Eikrem, K. S., G. Nævdal, and M. Jakobsen (2019). "Iterated extended Kalman filter method for time-lapse seismic full-waveform inversion". In: *Geophysical Prospecting* 67.2, pp. 379–394.

Emerick, A. A. and A. C. Reynolds (2013). "Ensemble smoother with multiple data assimilation". In: *Computers & Geosciences* 55, pp. 3–15.

Emerick, A. A. (2016). "Towards a hierarchical parametrization to address prior uncertainty in ensemble-based data assimilation". In: *Computational Geosciences* 20.1, pp. 35–47.

Evensen, G. (2009). *Data Assimilation*. Springer.

Evensen, G. (2018). "Analysis of iterative ensemble smoothers for solving inverse problems". In: *Computational Geosciences* 22.3, pp. 885–908.

Evensen, G. et al. (2019). "Efficient Implementation of an Iterative Ensemble Smoother for Data Assimilation and Reservoir History Matching". In: *Frontiers in Applied Mathematics and Statistics* 5, p. 47.

Fichtner, A. (2010). *Full Seismic Waveform Modelling and Inversion* (*Advances in Geophysical and Environmental Mechanics and Mathematics*). Springer.

Gineste, M. and J. Eidsvik (2015). "Framework for Seismic Inversion of Full Waveform Data Using Sequential Filtering". In: *Conference Proceedings, Petroleum Geostatistics 2015*. European Association of Geoscientists & Engineers.

Gineste, M. and J. Eidsvik (2017). "Seismic Waveform Inversion Using The Ensemble Kalman Smoother". In: *Conference Proceedings, 79th EAGE Conference and Exhibition 2017*. Vol. 2017. 1. European Association of Geoscientists & Engineers, pp. 1–5.

Gineste, M. and J. Eidsvik (2019). "Seismic Waveform Inversion of Elastic Properties Using an Iterative Ensemble Kalman Smoother". In: *Conference Proceedings, Petroleum Geostatistics 2019*. Vol. 2019. 1. European Association of Geoscientists & Engineers, pp. 1–5.

Gineste, M., J. Eidsvik, and Y. Zheng (2018). "Velocity Estimation in Layered Media Using Ensembled-based Sequential Filtering". In: *Conference Proceedings, 80th EAGE Conference and Exhibition 2018*. Vol. 2018. 1. European Association of Geoscientists & Engineers, pp. 1–5.

Gineste, M., J. Eidsvik, and Y. Zheng (2019). "Seismic Waveform Inversion Using an Iterative Ensemble Kalman Smoother". In: *Conference Proceed-*

ings, Second EAGE/PESGB Workshop on Velocities. Vol. 2019. 1. European Association of Geoscientists & Engineers, pp. 1–3.

Gineste, M., J. Eidsvik, and Y. Zheng (2020). "Ensemble-based seismic inversion for a stratified medium". In: *GEOPHYSICS* 85.1, R29–R39.

Hunt, B. R., E. Kalnay, et al. (2004). "Four-dimensional ensemble Kalman filtering". In: *Tellus A* 56.4, pp. 273–277.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh (2007). "Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter". In: *Physica D: Nonlinear Phenomena* 230.1-2, pp. 112–126.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Mathematics in Science and Engineering. New York, NY: Academic Press.

Johnson, R. A. and D. W. Wichern (2007). *Applied multivariate statistical analysis*. 6. ed. Pearson/Prentice Hall. 773 pp.

Kennett, B. L. N. (2005). *ERZSOL3*. `http://www.spice-rtn.org/library/software/ERZSOL3.html`. Accessed on 2017-09-01.

Kennett, B. L. N. (2011). *Seismic Wave Propagation in Stratified Media*. ANU Press.

Kennett, B. L. N. and N. J. Kerry (1979). "Seismic waves in a stratified half space". In: *Geophysical Journal International* 57.3, pp. 557–583.

Komatitsch, D. and J. Tromp (1999). "Introduction to the spectral element method for three-dimensional seismic wave propagation". In: *Geophysical Journal International* 139.3, pp. 806–822.

Livings, D. M., S. L. Dance, and N. K. Nichols (2008). "Unbiased ensemble square root filters". In: *Physica D: Nonlinear Phenomena* 237.8, pp. 1021–1028.

Madsen, H. (2007). *Time Series Analysis*. Chapman & Hall / CRC Texts in Statistical Science. CRC Press.

Mallick, S. and L. N. Frazer (1987). "Practical aspects of reflectivity modeling". In: *GEOPHYSICS* 52.10, pp. 1355–1364.

Müller, G. (1985). "The reflectivity method: a tutorial". In: *Journal of Geophysics* 58.1-3, pp. 153–174.

Nerger, L. et al. (2012). "A Unification of Ensemble Square Root Kalman Filters". In: *Monthly Weather Review* 140.7, pp. 2335–2345.

Petersen, K. B. and M. S. Pedersen (2012). *The Matrix Cookbook*. Version 20121115.

Pham, D. T. (1996). *A Singular Evolutive Interpolated Kalman filter for data assimilation in Oceanography*.

Pham, D. T. (2001). "Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems". In: *Monthly Weather Review* 129.5, pp. 1194–1207.

Raanes, P. N., A. S. Stordal, and G. Evensen (2019). "Revising the stochastic iterative ensemble smoother". In: *Nonlinear Processes in Geophysics* 26.3, pp. 325–338.

Sætrom, J. and H. Omre (2010). "Ensemble Kalman filtering with shrinkage regression techniques". In: *Computational Geosciences* 15.2, pp. 271–292.

Sakov, P. and L. Bertino (2011). "Relation between two common localisation methods for the EnKF". In: *Computational Geosciences* 15.2, pp. 225–237.

Sakov, P., G. Evensen, and L. Bertino (2010). "Asynchronous data assimilation with the EnKF". In: *Tellus A* 62.1, pp. 24–29.

Sakov, P. and P. R. Oke (2008). "Implications of the Form of the Ensemble Transformation in the Ensemble Square Root Filters". In: *Monthly Weather Review* 136.3, pp. 1042–1053.

Sakov, P., D. S. Oliver, and L. Bertino (2012). "An Iterative EnKF for Strongly Nonlinear Systems". In: *Monthly Weather Review* 140.6, pp. 1988–2004.

Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Institute of Mathematical Statistics Textbooks. Cambridge University Press.

Sheriff, R. E. and L. P. Geldart (1995). *Exploration Seismology*. 2nd ed. Cambridge University Press.

Själander, M. et al. (2019). "EPIC: An Energy-Efficient, High-Performance GPGPU Computing Research Infrastructure". In: *arXiv*.

Tarantola, A. (1984). "Inversion of seismic reflection data in the acoustic approximation". In: *GEOPHYSICS* 49.8, pp. 1259–1266.

Thurin, J., R. Brossier, and L. Métivier (2019). "Ensemble-based uncertainty estimation in full waveform inversion". In: *Geophysical Journal International* 219.3, pp. 1613–1635.

Tippett, M. K. et al. (2003). "Ensemble Square Root Filters". In: *Monthly Weather Review* 131.7, pp. 1485–1490.

Virieux, J., A. Asnaashari, et al. (2017). "An introduction to full waveform inversion". In: *Encyclopedia of Exploration Geophysics*. Geophysical Reference Series. SEG. Chap. 6, R1-1-R1–40.

Virieux, J. and S. Operto (2009). "An overview of full-waveform inversion in exploration geophysics". In: *GEOPHYSICS* 74.6, WCC1–WCC26.

Vogel, C. R. (2002). *Computational Methods for Inverse Problems*. Society for Industrial and Applied Mathematics.

Wang, X., C. H. Bishop, and S. J. Julier (2004). "Which Is Better, an Ensemble of PositiveNegative Pairs or a Centered Spherical Simplex Ensemble?" In: *Monthly Weather Review* 132.7, pp. 1590–1605.

Wikle, C. K. and L. M. Berliner (2007). "A Bayesian tutorial for data assimilation". In: *Physica D: Nonlinear Phenomena* 230.1-2, pp. 1–16.

Zhang, X. and A. Curtis (2020). "Variational full-waveform inversion". In: *Geophysical Journal International* 222.1, pp. 406–411.

Paper I

Ensemble-based seismic inversion for a stratified medium

Michael Gineste, Jo Eidsvik & York Zheng

Published in
GEOPHYSICS, 2020, Vol. 85, No. 1, 29-39
doi: 10.1190/GEO2019-0017.1

GEOPHYSICS, VOL. 85, NO. 1 (JANUARY-FEBRUARY 2020); P. R29–R39, 9 FIGS., 2 TABLES.
10.1190/GEO2019-0017.1

Ensemble-based seismic inversion for a stratified medium

Michael Gineste[1], Jo Eidsvik[1], and York Zheng[2]

ABSTRACT

Seismic waveform inversion is a nontrivial optimization task, which is often complicated by the nonlinear relationship between the elastic attributes of interest and the large amount of data obtained in seismic experiments. Quantifying the solution uncertainty can be even more challenging, and it requires considering the problem in a probabilistic setting. Consequently, the seismic inverse problem is placed in a Bayesian framework, using a sequential filtering approach to invert for the elastic parameters. The method uses an iterative ensemble smoother to estimate the subsurface parameters, and from the ensemble, a notion of estimation uncertainty is readily available. The ensemble implicitly linearizes the relation between the parameters and the observed waveform data; hence, it requires no tangent linear model. The approach is based on sequential conditioning on partitions of the whole data record (1) to regularize the inversion path and effectively drive the estimation process in a top-down manner and (2) to circumvent a consequence of the ensemble reduced rank approximation. The method is exemplified on a synthetic case, inverting for elastic parameters in a 1D medium using a seismic shot record. Our results indicate that the iterative ensemble method is applicable to seismic waveform inversion and that the ensemble representation indeed indicates estimation uncertainty.

INTRODUCTION

Seismic waveform inversion aims at obtaining a parameterized subsurface model of elastic attributes which, evaluated with a forward model representing the physics of seismic wave propagation, matches measured data to a satisfactory degree. In most cases, this inversion is treated as an optimization problem, in which the goal is to find the best subsurface model according to some matching criterion. Full-waveform inversion, see, e.g., Virieux and Operto (2009) and Fichtner (2010), requires minimum processing of the acquired seismic data. Other forms of seismic waveform inversion start the inversion on somewhat processed data, such as common-midpoint (CMP) gathers and angle stacks (Sheriff and Geldart, 1995).

One key challenge of waveform inversion is to guide the iterative search for the optimum subsurface parameters. In most situations, this relies on derivatives of the forward model obtained via an adjoint-based approach (Plessix, 2006). It appears useful to start by bringing parts of the model in place, such as low-frequency trends, to obtain a stable optimization (Bunks et al., 1995). One can also consider the inversion in a probabilistic setting.

Bayesian inversion aims at obtaining the posterior distribution of the subsurface parameters, conditioned to observed seismic waveform data. A Gaussian approximation of the posterior distribution is defined by a mean at the posterior mode and a covariance matrix defined by the inverse negative Hessian of the log objective function, evaluated at the posterior mode (Gouveia and Scales, 1998). This approach is not considered here because (1) it restricts solutions to Gaussian approximations and (2) it requires derivative calculations that are not always available in black-box forward models. Instead, posterior assessment using Monte Carlo sampling (Eidsvik et al., 2004; Bosch et al., 2007) appears to be more suitable for this kind of nonlinear problem. But the seismic forward model evaluation is computationally costly, and this limits the choice of sampling methods.

In this paper, an ensemble-based approach for seismic waveform inversion is presented. An ensemble consists of multiple realizations of a subsurface model and represents a sample set from a probability distribution. The ensemble representation allows and calls for parallel evaluation of the forward model, so this type of method leverages the availability of distributed computing resources. The inversion is performed by conditioning this ensemble to data, thereby obtaining a conditional ensemble representation. Notably, the suggested approach assimilates subsets of seismic data in incremental temporal intervals, which appears to stabilize the inversion. Rather than obtaining derivative information explicitly, it is extracted from the

Manuscript received by the Editor 22 January 2019; revised manuscript received 20 September 2019; published ahead of production 15 October 2019; published online 22 November 2019.
[1]Norwegian University of Science and Technology, Department of Mathematical Sciences, Trondheim, Norway. E-mail: michael.gineste@ntnu.no; jo.eidsvik@ntnu.no.
[2]BP Exploration Operating Company Ltd, Sunbury-on-Thames, UK. E-mail: york.zheng@bp.com.

covariances in the ensemble; hence, no adjoint model is required. The ensemble-based approach relies on correlations between subsurface model parameters and synthetic seismic data and finding a tangent linear model at each assimilation step using the ensemble. The method is demonstrated on a synthetic seismic record using the reflectivity method (Kennett and Kerry, 1979; Kennett, 2011) as the forward solver.

The new approach outlined in this paper resembles ensemble Kalman filtering techniques, in particular the iterative ensemble smoother (see Chen and Oliver, 2011; Evensen, 2018), which has shown very good results in several application domains. Whereas the usual filtering methods update a dynamic state vector from observations, the problem considered here has a static state. It is a parameter estimation problem much like that of extracting static reservoir properties from well production data, known as the history matching problem (for a review, see Aanonsen et al., 2009; Oliver and Chen, 2010). The problem of seismic inversion and history matching has similarities such as the forward model being a black box, which requires a restart at parameter updates. Another similarity is the potential multimodality of the posterior solution. Ensemble Kalman filtering by itself does not handle multimodality well, when the ensemble-based tangent linear model approximation is used (Evensen, 2018). When the posterior solution is sought and explored in terms of multimodality, the use of a Gaussian mixture models has been investigated as a means to accommodate the effect of nonlinearity (Stordal, 2014) or as a prior specification (Dovera and Della Rossa, 2011), and recently also more "model-free" approaches to multimodality have been presented (e.g., Zhang et al., 2018). Here, the multimodal solution is not studied, and with the use of Gaussian prior and ensemble linearization, the posterior solution will be unimodal. An immediate dissimilarity between history matching and seismic waveform inversion is the high data dimensionality and the complex wave propagation phenomena in the latter, and because of this, a straightforward implementation of the ensemble smoother will not work well. There is currently much interest in developing ensemble-based approaches for reliable seismic inversion (Gineste and Eidsvik, 2015; Thurin et al., 2017, 2019; Liu and Grana, 2018). The main contribution of the current paper is to investigate the applicability of an iterative ensemble-based smoother approach for elastic inversion of seismic data in a situation of stratified layers. Results demonstrate that the suggested method is useful for optimization and uncertainty assessment in seismic waveform inversion.

The paper is structured as follows: First, inverse problems are discussed in a probabilistic setting, followed by sequential methods for inversion, using only parts of the data at every conditioning step. Then, the iterative ensemble Kalman smoother (IEnKS) is presented in the context of seismic waveform inversion. This is applied on an example with synthetic seismic data generated from a well log, followed by a discussion of the sensitivities.

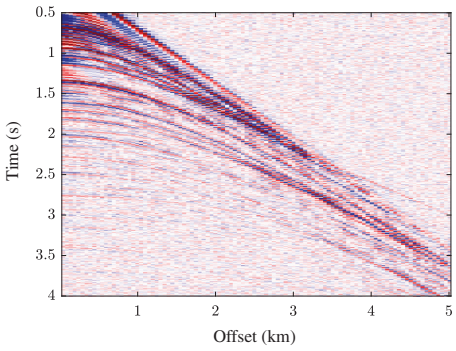


Figure 1. CMP gather of vertical displacement.

BAYESIAN INVERSION

A characteristic of seismic waveform inversion, and many similar problems in the geosciences, is the tremendous difference between forward modeling and inversion. The forward model is defined by geophysical modeling; it can be evaluated and understood. The inverse problem is much more difficult because there is generally no unique solution for the subsurface parameters as a function of the seismic waveform data (Tarantola, 2005). An example of prestack seismic data is provided by a gather of vertical displacement recordings over a range of time and receiver offsets (Figure 1). The wavefield is initiated by a known source, and the measurements contain the direct wave and the weaker reflected wavefield, caused by material inhomogeneity met by the wave when propagating through the subsurface. The latter holds the information from which the subsurface parameters are estimated. The forward model generating the synthetic seismic data is the reflectivity method, as implemented in the program/solver ERZSOL3 (Kennett, 2005). Under the assumption of a locally 1D subsurface with stratified, isotropic elastic properties, the full elastic wave equation can be transformed into an ordinary differential equation in the slowness-frequency domain. In this domain, an analytic solution is possible, involving the response of a stack of uniform layers that can be iteratively computed, and the solution is subsequently mapped to time-domain seismograms via the inverse Fourier transform. Therefore, synthetic seismic data are generated for the entire time length at once because there is no explicit time integration, whereas the number of layers and ranges for frequency- and slowness-domain integrations directly affects the program runtime.

Inversion of seismic waveform data is studied here under the assumption of (1) a stratified subsurface with laterally invariant properties represented by a parameter vector $\mathbf{x} \in \mathbb{R}^m$ with m, the number of layers, (2) a deterministic wave propagation forward model $h(\mathbf{x})$ that includes all initial and boundary conditions, as well as the data recording configuration, and (3) the observed data $\mathbf{y}$ are noisy observations of $h(\mathbf{x})$, contaminated by additive noise $\mathbf{e}$, so that $\mathbf{y} = h(\mathbf{x}) + \mathbf{e}$. Table 1 summarizes the notation used throughout this paper.

The most common solutions to such inverse problems are based on regularization of the model parameters and multiple evaluations of the forward model, where the mismatch with data is minimized. These approaches can be formalized in a coherent manner using the Bayesian framework. With a prior probability density $p(\mathbf{x})$ imposed on the subsurface parameterization along with a prescribed likelihood model $p(\mathbf{y}|\mathbf{x})$ involving the seismic forward model, Bayes' rule gives the sought solution in form of the posterior distribution $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$.

With Gaussian distributions for the model parameters and the observation error — $p(\mathbf{x}) = \mathrm{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ and $p(\mathbf{e}) = \mathrm{N}(\mathbf{0}, \mathbf{R})$ — the posterior distribution is

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}(\|\mathbf{y}-h(\mathbf{x})\|^2_{\mathbf{R}} + \|\mathbf{x}-\boldsymbol{\mu}_x\|^2_{\boldsymbol{\Sigma}_x})\right), \quad (1)$$

with the notation $\|\mathbf{a}\|^2_{\mathbf{B}} = \mathbf{a}^T\mathbf{B}^{-1}\mathbf{a}$. The maximum a posteriori (MAP) solution can be found by minimizing the negative log posterior, which has the weighted nonlinear least-squares cost/misfit form:

$$J(\mathbf{x}) = \frac{1}{2}\|\mathbf{y}-h(\mathbf{x})\|^2_{\mathbf{R}} + \frac{1}{2}\|\mathbf{x}-\boldsymbol{\mu}_x\|^2_{\boldsymbol{\Sigma}_x}. \quad (2)$$

The first term is the mismatch between the synthetic seismic data and the observations, whereas the second term contains the prior information regularizing the model solution to be not too distant from the prior mean $\boldsymbol{\mu}_x$ and with smoothness indicated by the prior covariance matrix $\boldsymbol{\Sigma}_x$. The prior and likelihood terms are possible to evaluate, but with the nonlinearity in the forward model, the posterior is not analytically tractable. Instead, one must resort to sampling-based methods to represent the posterior distribution. But, the forward model is often computationally costly, and the choice of Monte Carlo methods for seismic inversion must be realistic in the number of possible forward evaluations. Exact sampling based on Markov chain Monte Carlo (MCMC) sampling is usually impractical because one cannot evaluate the forward model as many times as is required. Approximate ensemble-based sampling is a practical solution, in which the ensemble represents possible posterior solutions, including the main trends and variability.

Below, a combination of sampling and iterative optimization is used to maintain the ensemble representation. To motivate this setting, the main idea is presented here. First, consider a linear forward model $h(\mathbf{x}) \rightarrow \mathbf{H}\mathbf{x}$, then the posterior distribution in equation 1 is Gaussian. One way of sampling from this posterior involves minimizing the quadratic cost function in equation 2 for one prior and one data residual sample. By substituting $\boldsymbol{\mu}_x$ in equation 2 with a prior sample $\mathbf{x}^f \sim p(\mathbf{x})$ and adding a sample from the noise distribution $\mathbf{e}_i \sim p(\mathbf{e})$ such that $\mathbf{y} \rightarrow \mathbf{y} + \mathbf{e}_i$, the minimization of the (stochastic) cost function $J(\mathbf{x})$ gives an exact sample from the posterior distribution. This is referred to as the randomized maximum likelihood (RML); see Gu and Oliver (2007) and Chen and Oliver (2011, 2013) for ensemble variants. A similar approach based on stochastic simulation and optimization can be used for approximate sampling in nonlinear problems.

SEQUENTIAL BAYESIAN INVERSION

Seismic waveform inversion for elastic parameters is a considerably nonlinear problem. The posterior surface is difficult to explore, and there can be many local posterior optima for the parameters $\mathbf{x}$. To perform robust inversion, a key component of the solution presented here is to partition the seismic data record $\mathbf{y}$ into K disjoint subsets $\mathbf{y}_k$ so that $\mathbf{y} = \cup_{k=1}^{K} \mathbf{y}_k$. These data subsets are then assimilated sequentially as

$$p(\mathbf{x}) \rightarrow p(\mathbf{x}|\mathbf{y}_1) \rightarrow p(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2) \rightarrow p(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3) \rightarrow \cdots \rightarrow p(\mathbf{x}|\mathbf{y}_1, \ldots, \mathbf{y}_K). \quad (3)$$

In the Bayesian framework, this means that the posterior from one assimilation cycle serves as prior for the next. For the situation with a seismic record, there is a natural geometric blocking structure as earlier data are reflected by shallower layers only, whereas later data are reflected from deeper depths. This argument leads to sequential conditioning on data of increasing arrival times to reduce the subset of parameters causing the observed data and thus enhancing their identifiability. This results in inverting for subsurface parameters in a top-down manner and the goal is that this divide-and-conquer strategy alleviates the nonlinearity of the problem. For the data example shown later, regular time windowing is used as data assimilation windows (DAWs).

Sequential conditioning to data is facilitated by embedding the parameters in a state space model as is commonly done in filtering problems. In the current situation, the order or index k indicates assimilation cycles over which the data partition $\mathbf{y}_k$ is used for model conditioning. The state vector holds the static parameters $\mathbf{x}$. Initiating with $p(\mathbf{x}_0) = p(\mathbf{x}) = \mathrm{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$, then for $k = 1, \ldots, K$, the system and observation equations become

$$\mathbf{x}_k = \mathbf{x}_{k-1}, \quad (4a)$$

$$\mathbf{y}_k = h_k(\mathbf{x}_k) + \mathbf{e}_k, \quad (4b)$$

Table 1. Summary of notations.

	Notations related to modeling aspects
$\mathbf{x}$	Subsurface model parameter, length m vector
$N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$	Gaussian density with mean $\boldsymbol{\mu}_x$ and covariance $\boldsymbol{\Sigma}_x$
$\mathbf{y}$	Seismic data
$h(\mathbf{x})$	Seismic forward model
$\mathbf{R}$	Covariance matrix of the seismic measurement noise
$p(\mathbf{y}\|\mathbf{x})$	Likelihood model of seismic data, given model
$p(\mathbf{x}\|\mathbf{y})$	Posterior model, given seismic data
	Notations related to blocking of seismic data
$\mathbf{y}_k$	Seismic data in block $k = 1, \ldots, K$
$h_k(\mathbf{x})$	Seismic forward model for data in block k
$\mathbf{R}_k$	Covariance matrix of seismic measurement noise in the block
$p(\mathbf{x}\|\mathbf{y}_1, \ldots, \mathbf{y}_k)$	Posterior model given seismic data up to block $k = 1, \ldots, K$
	Notations related to ensemble-based updating
$\mathbf{E} = \{\mathbf{x}_{[i]}\}_{i=1}^{n}$	Parameter ensemble of size $m \times n$
$\bar{\mathbf{x}}$	Ensemble mean of the model parameters
$\mathbf{X}_\mathrm{f}$	Forecast model anomalies around the mean $\bar{\mathbf{x}}^\mathrm{f}$, size $m \times n$ matrix
$\mathbf{Y}_\mathrm{f}$	Forecast seismic data anomalies around the mean, size $p \times n$ matrix
$\mathbf{w}$	Weight of the linear combination of model ensembles
$J(\mathbf{w})$	Objective function, negative log-posterior
∇J	Ensemble-based gradient of the objective function
$\mathcal{H}$	Ensemble-based Hessian of the objective function

where the forward observation operator $h_k(\mathbf{x})$ forecasts (simulates) a seismic record and extracts the kth subset of the data. The observation errors $\mathbf{e}_k$ are assumed to be independent zero-mean Gaussian variables with covariance matrix $\mathbf{R}_k$.

For the ensemble-based methods of sequential data assimilation, the following sub- and superscript notation is used: "f" indicates forecast (also known as prior or background) and "a" indicates analysis (i.e., posterior or update). Ignoring index k, this means that $\mathbf{E}^{\mathrm{f}} = \{\mathbf{x}^{\mathrm{f}}_{[i]}\}_{i=1}^{n}$ is an ensemble of n forecast states, where the bracketed index indicates the ith ensemble member and $\mathbf{E}^{a} = \{\mathbf{x}^{\mathrm{a}}_{[i]}\}_{i=1}^{n}$ is the ensemble of analysis states. At every assimilation step, the forecast ensemble is updated to the analysis ensemble. Applying the static state transition equation for $\mathbf{x}_k$ (equation 4a), the analysis ensemble becomes the forecast ensemble for the next assimilation step. The procedure is run through all of the K cycles.

The actual assimilation is achieved here by an approximate Monte Carlo approach similar to RML as mentioned in the previous section. This means that the forecast elastic parameters and associated forecast seismic data are combined in an objective function that is optimized. The solution to this defines the analysis ensemble.

ITERATIVE ENSEMBLE KALMAN SMOOTHER

The data assimilation method applied here is the IEnKS; see Bocquet and Sakov (2013, 2014) and Asch et al. (2016). Developed in the atmospheric sciences for the purpose of estimating dynamic states, it is here adapted to the case of static parameters in seismic inversion. The IEnKS merges the ensemble and variational approaches: (1) it iteratively solves the variational problem to obtain a good fit to observations over the DAW, and (2) it uses the ensemble to approximate the tangent linear observation operator $\partial_{\mathbf{x}} h(\mathbf{x})$ and obtains sensitivities for the variational formulation. The method is very suitable for cases in which the forward model/simulator is a black box.

The IEnKS uses a square-root filter (SRF)-type approach, separately updating the state mean and square-root of the covariance matrix. Moreover, it finds the Kalman gain in the ensemble subspace, which is computationally efficient when the ensemble space dimension is smaller than the observation space dimension. A requirement of the method is that the observation error covariance $\mathbf{R}$ is easily invertible, which is the case with independent observation errors.

The ensemble-based Kalman filtering framework has several causes for suboptimality, with optimality only in principle for Gauss linear models and an infinite ensemble size. The ensemble estimates of the first- and second-order moments replace the true but unknown counterparts from the prior distribution and assume that these ensemble moments are correct. This sampling error often causes the underestimation of the error covariance, and over successive analysis steps, this can lead to collapse of the ensemble, known as filter divergence. This negative consequence of a limited ensemble size is often practically handled through the use of inflation, in which the ensemble spread is artificially increased, formally as $\mathbf{x}_{[i]} \to \bar{\mathbf{x}} + \beta(\mathbf{x}_{[i]} - \bar{\mathbf{x}})$ with inflation factor $\beta > 1$. The finite-size ensemble approach derived by Bocquet (2011) and Bocquet et al. (2015) offers adaptive inflation to address this issue of sampling error. Without going into the details of its underlying theory, this is used here within the IEnKS, without any special consideration (Bocquet and Sakov, 2012).

A description of the components used in the IEnKS analysis step is now presented, followed by a description of the variational problem. A pseudocode describing the implementation of this approach is summarized in Algorithm 1, which is reproduced from Bocquet and Sakov (2012) and Asch et al. (2016) with some modifications for this particular use.

In the description, index k is dropped because the focus is on a single analysis cycle. It is implied that $\mathbf{E}^{\mathrm{f}} = \mathbf{E}^{\mathrm{f}}_k = \{\mathbf{x}^{\mathrm{f}}_{[i],k}\}_{i=1}^{n}$ consists of realizations from the predictive probability density $p(\mathbf{x}|\mathbf{y}_1, \ldots \mathbf{y}_{k-1})$ of elastic parameters. The sample mean and covariance of this forecast distribution are

$$\bar{\mathbf{x}}^{\mathrm{f}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}^{\mathrm{f}}_{[i]}, \quad \mathbf{P}^{\mathrm{f}} = \mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{T}, \tag{5}$$

where the $m \times n$ forecast anomaly (or perturbation) matrix $\mathbf{X}_{\mathrm{f}}$ has an ith column:

$$[\mathbf{X}_{\mathrm{f}}]_i = (n-1)^{-1/2}(\mathbf{x}^{\mathrm{f}}_{[i]} - \bar{\mathbf{x}}^{\mathrm{f}}). \tag{6}$$

This anomaly matrix $\mathbf{X}_{\mathrm{f}}$ is accordingly a square root of the forecast covariance matrix $\mathbf{P}_{\mathrm{f}}$. Similarly, a square root of the forecast covariance in observation space is defined as the $p \times n$ observation anomaly matrix $\mathbf{Y}_{\mathrm{f}}$:

$$[\mathbf{Y}_{\mathrm{f}}]_i = (n-1)^{-1/2}(\mathbf{y}^{\mathrm{f}}_{[i]} - \bar{\mathbf{y}}^{\mathrm{f}}), \tag{7}$$

where the size p synthetic seismic data are $\mathbf{y}^{\mathrm{f}}_{[i]} = h(\mathbf{x}^{\mathrm{f}}_{[i]})$, $i = 1, \ldots, n$ and $\bar{\mathbf{y}}^{\mathrm{f}} = 1/n \sum_{i=1}^{n} \mathbf{y}^{\mathrm{f}}_{[i]}$. So, the forecast observation anomalies $\mathbf{Y}_{\mathrm{f}}$ are deviations in waveform data predictions and represent a square root of the forecast observation covariance.

Algorithm 1. Iterative ensemble Kalman smoother.

Require: Prior ensemble $\mathbf{E}_0 = \mathbf{E}^{\mathrm{f}}_k$; DAW index k; algorithm parameters: $\varepsilon_J, j_{\max}$

$j = 0$, $\mathbf{w} = \mathbf{0}$, $\mathbf{T} = \mathbf{I}_n$

$\mathbf{x}_0 = \mathbf{E}_0\mathbf{1}/n$ ▷ prior mean

$\mathbf{X}_0 = (\mathbf{E}_0 - \mathbf{x}_0\mathbf{1}^T)/\sqrt{n-1}$ ▷ prior anomalies

repeat ▷ variational iteration

  $\mathbf{x} = \mathbf{x}_0 + \mathbf{X}_0\mathbf{w}$ ▷ iterate mean

  $\mathbf{E} = \mathbf{x}\mathbf{1}^T + \sqrt{n-1}\mathbf{X}_0\mathbf{T}$ ▷ iterate ensemble for sensitivities

  $\bar{\mathbf{y}} = h_k(\mathbf{E})\mathbf{1}/n$ ▷ observation mean

  $\mathbf{Y} = (h_k(\mathbf{E}) - \bar{\mathbf{y}}\mathbf{1}^T)\mathbf{T}^{-1}/\sqrt{n-1}$ ▷ observation anomalies

  $\nabla J = \frac{(n+1)\mathbf{w}}{n-1/n+\mathbf{w}^T\mathbf{w}} - \mathbf{Y}^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \bar{\mathbf{y}})$ ▷ gradient

  $\mathcal{H} = \frac{(n+1)((n-1/n+\mathbf{w}^T\mathbf{w})\mathbf{I}_n - 2\mathbf{w}\mathbf{w}^T)}{(n-1/n+\mathbf{w}^T\mathbf{w})^2} + \mathbf{Y}^T\mathbf{R}_k^{-1}\mathbf{Y}$ ▷ Hessian

  $\mathbf{w} := \mathbf{w} - \mathcal{H}^{-1}\nabla J$ ▷ Gauss-Newton step

  $\mathbf{T} = \mathcal{H}^{-\frac{1}{2}}$ ▷ iterate ETM

  $j := j + 1$

until: $(S_j < \varepsilon_J) \vee (j > j_{\max})$

$\mathbf{E}^{a}_k = \mathbf{x}_0\mathbf{1}^T + \mathbf{X}_0\mathbf{w}^* + \sqrt{n-1}\mathbf{X}_0\mathbf{T}^*$ ▷ analysis, equation 7

The SRF approach seeks to map the forecast square-root covariance to the analysis square-root covariance by using the Kalman equation for the covariance update. Thus, a transform $\mathbf{X}_\mathrm{f} \mapsto \mathbf{X}_\mathrm{a}$ is the goal with $\mathbf{X}_\mathrm{a}$ defined similarly as in equation 6. The Kalman covariance update equation with the sample (ensemble) covariance matrices inserted can, using the matrix inversion lemma, be formulated as

$$\mathbf{X}_\mathrm{a}\mathbf{X}_\mathrm{a}^T = \mathbf{X}_\mathrm{f}\left(\beta^{-1}\,\mathbf{I} + \mathbf{Y}_\mathrm{f}^T\mathbf{R}^{-1}\mathbf{Y}_\mathrm{f}\right)^{-1}\mathbf{X}_\mathrm{f}^T = \mathbf{X}_\mathrm{f}\mathbf{T}(\mathbf{X}_\mathrm{f}\mathbf{T})^T, \quad (8)$$

so by selecting the ensemble transform matrix (ETM) as $\mathbf{T} = (\mathbf{I} + \mathbf{Y}_\mathrm{f}^T\mathbf{R}^{-1}\mathbf{Y}_\mathrm{f})^{-1/2}$, the analysis square-root covariance can be updated as $\mathbf{X}_\mathrm{a} = \mathbf{X}_\mathrm{f}\mathbf{T}$. This approach is not unique, and SRF variants differ in how this transformation is found. The ETM being a matrix square root is not uniquely defined either, but here the version referred to as the symmetric square root is used (Sakov and Oke, 2008; Raanes et al., 2015). Furthermore, the multiplicative inflation factor β (Hunt et al., 2007) establishes a link to the adaptive inflation provided by the finite-size ensemble scheme.

The low-rank property of ensemble-based variants causes the analysis state to be a linear combination within the span of the forecast ensemble. This is stated explicitly by introducing a weight vector $\mathbf{w}$ so that the analysis mode is $\bar{\mathbf{x}}^\mathrm{a} \in \{\bar{\mathbf{x}}^\mathrm{f} + \mathbf{X}_\mathrm{f}\mathbf{w} | \mathbf{w} \in \mathbb{R}^n\}$. By finding the optimal analysis weight $\mathbf{w}^*$ and the corresponding $\mathbf{T}$ (i.e., evaluated at $\mathbf{w}^*$), the ensemble mean and perturbation matrix are updated separately as $\bar{\mathbf{x}}^\mathrm{a} = \bar{\mathbf{x}}^\mathrm{f} + \mathbf{X}_\mathrm{f}\mathbf{w}^*$ and $\mathbf{X}_\mathrm{a} = \mathbf{X}_\mathrm{f}\mathbf{T}$, and the analysis ensemble of the model parameters is

$$\mathbf{E}^\mathrm{a} = \bar{\mathbf{x}}^\mathrm{a}\mathbf{1}^T + \mathbf{X}_\mathrm{a} = \bar{\mathbf{x}}^\mathrm{f}\mathbf{1}^T + \mathbf{X}_\mathrm{f}(\mathbf{w}^*\mathbf{1}^T + (n-1)^{\frac{1}{2}}\mathbf{T}). \quad (9)$$

Now we consider the variational part of the methodology that specifies the optimal weight vector $\mathbf{w}$, i.e., the coordinate vector within the ensemble subspace, for seismic data in the kth DAW. When considering the cost function, equivalent to equation 2 in this subspace, the prior term becomes $1/2\|\mathbf{w}\|^2$ (Hunt et al., 2007). The change in prior for parameters $\mathbf{x}$ from the finite-size formulation can be propagated into the coefficient space $\mathbf{w}$ so that the cost function becomes

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{y}_k - h_k(\bar{\mathbf{x}}^\mathrm{f} + \mathbf{X}_\mathrm{f}\mathbf{w})\|^2_{\hat{\mathbf{R}}_k} + \frac{n+1}{2}\ln\left(1 + \frac{1}{n} + \frac{\|\mathbf{w}\|^2}{n-1}\right). \quad (10)$$

This cost function is minimized using an iterative method. The gradient ∇J and approximative Hessian $\mathcal{H}$ (lines 9 and 10 in Algorithm 1, respectively) use the simulated observation anomalies $\mathbf{Y}_\mathrm{f}$ in the terms originating from the seismic data misfit part of the objective function. The iterative scheme $\mathbf{w}^{(j+1)} = \mathbf{w}^{(j)} - \mathcal{H}^{-1}\nabla J$ is of a Gauss-Newton type. The iteration is stopped when either the relative change of a function is sufficiently small $S_j = |f_j - f_{j-1}|/|f_{j-1})| < \varepsilon_J$ or if a termination condition on the maximum number of iterations $j > j_\mathrm{max}$ is reached. The function f_j makes use of the objective function evaluation, and a standard choice would be $f_j = J(\mathbf{w}^{(j)})$. However, in our experience, this could lead to premature termination, especially at later assimilation cycles in which noise starts dominating the residuals and the objective value decrease during iterations is more moderate. Therefore, a weighting functional is used instead as $f_j = \sum_{m=0}^{2} J(\mathbf{w}^{(j-m)})$ to make the objective value level out before ending.

At each iteration j, an ensemble $\mathbf{E}^{(j)}$, based on equation 9 with iterates $\mathbf{w}^{(j)}$ and $\mathbf{T}^{(j)}$ initialized as $\mathbf{w}^{(0)} = \mathbf{0}$ and $\mathbf{T}^{(0)} = \mathbf{I}$, is evaluated to obtain the synthetic seismic observation anomalies $\mathbf{Y}_\mathrm{f}^{(j)} \equiv \mathbf{Y}_\mathrm{f}^{(j)}(\mathbf{T}^{(j)})^{-1}$. The ETM $\mathbf{T}^{(j)}$ involves only the observation anomalies via the underlying linearization $\mathbf{Y}_\mathrm{f} \equiv \mathbf{H}\mathbf{X}_\mathrm{f}$, with $\mathbf{H}$ being the tangent linear model $\partial h_k/\partial \mathbf{x}$ of the observation operator h_k. This loop is depicted in Figure 2. When an analysis step is initiated, this linearization might not be very good, but it improves significantly over a few iterations; thus, the effect of iterating not only improves the mean but also the Hessian approximation and thereby the ETM.

The quality of the observed tangent linear approximation and the associated Hessian depend on the DAW span, the ensemble size, and the prior ensemble itself. For instance, if the data partition in the current cycle cost function is relatively localized in time/offset (i.e., a small number of observations), the first iteration estimates the relevant parameters accurately and subsequent iterations do not bring much change to the analysis mean or the anomalies. The iteration permits larger DAW and smaller ensembles. For the transform version of the IEnKS, the first iteration is equivalent to the 4D-ETKF of Hunt et al. (2004), which is also equivalent to the asynchronous EnKF (AEnKF) of Sakov et al. (2010) when these methods are adapted for static parameter estimation. However, if it is used for assimilating observations at a single time instance, it is like a maximum likelihood ensemble filter (Zupanski, 2005).

Because the method seeks to fit the velocity profile by a linear combination within the ensemble subspace, the depth range reflected within the span of the DAW cannot be too large. It is problematic to obtain a weight vector to adequately approximate a large depth section profile. For the case below, a depth window of constant length is used. Based on the prior mean, seismic traveltime intervals are constructed so that these approximately cover the specified depth span by the assimilation cycle. With increasing velocity as a function of depth, the time windows will be larger at earlier arrival times.

The prior in an assimilation cycle defines the exploration space, and as such it must have sufficient variability. Otherwise, it will be difficult to obtain an analysis mean $\bar{\mathbf{x}}^\mathrm{a} = \bar{\mathbf{x}}^\mathrm{f} + \mathbf{X}_\mathrm{f}\mathbf{w}$, resulting in synthetic waveform data that are sufficiently close to the observations within the DAW. A larger ensemble size is always helpful, but a trade-off exists with the realistic time constraint set by the complex forward model.

EXAMPLE

The example consists of seismic CMP gather data that are constructed by forward modeling a well log. The aim is to predict the true smoothed well-log profile of elastic parameters and conduct some degree of uncertainty quantification.

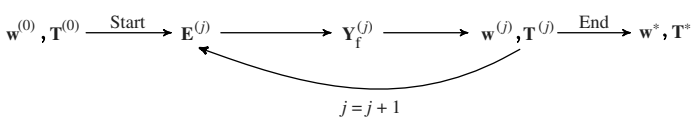


Figure 2. Diagram indicating the iterative structure of an IEnKS analysis cycle.

The target region of the subsurface is 4 km in depth, and it is parameterized by $l = 40$ layers of 100 m thickness each. These layers have homogeneous elastic parameters of P-wave velocity V_P, S-wave velocity V_S, and density ρ, along with compressional and shear attenuation factors Q_P and Q_S that are held fixed. Hence, the parameter vector for the inversion is the l layer distribution of P-wave velocity $\mathbf{V}_P = [V_{P,1}, V_{P,2}, \ldots, V_{P,l}]^T$ with $V_{P,i}$ being the ith layer value and similarly for shear velocity $\mathbf{V}_S$ and density $\boldsymbol{\rho}$. To ensure the positivity of the elastic properties, the unknown $m = 3l$ dimensional parameter state is defined as $\mathbf{x} = [\log \mathbf{V}_P, \log \mathbf{V}_S, \log \boldsymbol{\rho}]^T$.

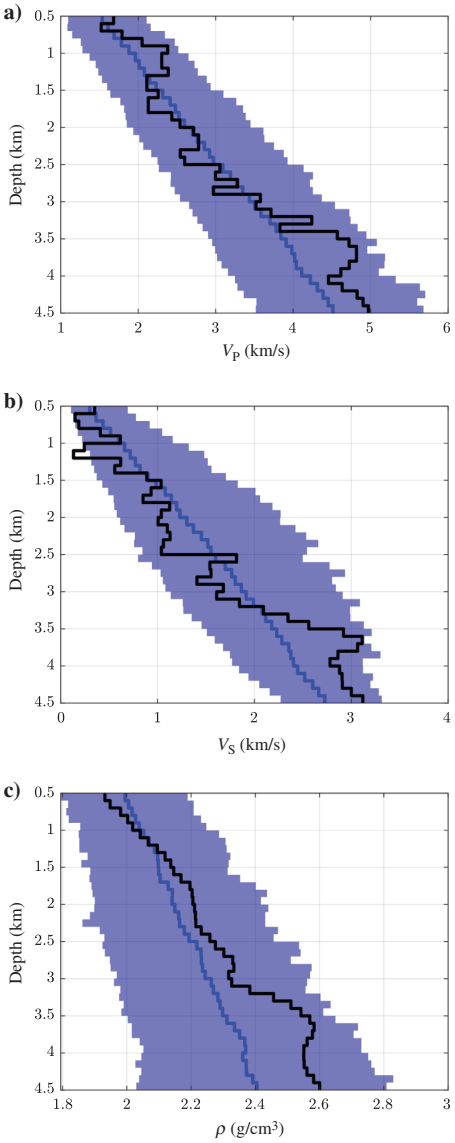


Figure 3. Prior ensemble distribution: (a) acoustic velocity, (b) shear velocity, and (c) density. The black line indicates the true model; the colored line indicates the ensemble median, and the shaded area covers 90% of the ensemble span.

The initial prior $p(\mathbf{x})$ is defined as a m-variate Gaussian distribution $N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$. Its mean $\boldsymbol{\mu}_x$ is constructed as a linear depth trend in the P- and S-wave velocities and density and then log-transformed, whereas the covariance matrix is defined in the log domain. An assumption of increasing (marginal) variance with depth is used for each of the properties. They share a common depth correlation structure, defined via a Matérn (of order 3/2) function (Chiles, 2012). The correlation length parameter is set to obtain a correlation of 5% at 500 m. Because it is expected that the P- and S-wave velocities are dependent in the true depth profile, the correlation coefficient between them is set as $\eta_{PS} = 0.5$, whereas density is considered independent of the velocities. The prior covariance is then formed as

$$\boldsymbol{\Sigma}_x = \mathrm{diag}(\boldsymbol{\sigma}_x)\left[\begin{bmatrix} 1 & \eta_{PS} & 0 \\ \eta_{PS} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \otimes \boldsymbol{\Gamma}\right]\mathrm{diag}(\boldsymbol{\sigma}_x), \quad (11)$$

where $\mathrm{diag}(\cdot): \mathbb{R}^m \mapsto \mathbb{R}^{m\times m}$ maps the marginal standard deviation vector $\boldsymbol{\sigma}_x$ into a diagonal matrix and $\otimes$ is the matrix Kronecker product. The size $l \times l$ correlation matrix $\boldsymbol{\Gamma}$ holds the correlations between variables at different depths.

For the initial ensemble, $n = 300$ samples are drawn from this prior distribution. Figure 3 displays in the empirical median line and 90% limits as the shaded area, in the physical domain. The nonsymmetry is due to the exponential transformation. The true profile (the black line) is covered by the initial ensemble. Having a large ensemble size provides a better basis to form the analysis estimate. With several elastic parameters covering the same depth space causing the reflection signal, one obtains more degrees of freedom to estimate all of the parameters within the relevant depth range. With a small ensemble, the dominant elastic property would be favored in the estimation, most likely V_P, and leaving less, if any, freedom for the other parameters.

For this numerical example, synthetic data are generated using the ERZSOL3 forward model on a smoothed version of the well-log data. A sample from the Gaussian observation error is added to this "true" seismic gather. The noise covariance matrix $\mathbf{R}$ has a diagonal structure with a constant variance $\mathbf{R} = \sigma_e^2\mathbf{I}$, so the observation noise signal has zero cross-covariance and equal power throughout the CMP gather. The noise variance was set to obtain a signal-to-noise level of 10 dB, where the reference signal power is based on an average signal power over offsets in the range 0–3 km in the time interval of 1–3 s in the true model. With such a constant noise variance, the noise contribution will vary a lot through time and offsets, and toward the end of the gather, it will be difficult to distinguish the signal from the noise.

The seismic point source is located in the topmost water layer at depth 5 m below the free surface. The source time function is a fifth-order minimum-phase Butterworth wavelet with a bandpass in the range of 2–50 Hz, albeit the seismograms are limited to a frequency response within 2–30 Hz. Reducing the output frequency bandwidth is primarily done for reasons of computational resources, but it also reduces the nonlinearity of the cost function. The output temporal sampling is 2 ms. The receiver array is located at the same depth as the source with spacing of 50 m. There are 100 receiver

stations giving an offset range of 50 m to 5 km. The water layer has a thickness of 500 m and an acoustic velocity of 1.495 km s^{-1}.

Gather data are partitioned into time intervals such that each assimilation cycle is supposed to condition on data that are reflected from a limited depth range. This is archived using simple depth-to-time conversion. With the velocity profile (V_P only) of the initial prior mean as reference, the ith layer is attributed a zero-offset, two-way traveltime as $t_i = 2\sum_{j=0}^{i} T_j/V_{P,j}$ where T_j's are layer thickness. With increasing velocity, these time intervals become shorter with depth, and a minimum size of 250 ms is enforced for later arrival times. Here, depth spans of four layers covering 400 m are used, resulting in 11 time intervals. Because the first 0.66 s is roughly the two-way traveltime for the first reflection from the seabed, the first time values start at approximately 0.6 s; thus, data between 0.6 and 4 s are used.

This division in time intervals is displayed in Figure 4, where a mute line is also indicated. The line defines an outer mute region in which observations are discarded, and it is found as the normal moveout for the top water layer. This mute region does exclude refraction data at far offsets, containing valuable information on shallower layers. But these parameters should already be estimated previously from the reflection data at earlier times, so it does not affect the estimation.

Shortening the DAW length as the estimation area moves downward, reduces the nonlinearity of the current cycles' minimization problem, and thus increases the chance of obtaining convergence for data with an increasing noise level compared with signal strength. The tolerance for stopping the iteration is $\varepsilon_J = 1 \times 10^{-3}$, and the termination criteria are set to $j_{\max} = 15$.

The estimation results are shown next. Results are presented in the physical domain $\mathbf{m} = \exp(\mathbf{x})$ — with $\mathbf{m}$ being either $\mathbf{V}_P$, $\mathbf{V}_S$, or $\boldsymbol{\rho}$ — where they are log-normal distributed and tend to be non-symmetric. The best-guess estimate $\hat{\mathbf{m}}$ in this domain is taken as the median of the mapped ensemble. The absolute error statistic is $|\hat{\mathbf{m}} - \mathbf{m}^t|$ with $\mathbf{m}^t$ being the true model. Some of the figures present standard deviations that indicate the scale parameters of the mapped ensemble distribution.

After the final assimilation cycle (Figure 5), the ensemble is hard to distinguish from the true model in the depth span of 0.5–3.5 km, and the prediction appears to be very good. Summary results from these posterior ensembles are displayed in Figure 6 as the estimation bias and the (ensemble) standard deviation. The estimation bias is seen to correlate well with the ensemble standard deviation. Whereas the estimation bias in field data application is unknown, the standard deviations in Figure 6 will be available and indicates that from depth approximately 3.5 km and downward, the estimation starts being more uncertain, which is also depicted in Figure 5. This increase in uncertainty might be due to the sharp contrast in

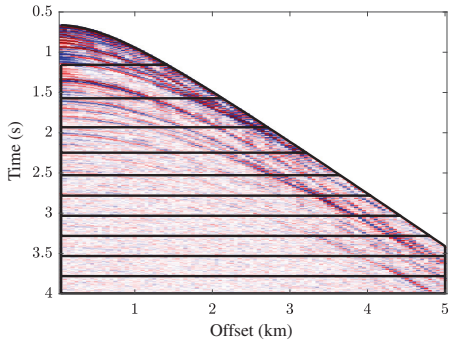


Figure 4. Seismic data with normal moveout line used for muting and boxes indicate time interval blocking.

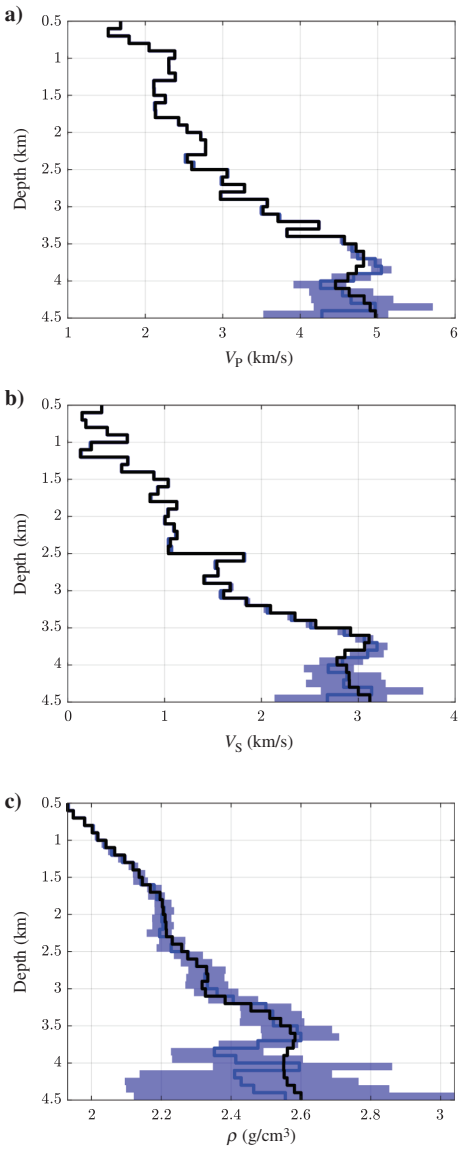


Figure 5. Posterior ensemble distribution: (a) acoustic velocity, (b) shear velocity, and (c) density. The black line indicates the true model; the colored line indicates the ensemble median, and the shaded area covers 90% of the ensemble span.

P-wave velocity, but there are also less informative reflection data at these depths.

The estimation bias and uncertainty of S-wave velocity are surprisingly large at depths of 0.7–0.9 and 1–1.2 km. Both regions are associated with a drop in the S-wave velocity, whereas the acoustic velocity is smoother. The relatively large bias of V_S is hence interpreted as a consequence of the preferential estimation of V_P before V_S, in addition to the solution being a linear combination within the ensemble subspace and the imposed correlation between elastic velocities in the prior.

The inclusion of the finite-size ensemble formulation is intended to counteract the resulting ensemble deflation in each analysis update due to the limited ensemble size. Figure 7a shows the adaptive inflation effect (equation 8). The dots are the values this factor takes during the iterative solution at each cycle, whereas the line connects the value at the last iteration and hence is the factor that enters the actual analysis ensemble update. The inflation factor depends on the weight norm $\mathbf{w}^T\mathbf{w}$, and the range it takes during a cycle thus indicates how much the weights change during the iteration. Whereas the first value at each iteration is the same (as $\mathbf{w}^{(0)} = \mathbf{0}$) and corresponds to a slight deflation, this changes quickly after the first iteration depending on the norm of the weight update and essentially reflects how much the ensemble mean needs to move. The effective values of inflation are seen to vary quite a lot, and being a function of ensemble and DAW size, they would be difficult to establish manually.

This kind of inflation aims at maintaining the ensemble spread. Figure 7b displaying the ensemble standard deviation for V_P over the course of assimilation cycles shows that the parameter estimation works its way down, and the intended behavior of keeping the

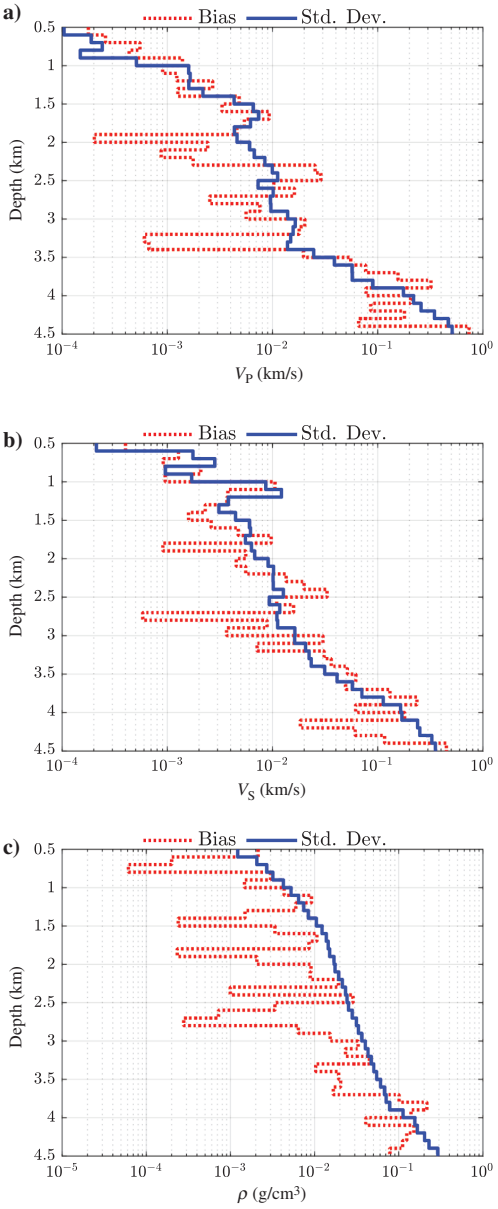


Figure 6. Estimation results as bias $|\hat{\mathbf{m}} - \mathbf{m}^t|$ and standard deviation around $\hat{\mathbf{m}}$: (a) acoustic velocity, (b) shear velocity, and (c) density.

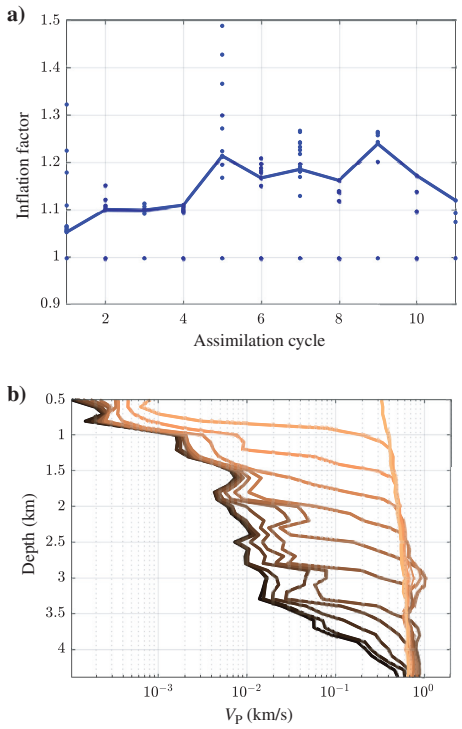


Figure 7. Adaptive inflation over assimilation cycles. (a) Adaptive inflation factor β. (b) Ensemble standard deviation. The order of assimilation cycles goes from light (first) to dark (last).

ensemble spread in place at lower depths is clearly fulfilled. This prevents an unnecessary shrinkage of prior parameter spread before the reflections from these are observed at later times. (The behavior is similar for the other elastic properties and is omitted here).

DISCUSSION

There are many components within this methodology that can be discussed; gaining geophysical insight about the inversion properties, understanding the effects of the DAW, and of course various elements with the prior and likelihood modeling, including differences in the true forward model and the implemented model used in the inversion. Here, the discussion is restricted to a comparison of acoustic and elastic inversion and the amount of data included in an assimilation cycle.

To compare a case of acoustic inversion to the elastic case, an inversion with shear velocity and density fixed at their true value is performed. The acoustic case uses the same initial prior ensemble for V_P, the same set of noise added to the observations, identical DAWs, and stopping criteria. The resulting estimation bias and ensemble standard deviation are compared in Figure 8, where the results for the elastic case are identical to the ones presented in Figure 6a. As expected, the acoustic case shows better estimation and lower uncertainty, as this case is relying on the true shear velocity and density. However, the trends are similar, and the acoustic case shows a similar increase in the standard deviation from approximately 3.5 km and downward, albeit to a smaller degree. Because the ensemble size is the same, the acoustic case has more degrees of freedom to fit the true profile, and the forecast observations are not influenced by variability of the other elastic properties. Both cases ran on 20 cores on the available moderate-size computing facility; the acoustic case finished in 3 h and 8 min, and the elastic case in 4 h and 41 min. This difference is solely attributed to the elastic case requiring more iterations before termination. These calculation times should be viewed in a relative sense because they are influenced by several factors. The reflectivity method (the forward model) can be implemented using different approaches and the ERZSOL3 program implementation might not be the most efficient. Other than that, in an industrial setting, the number of cores for parallel evaluation of the ensemble would be much larger. Finally, the overall execution time depends on the number of assimilation cycles, so how the data partitioning is carried out has a significant impact. The DAW sizes used in this example could have been longer, making the estimation more efficient.

The limitation on the DAW size is due to a detrimental effect when including too many observations in the analysis step. The data partitioning used in the example was constructed to localize certain depth ranges; it was a means to an end, which is to avoid a critical amount of data. Accordingly, this discussion will exemplify further the necessity of splitting the gather data into smaller blocks and condition on these sequentially.

For the sake of avoiding external effects, the finite size approach with inflation is not used in this comparison. The first term in the Hessian then becomes the identity matrix; thus, the ETM is $\mathbf{T} = (\mathbf{I} + \mathbf{Y}_f^T \mathbf{R}^{-1} \mathbf{Y}_f)^{-1/2}$.

The p-dimensional observation perturbations $\mathbf{Y}_f$ are mapped (projected) onto the n-dimensional ensemble subspace with $n \ll p$. Whereas p is large, the effective dimension will be smaller due to a large degree of multicollinearity within the waveform data. Still, the forecast observation covariance needs to be sufficiently resolved within the provided n degrees of freedom; otherwise, it is underestimated and will likely lead to filter divergence.

The experiment is as follows: For a particular sample of prior ensemble and observation noise, a first analysis cycle is carried out with stopping criteria as in the previous section, thus obtaining an analysis ensemble. This ensemble forms the prior for the next cycle, which is what is evaluated here. For each sample (ensemble and noise), two cases of time window lengths are compared. The top is identical, but one is 100 ms longer than the other. Hence, the amount of data that is projected onto the ensemble subspace is larger for the longer window. The lengths of the windows are chosen such that the shorter one is stable, whereas the longer one starts to be unstable.

The notion of too many data points is relative to the ensemble size and the waveform signal content in the ensemble. Thus, there will be a difference between a case of only acoustic velocity variations or variation in all of the elastic parameters. Therefore, these two cases are compared for the same time windows.

In each case, the analysis is replicated 20 times, for elastic and acoustic. The ensemble size is reduced to 100 here because it does not affect the example, but only the critical length of the time window.

The results are shown in Figure 9 with the left column being the shorter time window and the right column the longer one. Shaded areas are ranges over the set of 20 replicates. The top rows show the ETM eigenspectrum after the first ensemble evaluation. Here, λ_i is the ith eigenvalue of $(\mathbf{Y}_f^{(0)})^T \mathbf{R}^{-1} \mathbf{Y}_f^{(0)}$. The bottom rows show the

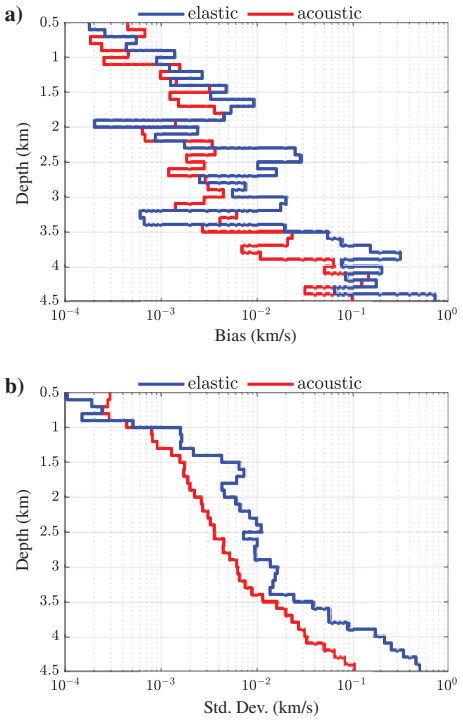


Figure 8. Comparison of $\mathbf{V}_P$ estimation between the acoustic and elastic case. (a) Bias $|\hat{\mathbf{V}}_P - \mathbf{V}_P^t|$. (b) Standard deviation around $\hat{\mathbf{V}}_P$.

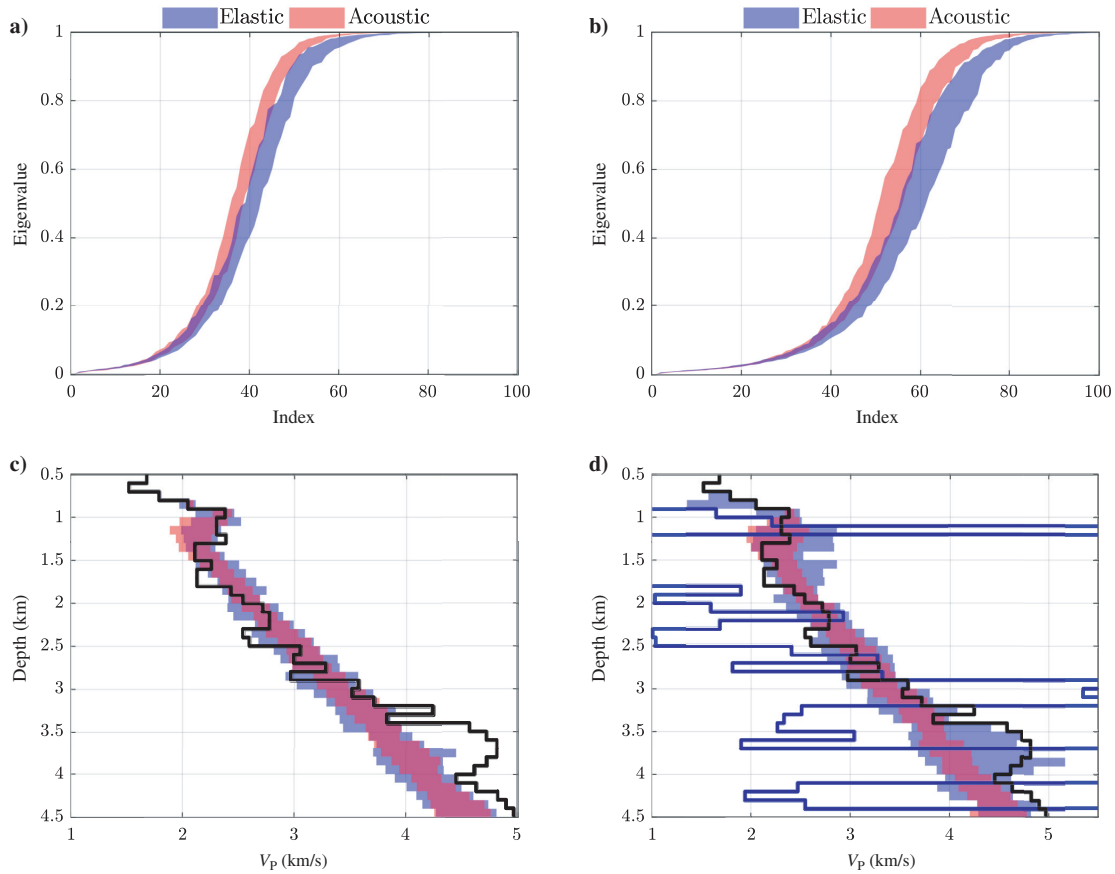


Figure 9. Replicate set of analysis results over different time window lengths. The left column represents the shorter time window, whereas the right column represents the longer one. The top row shows ETM eigenspectrum, whereas the bottom row shows the analysis mean mapped onto the physical domain. The black line indicates the true profile.

analysis ensemble estimate (mean) for each replicate with the shaded area ranging over the set of replicates. This is used as a proxy for the appropriateness of the analysis ensemble for use in the next assimilation cycle, as this ideally should be fairly similar over replicates and close to the mean of the initial prior for larger depths.

A comparison of Figure 9a and 9b shows that the ETM has fewer eigenvalues close to one when a longer DAW is used (right column). For this case (Figure 9d), the acoustic analysis mean through depth still remains within a confined region and the estimation goes a little further down due to the longer time window. But, for the elastic case, there is much more variability among the replicates, indicating large sensitivity to the ensemble sample. For this elastic case, the shaded areas only cover 15 of 20 replicates as the remaining five replicates terminate with wild updates, of which a single replicate is visualized. Such an analysis would be useless as prior for subsequent data assimilation cycles, and the forward model solver might not accept such input. An overshooting update in the mean is accompanied by collapse in the preconditioned ensemble used for the sensitivities and the solution diverges quickly within the iteration.

In the cases presented here, the synthetic seismograms were limited to a certain frequency bandwidth as part of the forward modeling. Enlarging this bandwidth and thereby introducing higher frequency waveforms into the forecast data would further increase the need to reduce the data amount to adequately resolve the information content within the ensemble subspace.

CONCLUSION

In this paper, an ensemble-based method is used for elastic inversion of seismic reflectivity data. This is a novel approach for seismic waveform inversion, and results from applying the IEnKS method are very promising. In this context, the assimilation is performed over blocks of seismic data, which regularizes the estimation and reduces the potential problem of overfitting. Moreover, the assimilation step is performed in an iterative manner, where a tangent linear model is implicitly defined and iteratively refined to approximate the seismic observation model. The ensemble-based approach means that the gradients and Hessian expressions are

specified via the sample variability. The results appear to give a fair estimation and reasonable levels of uncertainties in the solution for elastic parameters.

Even though ensemble-based methods have shown large success in several applications, it is not obvious how to transfer them to this application with seismic data. One challenge is the enormous volume of seismic data, another complexity is that of the wave equation forward model that complicates the likelihood model for the data. In this paper, the main results are augmented by studying the statistical properties of predictions, and this kind of analysis fosters insight about the potential applicability of the method.

A numerical experiment is used to provide a proof-of-concept study. Further work includes extensions to more complicated waveform data, as well as applications with 2D subsurface models. In the end, the assimilation choices would depend on the nonlinearity inherent in the problem and the geometry. Some kind of real-time diagnostic plots could be used to inspect the quality of the current inversion, and this might facilitate more adaptive data assimilation. Such approaches must also consider the evaluation costs that should not grow too large.

ACKNOWLEDGMENTS

We thank partners of the Uncertainty in Reservoir Evaluation project at NTNU and BP for data and discussions.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are confidential and cannot be released.

REFERENCES

Aanonsen, S. I., G. Nævdal, D. S. Oliver, A. C. Reynolds, and B. Vallès, 2009, The ensemble Kalman filter in reservoir engineering — A review: SPE Journal, **14**, 393–412, doi: 10.2118/117274-PA.

Asch, M., M. Bocquet, and M. Nodet, 2016, Data assimilation: Methods, algorithms, and applications: SIAM.

Bocquet, M., 2011, Ensemble Kalman filtering without the intrinsic need for inflation: Nonlinear Processes in Geophysics, **18**, 735–750, doi: 10.5194/npg-18-735-2011.

Bocquet, M., P. N. Raanes, and A. Hannart, 2015, Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation: Nonlinear Processes in Geophysics, **22**, 645–662, doi: 10.5194/npg-22-645-2015.

Bocquet, M., and P. Sakov, 2012, Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems: Nonlinear Processes in Geophysics, **19**, 383–399, doi: 10.5194/npg-19-383-2012.

Bocquet, M., and P. Sakov, 2013, Joint state and parameter estimation with an iterative ensemble Kalman smoother: Nonlinear Processes in Geophysics, **20**, 803–818, doi: 10.5194/npg-20-803-2013.

Bocquet, M., and P. Sakov, 2014, An iterative ensemble Kalman smoother: Quarterly Journal of the Royal Meteorological Society, **140**, 1521–1535, doi: 10.1002/qj.2236.

Bosch, M., L. Cara, J. Rodrigues, A. Navarro, and M. Díaz, 2007, A Monte Carlo approach to the joint estimation of reservoir and elastic parameters from seismic amplitudes: Geophysics, **72**, no. 6, O29–O39, doi: 10.1190/1.2783766.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic wave-form inversion: Geophysics, **60**, 1457–1473, doi: 10.1190/1.1443880.

Chen, Y., and D. S. Oliver, 2011, Ensemble randomized maximum likelihood method as an iterative ensemble smoother: Mathematical Geosciences, **44**, 1–26, doi: 10.1007/s11004-011-9376-z.

Chen, Y., and D. S. Oliver, 2013, Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification: Computational Geosciences, **17**, 689–703, doi: 10.1007/s10596-013-9351-5.

Chiles, J.-P., 2012, Geostatistics, 2nd ed.: John Wiley & Sons Inc, Wiley Series in Probability and Statistics.

Dovera, L., and E. Della Rossa, 2011, Multimodal ensemble Kalman filtering using Gaussian mixture models: Computational Geosciences, **15**, 307–323, doi: 10.1007/s10596-010-9205-3.

Eidsvik, J., P. Avseth, H. Omre, T. Mukerji, and G. Mavko, 2004, Stochastic reservoir characterization using prestack seismic data: Geophysics, **69**, 978–993, doi: 10.1190/1.1778241.

Evensen, G., 2018, Analysis of iterative ensemble smoothers for solving inverse problems: Computational Geosciences, **22**, 885–908, doi: 10.1007/s10596-018-9731-y.

Fichtner, A., 2010, Full seismic waveform modelling and inversion (advances in geophysical and environmental mechanics and mathematics): Springer.

Gineste, M., and J. Eidsvik, 2015, Framework for seismic inversion of full waveform data using sequential filtering: Presented at the 3rd Petroleum Geostatistics Conference.

Gouveia, W. P., and J. A. Scales, 1998, Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis: Journal of Geophysical Research, **103**, 2759–2779, doi: 10.1029/97JB02933.

Gu, Y., and D. S. Oliver, 2007, An iterative ensemble Kalman filter for multiphase fluid flow data assimilation: SPE Journal, **12**, 438–446, doi: 10.2118/108438-PA.

Hunt, B. R., E. Kalnay, E. J. Kostelich, E. Ott, D. J. Patil, T. Sauer, I. Szunyogh, J. A. Yorke, and A. V. Zimin, 2004, Four-dimensional ensemble Kalman filtering: Tellus A: Dynamic Meteorology and Oceanography, **56**, 273–277, doi: 10.1111/j.1600-0870.2004.00066.x.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007, Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter: Physica D: Nonlinear Phenomena, **230**, 112–126, doi: 10.1016/j.physd.2006.11.008.

Kennett, B., 2005, ERZSOL3, http://www.spice-rtn.org/library/software/ERZSOL3.html, accessed 10 September 2017.

Kennett, B., 2011, Seismic wave propagation in stratified media: ANU Press.

Kennett, B. L. N., and N. J. Kerry, 1979, Seismic waves in a stratified half space: Geophysical Journal International, **57**, 557–583, doi: 10.1111/j.1365-246X.1979.tb06779.x.

Liu, M., and D. Grana, 2018, Stochastic nonlinear inversion of seismic data for the estimation of petroelastic properties using the ensemble smoother and data reparameterization: Geophysics, **83**, no. 3, M25–M39, doi: 10.1190/geo2017-0713.1.

Oliver, D. S., and Y. Chen, 2010, Recent progress on reservoir history matching: A review: Computational Geosciences, **15**, 185–221, doi: 10.1007/s10596-010-9194-2.

Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503, doi: 10.1111/j.1365-246X.2006.02978.x.

Raanes, P. N., A. Carrassi, and L. Bertino, 2015, Extending the square root method to account for additive forecast noise in ensemble methods: Monthly Weather Review, **143**, 3857–3873, doi: 10.1175/MWR-D-14-00375.1.

Sakov, P., G. Evensen, and L. Bertino, 2010, Asynchronous data assimilation with the EnKF: Tellus A: Dynamic Meteorology and Oceanography, **62**, 24–29, doi: 10.1111/j.1600-0870.2009.00417.x.

Sakov, P., and P. R. Oke, 2008, Implications of the form of the ensemble transformation in the ensemble square root filters: Monthly Weather Review, **136**, 1042–1053, doi: 10.1175/2007MWR2021.1.

Sheriff, R. E., and L. P. Geldart, 1995, Exploration seismology, 2nd ed.: Cambridge University Press.

Stordal, A. S., 2014, Iterative Bayesian inversion with Gaussian mixtures: Finite sample implementation and large sample asymptotics: Computational Geosciences, **19**, 1–15, doi: 10.1007/s10596-014-9444-9.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.

Thurin, J., R. Brossier, and L. Métivier, 2017, Ensemble-based uncertainty estimation in full waveform inversion: 79th Annual International Conference and Exhibition, EAGE, Extended Abstracts, Tu P1 07.

Thurin, J., R. Brossier, and L. Métivier, 2019, Ensemble-based uncertainty estimation in full waveform inversion: Geophysical Journal International, **219**, 1613–1635, doi: 10.1093/gji/ggz384.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, doi: 10.1190/1.3238367.

Zhang, J., G. Lin, W. Li, L. Wu, and L. Zeng, 2018, An iterative local updating ensemble smoother for estimation and uncertainty assessment of hydrologic model parameters with multimodal distributions: Water Resources Research, **54**, 1716–1733, doi: 10.1002/wrcr.v54.3.

Zupanski, M., 2005, Maximum likelihood ensemble filter: Theoretical aspects: Monthly Weather Review, **133**, 1710–1726, doi: 10.1175/MWR2946.1.

Paper II

Regularizing the data assimilation window in the iterative ensemble Kalman smoother for seismic inversion

Michael Gineste & Jo Eidsvik

Submitted to
Computational Geosciences, 2020

Computational Geosciences manuscript No.
(will be inserted by the editor)

Regularizing the data assimilation window in the iterative ensemble Kalman smoother for seismic inversion

Michael Gineste · **Jo Eidsvik**

Received: date / Accepted: date

Abstract An ensemble-based method for seismic inversion of elastic attributes is considered, namely the iterative ensemble Kalman smoother. The main focus of this work is the challenge associated with ensemble-based inversion of seismic waveform data. The amount of seismic data is large and, depending on ensemble size, cannot be processed in a single batch. Instead a solution strategy of partitioning the data recordings in time windows and processing these sequentially is suggested. This work demonstrates how this partitioning can be done adaptively, with a focus on reliable and efficient estimation. The adaptivity relies on an analysis of the iterative update and an interpretation of contributions from prior and likelihood to this update. The idea is that these must balance; if the prior dominates, the estimation process is inefficient while if data dominates, the estimation if likely to overfit and diverge. Two alternative interpretations are formulated and evaluated, and only one is found to sufficiently regularize the data window. Although no guarantees for avoiding ensemble divergence are provided in the paper, the results of this adaptive procedure indicate that robust estimation performance can be achieved for ensemble-based inversion of seismic waveform data.

Keywords Ensemble smoother · Iterative Ensemble Kalman Smoother · Data assimilation · Seismic inversion

Mathematics Subject Classification (2010) 62L12 · 86-08

1 Introduction

The motivation behind this work is seismic waveform inversion, where the goal is to predict the elastic attributes of the subsurface, in the form of acoustic- and shear velocity and density, conditional on records of seismic reflection data. Seismic inversion thus provides an image of the subsurface and its interpretation can, combined with other geophysical analysis, be used to establish a geological model.

Phrased in a Bayesian setting, where initial knowledge is incorporated via a prior probability distribution and a likelihood model is used for the specific data, the solution to this Bayesian inversion problem is available as the posterior probability distribution. However, with the non-linearity and complexity of the forward model, there is no exact solution to this posterior.

With the growing availability of diverse data types in complex spatio-temporal systems, there is currently much focus on data assimilation methods that scales well with high-dimensional spaces. Such is the ensemble Kalman framework [1, 5, 10] that is increasingly applied to problems in the geosciences [4] and has a succesful track record in history matching applications. In particular, the method referred to as the iterative ensemble Kalman smoother (IEnKS), introduced by Bocquet and Sakov [2], is here used for the case of static parameter estimation. The IEnKS combines aspects of ensemble-based and variational approach to data assimilation. Notably, it avoids the need for tangent linear models, an attractive feature when using black-box forward models. Instead the sensitivity is indirectly provided by the ensemble evaluation. This implies that the ensemble is updated using a common sensitivity, and excludes the possibility of multimodality in the posterior ensemble.

Recent developments indicate that ensemble-based approaches can be used for inversion of seismic waveform data [7, 17], but it is not obvious how to assimilate the massive data in a reliable manner. The inversion is then formulated as a sequential data assimilation problem, where disjoint subsets of the seismic data records, denoted by data assimilation windows (DAWs), are used to update the ensemble at every assimilation cycle. If the DAWs are too large, the IEnKS

Department of Mathematical Sciences, NTNU, Trondheim, Norway
E-mail: michael.gineste@ntnu.no, jo.eidsvik@ntnu.no

analysis will diverge. If the DAWs are too small, the inversion run time will grow because of the time-requirements of the forward model. A key issue here is hence to find efficient DAW sizes, and to do so automatically. An efficient DAW size is maximizing amount of data to be assimilated while providing a good starting point for convergence. This is what is meant by regularizing the DAW.

Analysis of the iterative update as a vector in the Hessian eigenbasis has been considered by others, e.g. [16] that used it to guide the choice of Levenberg-Marquardt regularization parameter. The focus in this work is also on the vector coefficients whereas the angle of analysis here is on how they are made up of contributions from prior and likelihood, and how these changes when the amount of data increases. The prior is a regularizing component and maintaining enough of its influence is the aimed balance.

An example with synthetic seismic data is used throughout the paper to give intuition around concepts and methods. The methodological developments should also be of interest in other applications.

A general issue with ensemble-based methods is rank deficiency of estimated covariance matrices, which is often addressed using localization and/or inflation. Neither of these techniques are considered here, as the focus is solely on the data dimension aspect and how to select this appropriately. While still in a synthetic setting, the challenges addressed are realistic and so are the described solution. Hence this paper describes a remedy to enable ensemble-based seismic waveform inversion and how an ensemble-based method can be applied to an inverse problem when its knobs are tweaked and adjusted in light of particularities of the applied problem.

The paper is structured as follows: In section Section 2 the main building blocks for sequential seismic inversion are introduced. In Section 3 the IEnKS method is outlined. In Section 4 two alternative methods for adaptive DAW are presented. In Section 5 one of the methods are applied to examplify the seismic inversion problem. In Section 6 a discussion and analysis of the two methods is provided.

2 Seismic inversion by sequential data assimilation

2.1 Seismic waveform model and Bayesian inversion

The motivation for this work is probabilistic inversion of seismic waveform data. The seismic inverse problem is that of inferring the subsurface properties from measured seismic reflection data, in the light of a physical model supposed to be able to predict the seismic experiment. This inverse problem is ill-posed and can have multiple solutions. This non-uniqueness of surface properties resulting in nearly identical seismic traces in some time window, poses problems to seismic inversion.

The elastic properties sought inferred are acoustic wave velocity v_p, shear wave velocity v_s and density ρ. As seismic data providing information to infer these from, common mid point (CMP) gathers are considered here. Such gathers represent partly processed waveform data, obtained by stacking shot-receiver data to a common mid-point location along the seismic acquisition line and sorted in the time-offset domain [15]. Assuming a subsurface consisting of layers, such data can be simulated by a seismic forward model that maps a depth profile of layer thickness' with associated elastic properties, to seismograms at offset points from the source. A commonly used forward model is the reflectivity method [9]. Under the assumption of a layered subsurface, the elastic wave equation can be transformed and solved in the slowness-frequency domain, and mapped to time domain seismograms via (inverse) Fourer transformation. The full recording time of the gather is thus calculated at once. Several implementations of the reflectivity method exists, the one used here is ERZSOL3 [8]. While the reflectivity method is quite fast compared to other numerical methods for elastic wave propagation, it is still time consuming and the realistic number of simulations as part of solving the the inverse problem is limited.

The seismic gather that will provide the example in this paper is shown in Fig. 1. Here, the seismic CMP data are

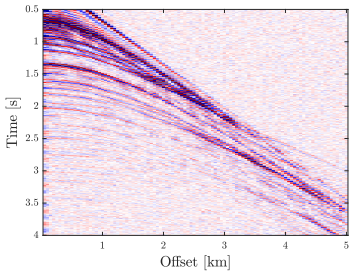


Fig. 1: Example of seismic CMP gather data

semi-synthetic in the sense that processed data from a well log of elastic measurements have been forward propagated with the same forward model. The strongest reflections in Fig. 1 represent major gradients in the elastic properties at shallower layers. These reflections appear as hyperbolic lines in the time-offset plot as the seismic waves take longer time to reach far offsets. At far offsets and earlier arrival time, the gather contains no records of a reflected wave, and data in this region will be excluded by imposing a mute region. This includes ignoring the direct source-to-receiver wave propagation. The seismic gather record has a set end time after which measurements are no longer used for inversion.

Because the pressure wave travels faster than the shear wave, the earlier parts of the data are dominated by the pressure wave differences in the subsurface, while the differences in shear properties arrive later in the data and with smaller amplitude. Notably, the seismic reflection data indicate products or sums of the elastic properties: For instance, the acoustic impedance is defined as the product of pressure wave velocity and density. It is hence difficult to split a reflection event in a causal underlying change in lower pressure wave velocity and higher density, or vice versa. The data sensitivities to density perturbations are somehow masked behind the sensitivity to wave velocities, and these must be known quite accurately before one can target density. These ambiguities are smaller with plenty of angle information available in time-offset plots as in Fig. 1, but it is still difficult to infer elastic properties from seismic CMP data.

Framed as a Bayesian inverse problem, the random variable of interest is the parameter state vector $\mathbf{x}$, with an assigned prior probability density function $p(\mathbf{x})$. The state consists of elastic properties in l homogeneous layers of the subsurface, and the prior distribution is represented by a multivariate Gaussian process with mean and covariance terms specified from initial knowledge. To ensure positivity of the elastic attributes, the state vector is the log-transform of these, $\mathbf{x} = \left[\log \mathbf{v}_p, \log \mathbf{v}_s, \log \boldsymbol{\rho}\right] \in \mathbb{R}^m$, where the parameter dimension m is three times the number of layers l. The number of layers and the thickness' depth profile are held fixed.

Data are denoted $\mathbf{y}$ and are measurements of the reflected wave amplitude as function of, besides the subsurface model, arrival time and offset relative to source position, as well as the source signal and boundary conditions. The data are linked to the state via the forward model $h(\mathbf{x})$, which represents the elastic wave propagation as simulated by the reflectivity method. The observation model for data $\mathbf{y}$ is assumed to be unbiased according to the forward model (perfect model assumption) and with additive noise, i.e. $\mathbf{y} = h(\mathbf{x}) + \mathbf{e}$, where $\mathbf{e}$ is an independent zero-mean Gaussian measurement noise vector with covariance matrix $\mathbf{R}$. The resulting likelihood is $p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(h(\mathbf{x}), \mathbf{R})$. The solution to the probabilistic inverse problem is then, from Bayes' rule, the posterior probability density function $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})$. Due the non-linear relationship between parameter state and observations, the posterior distribution is not directly available.

The non-uniqueness of the inverse problem means that the posterior distribution principally can be multimodal at certain depth regions. The ensemble Kalman method is incapable of representing multimodality as such, and the resulting posterior ensemble is likely to converge to a local mode.

2.2 Sequential data integration

The seismic gather can be split in disjoint subsets $\mathbf{y}_k, k = 1,\ldots,K$ such that $\mathbf{y} = \{\mathbf{y}_1,\ldots,\mathbf{y}_K\}$. In this example with seismic waveform data, the partitions of data cover different arrival times, as defined by a data assimilation window (DAW). Partition k of data is extracted by selecting suitable elements of the forward operator for the data, denoted by $h_k(\mathbf{x})$. Assuming conditionally independent measurement noise terms, given the state vector, the likelihood function can also be partioned $p(\mathbf{y}|\mathbf{x}) = \prod_k p(\mathbf{y}_k|\mathbf{x})$, with the k-th DAW likelihood $p(\mathbf{y}_k|\mathbf{x}) = \mathcal{N}(h_k(\mathbf{x}), \mathbf{R}_k)$ where the k-th DAW noise covariance matrix is $\mathbf{R}_k$. When data is partitioned in such windows of arrival time, the reflected wave measured in a window will stem from a certain depth region which is here referred to as the observed depth region.

Data are assimilated sequentially over these disjoint subsets. At the first assimilation cycle, the prior forms the *forecast* model, which is updated using data $\mathbf{y}_1$ in the first DAW. The *analysis* model from the first cycle is then $p(\mathbf{x}|\mathbf{y}_1)$. This procedure of going from a forecast model to an analysis model continues at the subsequent cycle, and from Bayes' rule:

$$p(\mathbf{x}|\mathbf{y}_1,\ldots,\mathbf{y}_k) \propto p(\mathbf{y}_k|\mathbf{x})\, p(\mathbf{x}|\mathbf{y}_1,\ldots,\mathbf{y}_{k-1})\,, \tag{1}$$

for $k = 2,\ldots,K$. At the last K cycle, all data has been assimilated. The main contribution of this paper is to robustly scale the size of the DAW when initiating an assimilation cycle. With this focus, the cycle index k is ignored in the following where the method is outlined for one assimilation cycle only.

3 Iterative Ensemble Kalman Smoother

This section introduyces the IEnKS and its components, of which some are fundamental for the adaptive DAW selection. First, its ensemble aspect is presented followed by outlining the iterative solution to the variational problem.

3.1 Ensemble-based Data Assimilation

The density functions in Eq. 1 are considered approximated by ensembles of realizations from these distributions, and their moments approximated by sample moments. In an assimilation cycle the forecast ensemble is input while the output is an analysis ensemble. For this static parameter estimation problem, the analysis ensemble then forms the forecast for the next cycle, and this continues until all DAWs are processed.

The members (columns) of a forecast ensemble $\mathbf{E}^{\mathrm{f}}$ are denoted $\mathbf{x}_i^{\mathrm{f}}, i = 1,\ldots,n$, where n is the ensemble size. The

state estimate is the ensemble mean:

$$\bar{\mathbf{x}}^{\mathrm{f}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{\mathrm{f}}, \tag{2}$$

and along with the $m \times n$ normalized state anomaly matrix $\mathbf{X}_{\mathrm{f}}$

$$\mathbf{X}_{\mathrm{f}} = \left(\mathbf{E}^{\mathrm{f}} - \bar{\mathbf{x}}^{\mathrm{f}}\mathbf{1}^{\mathrm{T}}\right)/(n-1)^{1/2}, \tag{3}$$

as the square root of the error covariance estimate

$$\mathbf{P}^{\mathrm{f}} = \mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\mathrm{T}}, \tag{4}$$

the two moments are defined via the ensemble. Turning (3) around, the forecast ensemble matrix assembles the mean and anomalies as

$$\mathbf{E}^{\mathrm{f}} = \bar{\mathbf{x}}^{\mathrm{f}}\mathbf{1}^{\mathrm{T}} + (n-1)^{1/2}\mathbf{X}_{\mathrm{f}}. \tag{5}$$

With the available data $\mathbf{y}$ in the considered DAW, the IEnKS updates the forecast ensemble to get an analysis ensemble, described by the analysis mean $\mathbf{x}^{\mathrm{a}}$ and anomaly matrix $\mathbf{X}_{\mathrm{a}}$.

The analysis state is found as a linear combination $\mathbf{x}^{\mathrm{a}} \in \left\{\bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w} \,|\, \mathbf{w} \in \mathbb{R}^n\right\}$ within the span of the ensemble anomalies $\mathbf{X}_{\mathrm{f}}$. With this parameterization of the analysis state, the control vector $\mathbf{w}$ replaces the elastic parameters as the variable of interest. The state that maximizes the posterior distribution of the Bayesian update in Eq. 1, is equivalent to the state that minimizes the following objective function

$$J(\mathbf{w}) = \tfrac{1}{2}\|\mathbf{y} - h(\bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w})\|_{\mathbf{R}}^2 + \tfrac{1}{2}\|\mathbf{w}\|^2. \tag{6}$$

Here, the notation $\|\mathbf{a}\|_{\mathbf{B}}^2 = \mathbf{a}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{a}$ is used. The analysis mean is $\mathbf{x}^{\mathrm{a}} = \bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w}^{\mathrm{a}}$ with the optimal weight vector being the solution $\mathbf{w}^{\mathrm{a}} = \arg\min_{\mathbf{w}} J(\mathbf{w})$. The analysis anomaly matrix is updated using an ensemble transform matrix $\mathbf{T}$ such that $\mathbf{X}_{\mathrm{a}} = \mathbf{X}_{\mathrm{f}}\mathbf{T}$. This transform matrix is the inverse square-root of the Hessian of the objective function (6), evaluated at $\mathbf{w}^{\mathrm{a}}$. With analysis mean $\mathbf{x}^{\mathrm{a}}$ and anomalies $\mathbf{X}_{\mathrm{a}}$ in place, the analysis ensemble is assembled similarly to Eq. 5 and thereby closes an analysis cycle.

3.2 Iterative procedure

The optimal weight vector is found by an iterative approach. Letting index j indicate iteration number, this variational problem is solved as $\mathbf{w}_{j+1} = \mathbf{w}_j + \Delta\mathbf{w}_{j+1}$, where the current iteration search direction is defined by $\Delta\mathbf{w}_j = -\mathbb{H}_j^{-1}\nabla J_j$. This involves the $n \times 1$ gradient (Jacobian) ∇J and the $n \times n$ (approximative) Hessian $\mathbb{H}$, of the objective function, and the iterative scheme is a Gauss-Newton method.

The gradient and Hessian calculation notably involves ensemble evaluations only. The observation anomalies are assumed to be the image of the prior state anomalies, mapped through the forward model gradient (or tangent linear) $\mathbf{Y}_j = \frac{\partial h}{\partial x}\big|_{x_j}\mathbf{X}_{\mathrm{f}}$. This is achieved using the transform variant of IEnKS [14]. Therein a preconditioned state ensemble is used to evaluate observation anomalies, followed by the latter being rescaled or un-transformed (or "de-conditioned", [12]), before being used to form the sensitivities. At iteration j, this preconditioned ensemble is centered around the current mean $\mathbf{x}_j = \bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w}_j$ and the transform matrix from the previous iteration. This iterative ensemble

$$\mathbf{E}_j = \mathbf{x}_j\mathbf{1}^{\mathrm{T}} + (n-1)^{1/2}\mathbf{X}_{\mathrm{f}}\mathbf{T}_j, \tag{7}$$

is used to evaluate the $p \times n$ observation anomaly matrix

$$\mathbf{Y} = \left(h(\mathbf{E}_j) - \bar{\mathbf{y}}_j\mathbf{1}^{\mathrm{T}}\right)/(n-1)^{1/2},$$

with $\bar{\mathbf{y}}_j = h(\mathbf{E}_j)\mathbf{1}/n$ being the mean of the observation ensemble. As the observation anomalies are conditional to the iterative ensemble but should relate to the prior, they are therefore unconditioned as $\mathbf{Y}_j = \mathbf{Y}\mathbf{T}_j^{-1}$ before entering the sensitivities:

$$\nabla J_j = \mathbf{w}_j - \mathbf{Y}_j^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \bar{\mathbf{y}}_j\right), \tag{8a}$$

$$\mathbb{H}_j = \mathbf{I}_n + \mathbf{Y}_j^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_j. \tag{8b}$$

The assimilation cycle is initialized with $\mathbf{w}_0 = \mathbf{0}$ and $\mathbf{T}_0 = \mathbf{I}$, and sensitivities used in the first iteration are based on the prior ensemble. The preconditioned/iterative ensembles (Eq. 7) thus represent a sequence of ensembles going from forecast to analysis, and the process reflects the gradual change in both the parameter estimate and its uncertainty. The gradually changing preconditioning ensemble is important for the evaluation of sensitivities. For the ensemble averaging to approximate the tangent linearization of the forward model satisfactorily, the cross-covariances between variables in the observed region and data arennn required to be adequately estimated. This would not be the case if variability in parameters above the currently observed depth region was still large. The top-down construction of DAWs enables sequential reduction in this variability, and allows the algorithm to focus naturally on the observed regions.

The matrix power operations applied to the Hessian (8b) are facilitated when this matrix is factorized in a eigen-decomposition. In this work, the singular value decomposition (SVD) is applied to the $p \times n$ ensemble of standardized observation anomalies, and with $p \gg n$ generally being the case, 'economic' SVD offers significant computational savings. Ignoring the subscript j, the decomposition is

$$\mathbf{R}^{-1/2}\mathbf{Y} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} = \sum_{i=1}^{n}\lambda_i\mathbf{u}_i\mathbf{v}_i^{\mathrm{T}}, \tag{9}$$

where the $p \times n$ matrix $\mathbf{U}$ has left singular vector $\mathbf{u}_i$ as ith column, and correspondingly for the $n \times n$ matrix of right singular vectors $\mathbf{V}$. The $n \times n$ diagonal matrix $\mathbf{S}$ holds the

sorted singular values $(\mathbf{S})_{ii} = \lambda_i$, $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$. Eq. 9 involves the inverse square root of the error covariance matrix $\mathbf{R}^{-1/2}$, which is straightforward to compute when $\mathbf{R}$ is a diagonal matrix. Inserting this decomposition into Eq. 8b, and using the orthogonal properties $\mathbf{V}\mathbf{V}^{\mathrm{T}} = \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$ which holds when $p > n$, the Hessian becomes

$$\mathbb{H} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}} \text{ with } \mathbf{\Lambda} = \mathbf{I} + \mathbf{S}^{\mathrm{T}}\mathbf{S}\,.$$

Here, the diagonal matrix $\mathbf{\Lambda}$ has elements $(\mathbf{\Lambda})_{ii} = (1 + \lambda_i^2)$. The inverse and square root of the Hessian are then the corresponding operations on the diagonal matrix $\mathbf{\Lambda}$:

$$\mathbb{H}^{-1} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^{\mathrm{T}} \text{ and } \mathbb{H}^{-1/2} = \mathbf{V}\mathbf{\Lambda}^{-1/2}\mathbf{V}^{\mathrm{T}},$$

which are used for the search direction and as the transform matrix, respectively.

3.3 Stopping criteria

A termination criteria is needed to stop the iteration process when continuing this does not improve the solution significantly. Such a criteria is most often expressed as an absolute or relative change in either objective function evaluation or some norm of the control variable, falling below a given tolerance level. In the considered application, it was challenging to set the appropriate tolerance level for any commonly used measure (on $J(\mathbf{w}_j)$ or $\mathbf{w}_j$) that resulted in consistent termination across varying data dimension, signal-to-noise ratios and ensemble sizes.

The reflection data assimilated in an analysis cycle is related to a (localized depth) observation region. The (global depth analysis) optimal control variable $\mathbf{w}^{\mathrm{a}}$ is supposed to form the estimate of the elastic parameters in this observed region, while keeping the prior mean more or less unchanged outside this region. The scale of the cost function is dominated by the data misfit term and is insensitive to (smaller) adjustments in $\mathbf{w}$ that acts on the state mean $\mathbf{x}$ outside the observed region. Thus basing termination on changes in $J(\mathbf{w}_j)$ does not necessarily express that a steady global mean has been reached. Norms on either $\mathbf{w}_j$ or $\Delta\mathbf{w}_j$ are very dependent on the DAW length and on the "distance" between forecast and analysis mean. Suitable threshold on changes in these would vary for different depths of observed region, making them difficult to set beforehand.

One measure was observed to have a consistent behavior across the data window length and its position within the gather and across different ensemble sizes. Most importantly, the measure might have different scale but behaved similarly when it seemingly was a good time to stop iterating, i.e. when $\|\mathbf{w}_j\|$ or $\|\Delta\mathbf{w}_j\|$ had reached stationary levels. This measure is the mutual information (or Shannon information content), originating from information theory but also used within data assimilation [13, 6], which addresses the reduction in entropy/improvement of knowledge. It can be evaluated from the eigenvalues of the $n \times n$ matrix $\mathbf{Y}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}$, referred to as the information matrix in ensemble subspace [18], as

$$\mathrm{MI} = \frac{1}{2}\sum_{i=1}^{n} \log\left(1 + \lambda_i^2\right), \tag{10}$$

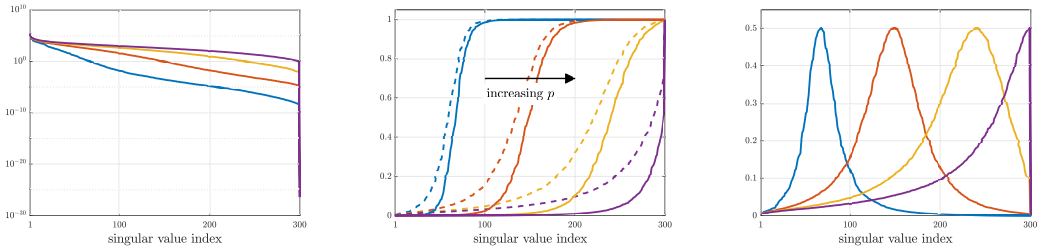
with λ_i the singular values of Eq. 9. This quantity decreases during (converging) iterations, flattens out and eventually increase slightly. The point where the measure reaches a minimum level is associated with stationarity in that the eigenvalues λ_i^2 does not change, meaning $\mathbf{Y}_j$ and thus $\mathbf{T}_j$ does not change either. Hence the stopping rule was formulated as if $\mathrm{MI}_j > \mathrm{MI}_{j-1}$, iterations are terminated. This stopping criterion is complemented by a maximum allowed number of iterations.

4 Selection of DAW length

The amount of data included in an analysis step can be a challenge for ensemble-based methods. In the application with seismic gathers, the amount of waveform data available to infer the subsurface parameters from is massive, and one needs to assimilate this carefully to avoid problems. First, the ensemble linearization becomes a limiting assumption when a large time horizon of data are integrated, and then the iterative procedure might not converge. Second, a large dataset can lead to over-fitting as in under-estimation of the state uncertainty. In what follows, methods for automatic selection of the DAW are presented. At an analysis cycle, performed as explained in the previous section, the data partition is set though h_k, and each analysis cycle should be proceeded by a search for an adequate DAW. In the case with seismic data, the data have a traveltime reference, and it is natural to order the data partitioning by time intervals. This choice of DAW, by the nature of the reflection signal, also limits the influence region from the subsurface parameters to a limited depth range and thereby regularizes the problem of estimating the parameters.

4.1 Spectral analysis

Throughout the following analysis, the singular value spectrum of the normalized observation anomalies in Eq. 9 is used, along with the resulting eigenspectrum of key quantities. The displays in Fig. 2 shows these eigenspectra. These result are obtained by using a sequence of time windows with identical start and increasing length, and thereby an increasing the number of observations p. These plots provide a useful background visualization of the key findings



(a) λ_i^2 (b) $\left(1+\lambda_i^2\right)^{-1}$ and $\left(1+\lambda_i^2\right)^{-1/2}$ (c) $\lambda_i\left(1+\lambda_i^2\right)^{-1}$

Fig. 2: Eigenvalue distributions as a function of singular values λ_i for; (a) information matrix, (b) inverse Hessian (full) and transform matrix (dashed), (c) weight factor for projected innovations, as a function of increasing data dimension p. Colors are for same p across plots.

presented next: the distribution in Fig. 2a is the information matrix eigenspectrum λ_i^2 from which eigenvalues of the Hessian $1+\lambda_i^2$, inverse Hessian $\left(1+\lambda_i^2\right)^{-1}$ and the transform matrix eigenvalues $\left(1+\lambda_i^2\right)^{-1/2}$ are derived. Their eigenbasis are the right singular vectors of $\mathbf{R}^{-1/2}\mathbf{Y}$, as explained in Section 2. While the specific shape of the λ_i^2 distribution and hence also of the derived spectra will depend on the prior ensemble characteristics and configuration of the forward model, the general behavior is well represented by these displays.

Inserting the singular value decomposition (9) into the gradient and Hessian expressions (8) along with the normalized innovations $\Delta\tilde{\mathbf{y}} = \mathbf{R}^{-1/2}(\mathbf{y}-\tilde{\mathbf{y}})$, the control vector search direction can be written as

$$\begin{aligned}\Delta\mathbf{w} &= -\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{\mathrm{T}}\left(\mathbf{w}-\mathbf{V}\boldsymbol{\Sigma}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\Delta\tilde{\mathbf{y}}\right)\\ &= \sum_{i=1}^{n}\mathbf{v}_i\left(\frac{-\mathbf{v}_i^{\mathrm{T}}\mathbf{w}}{1+\lambda_i^2}+\frac{\lambda_i(\mathbf{u}_i^{\mathrm{T}}\Delta\tilde{\mathbf{y}})}{1+\lambda_i^2}\right).\end{aligned} \tag{11}$$

The update direction is explicitly within the span of right singular vectors, the Hessian eigenbasis. The coefficient of each singular vector $\mathbf{v}_i$ is composed of (projected) contributions from the prior (via $\mathbf{w}$) and from the likelihood (via $\Delta\tilde{\mathbf{y}}$). The update vector can be split into the vector components $\Delta\mathbf{w}_x = \mathbf{V}\mathbf{a}$ and $\Delta\mathbf{w}_y = \mathbf{V}\mathbf{b}$, with ith coefficient $\mathbf{a}_i = -(\mathbf{v}_i^{\mathrm{T}}\mathbf{w})(1+\lambda_i^2)^{-1}$ and $\mathbf{b}_i = \lambda_i(\mathbf{u}_i^{\mathrm{T}}\Delta\tilde{\mathbf{y}})(1+\lambda_i^2)^{-1}$ respectively, which will be used later.

The reduced rank property of the forecast error covariance means that when the number of observations increases, the ensemble covariance cannot properly represent the variability in the (ensemble) waveform data. As a result, the forecast error covariance is underestimated and observations are given more weight in the analysis. The expression (11) shows how this enters through the weighting of the two components. When the amount of data is increased, the influence of the prior component on the update $\Delta\mathbf{w}$ is reduced as more and more $(1+\lambda_i^2)^{-1} \to 0$ (Fig. 2b). Where the weighting of the normalized innovations projections $\lambda_i(1+\lambda_i^2)^{-1}$ centers around $\lambda^2 = 1$ with a maximum of 0.5 (Fig. 2c), the increase of data shifts the weighting of the likelihood component towards higher singular components and generally increases the likelihood contribution to the analysis.

Referring to the information matrix, Zupanski et al. [18] divide the eigenspectrum into a region of signal, where $\lambda_i^2 \geq 1$ and the forecast errors are larger than observational noise, and a noise region with $\lambda_i^2 < 1$ and the forecast errors are smaller than the noise. According to this, the weighting curves $(1+\lambda_i^2)^{-1}$ and $\lambda_i(1+\lambda_i^2)^{-1}$ have relation to the influence from observation and from prior, to the left and right of $\lambda_i^2 = 1$ respectively. This division of eigenvalues into regions of influence also applies to the transform matrix. The eigenvalues $(1+\lambda_i^2)^{-1/2}$ affects the amount of shrinkage of the preconditioning ensemble (7), where components with eigenvalues approaching 1 contributes to maintained ensemble spread. They are associated with states where observations are non-informative and the prior should dominate. The opposite occurs for components approaching 0, which are related to the influence from data, providing information to improved estimation and reduced uncertainty.

The singular value index i for which $\lambda_i^2 \simeq 1$ is close in value to another information theoretic measure referred to as the degree of freedom for signal d_{s}. This measure can be viewed as the influence of observations to the analysis [11, 3], and can be expressed as $d_{\mathrm{s}} = \mathrm{trace}\left(\mathbf{I}_m - \mathbf{P}^{\mathrm{a}}(\mathbf{P}^{\mathrm{f}})^{-1}\right)$ where these error covariances are full-rank versions. The corresponding quantity when considering the ensemble subspace can be found as $d_{\mathrm{s}} = \sum_{i=1}^{n}\lambda_i^2\left(1+\lambda_i^2\right)^{-1}$ [18]. The complementary degrees of freedom for noise can be interpreted as a constrain imposed by the prior.

From this interpretation of the eigenvalues, it makes some sense to base regularization of the initial update $\Delta\mathbf{w}_0$ on the data window length and therefore dimension p, via an anal-

ysis of the effect p has on various elements of IEnKS. Two strategies with different angle on (11) that seeks to address what an appropriate data dimension is follows next.

4.2 DAW selection strategy

When entering the kth analysis cycle, the data partition expressed through h_k must be set so the determination of this is the first step when a cycle starts. The focus is on choosing a partition size that regularizes the initial control variable updates. The DAW selection searches for an end time $t_{k,L}$ of a time window so that the partition is $h_k : y(t,\cdot), t \in [t_{k,0}, t_{k,L}]$, that is, reflection data for times $t_{k,0}$ to $t_{k,L}$ for all offsets. The exclusion of data in the mute region is implicit in h_k. The subsequent analysis cycle then starts at $t_{k+1,0} = t_{k,L} + \Delta t$. Simplicity is sought by using a minimum of user-supplied tuning parameters. The approach involves repeated use of matrix decomposition, but this cost is negligible compared to the benefit of conducting an analysis cycle with adequate block sizes for the seismic data.

The principle of the algorithm is to keep extending the time window until a criterion is no longer respected. The extension is done via a step size in traveltime. The approach requires that the forecasted ensemble of simulated data gathers are available for increasing arrival times, but because the reflectivity method synthesizes the seismograms over the full recording time at once, the step-wise increase of time window does not require additional forward model ensemble evaluations. For efficiency a larger step size for the time window growth can be used initially. Once the criterion is no longer satisfied (or the end time of the gather is reached), the current window increase is reverted, the step size decreased and the loop repeats until a minimum step size is reached. So the window length search is independent of the expansion acceptance criterion.

Weight criteria

When entering an analysis cycle, the control vector is initialised by $\mathbf{w}_0 = \mathbf{0}$ and the basis coefficients $\mathbf{b}$ determines the first update. With increasing DAW size, the smallest singular values increases in value and the index at which $\lambda_i^2 \approx 1$ increases. Consequently, the region of peak weighting projections moves to higher indices, thus amplifying singular components with more high-frequency content, relative to components with lower indices. Basing the search direction $\Delta\mathbf{w}_0$ on projections of $\Delta\tilde{\mathbf{y}}$ onto higher-frequency $\mathbf{u}_i$-components can over-fit to noise rather than structural eigenbasis components of the observation error covariance. This could render the mean update $\mathbf{x}_1$ highly varying with values that are unacceptable for the forward solver, or so far from the true profile that the error covariances evaluated around this new mean are useless.

A reasonable location of this (asymmetrically) bell-shaped curve is argued to be of importance as an objective for determining a properly selected DAW size. The focus of this first criterion is thus to balance the weighting curve $\lambda_i(1+\lambda_i^2)^{-1}$ over the index interval $[1,n]$. As explained, the point $\lambda^2 = 1$ has a central location on different curves that can be useful for controling the distribution. The relative location of this point is therefore taken as the key variable in the window expansion criterion, in order to balance the prior and likelihood weighting curves in Eq. 11 from the start of the iterations. Setting $i_\mathrm{C} = \max\left\{i \,|\, \lambda_i^2 \geq 1; i = 1,\ldots,n\right\}$ the weighting curve criterion is expressed by

$$\frac{n - i_\mathrm{C}}{i_\mathrm{C}} \leq \beta, \tag{12}$$

where $\beta = 1$ means $i_\mathrm{C} \approx n/2$. The ratio parameter β must naturally be positive, and for $\beta > 1$ the window size selection will be smaller than for $\beta = 1$ and vice versa for $\beta < 1$.

Norm criteria

Another approach to selecting the data partition size, still using an objective of balancing contributions from prior and likelihood. This alternative approach is also based on the control variable update direction (11), but the focus of argument is shifted from the weighting curves to the prior and likelihood vector components. With $\Delta\mathbf{w} = \Delta\mathbf{w}_x + \Delta\mathbf{w}_y$ all that is known is that the inequality $\|\Delta\mathbf{w}\| \leq \|\Delta\mathbf{w}_x\| + \|\Delta\mathbf{w}_y\|$ holds, using the standard 2-norm $\|\mathbf{z}\| = \left[\sum_i |z_i|^2\right]^{1/2}$. The second strategy to select the DAW size is based on the criteria

$$\frac{\|\Delta\mathbf{w}_x\|}{\|\Delta\mathbf{w}_y\|} \leq \beta, \tag{13}$$

where again β is a preset threshold parameter. This indirectly sets a bound on the ratio $\|\Delta\mathbf{w}\|/\|\Delta\mathbf{w}_y\|$ and consequently regularizes the likelihood component contribution to the update vector. As mentioned an analysis cycle is initiated with $\mathbf{w}_0 = \mathbf{0}$ and $\mathbf{T}_0 = \mathbf{I}$, and the first update direction is solely determined by the likelihood contribution. Seeing these initial values as the imposition of a standard normal distribution as prior on $\mathbf{w}_0$, a Monte Carlo estimate of a fictious prior component vector $\widehat{\mathbf{a}}$ is generated. For a given DAW selection with associated set of singular components $\{\lambda_i, \mathbf{u}_i\}$, a large batch of B samples $\mathbf{w}^b \sim \mathcal{N}(0, \mathbf{I})$ is used to form

$$\widehat{\mathbf{a}}_i = \frac{1}{B}\sum_{b=1}^{B}\left|\frac{-\mathbf{u}_i^\mathrm{T}\mathbf{w}^b}{1+\lambda_i^2}\right|, \tag{14}$$

which subsequently sets $\|\Delta\mathbf{w}_x\| = \|\widehat{\mathbf{a}}\|$ used to evaluate the criteria Eq. 13. The Monte Carlo expectation is taken on the norm argument instead of the norm due to Jensen's inequality. The absolute value operator of the 2-norm is taken within

the estimator (14) as otherwise the expectation would average out to nearly a zero vector.

At the initial window expansion (as described in Section 4.2), the smaller observation dimension makes the fictious prior component norm dominate. Continuing the window expansion with increasing p, enlarges the likelihood component norm and eventually this dominates, consistent with the evolution of the weight curves in Fig. 2b and 2c. In contrast to the weight curve approach, this does include the information from the innovations, which makes it more adaptive.

The actual $\Delta\mathbf{w}_x|_{j=1}$ will still be zero as $\mathbf{w}_0 = \mathbf{0}$ but the aim is that at the following iteration, the magnitude of the prior component will be in the vicinity of that of the likelihood component and thereby reduce the risk of observations dominating the update on behalf of the prior constrainment.

5 Numerical example

Results of applying the IEnKS to seismic data inversion are presented next. First, the seismic data acquisition design is outlined, followed by the prior model description, and then results of the adaptive DAW approach for sequential inversion. This section uses the norm criterion with a ratio parameter of $\beta = 1$, while a result using the weight criterion is discussed in 6.1.

5.1 Description of setup

The measurement configuration consists of a 100 receiver locations, at offsets distributed in the range 50 m to 5 km with a uniform spacing of 50 m. The source is located in the top layer at 5 m below the top surface, which has the boundary condition of a free surface. This top layer of 500 m depth has fixed properties. The source time signal is a fifth order Butterworth wavelet with frequency bandpass 2–50 Hz and time sampling is 2 ms. The seismic traces has a limited frequency bandwidth compared to the source signal, where these are generated with a frequency content 5–32 Hz, with linear in- and out-tapering from 5–7 Hz and 30–32 Hz. The gather data up to 4 s is used for the inversion, excluding data in a mute region defined by normal move-out in the top layer. The total number of data points is $\sim 10^5$.

Data from a processed well log are used as the true subsurface model $\mathbf{m}^{\mathrm{t}}$ and using this as input for the forward model a simulated data set is considered the true seismic CMP gather, with a sample of measurement noise added. A constant noise level $\mathbf{R} = \sigma_{\mathrm{e}}^2\mathbf{I}$ is used in the measurement model, where the noise variance σ_{e}^2 is set to have a signal-to-noise ratio of 13 dB with respect to a reference signal power. This reference signal power is set as an averaged power in the time interval 1–3 s and offset range 0–3 km of the true seismic gather. As the amplitude of the seismic signal decays with time, this means a very non-uniform signal to noise ratio will be present in the data to be assimilated.

Larger ensemble size had a tendency to make the system unstable. The reason is growth of the largest eigenvalues of the information matrix during iterations. This propagates into a corresponding largest eigenvalue of the inverse transform matrix, so that observation anomalies were upscaled unreasonably causing problems for the used SVD routine. This is a numerical issue and was handled by clipping eigenvalues of the transform matrix below a certain threshold which propagates into its inverse. This approach is exactly the same as applied in [14].

5.2 Specification of prior

As the ensemble mean is a linear combination within the ensemble subspace, the trends and smoothness specified in the prior structure influence the ability to form combinations of sufficient variability to fit the true underlying profile of elastic parameters.

The prior ensemble is here specified by samples from a multivariate Gaussian distribution of the log-elastic parameters. This initial distribution for the ensemble is defined through a mean vector and a covariance matrix. The log-units do not have the most intuitive domain for prior specification, and instead, linear depth trends for the mean and standard deviation are set in the physical domain of the three elastic parameters. Using the relation between arithmetic moments of normal and log-normal distributions, these trends are mapped into normal-domain mean $\boldsymbol{\mu}_x$ and standard deviation $\boldsymbol{\sigma}_x$. A cross-correlation structure between the three parameters and a spatial correlation must also be specified. Using a separable structure, the final covariance matrix is

$$\operatorname{cov}\mathbf{x} = \boldsymbol{\Sigma}_x = \operatorname{diag}(\boldsymbol{\sigma}_x)\left[\begin{bmatrix} 1 & \eta_{ps} & 0 \\ \eta_{ps} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \otimes \boldsymbol{\Gamma}\right]\operatorname{diag}(\boldsymbol{\sigma}_x)\,,$$

where the spatial (depth) correlation structure $\boldsymbol{\Gamma}$ is taken as a Matérn function of order $3/2$ with a range parameter such that correlation is 5 % at 500 m distance. The cross-correlation between velocities is $\eta_{ps} = 0.5$, and $\otimes$ is the Kronecker product.

Initially, the ensemble consists of n independent samples from this Gaussian model. For the benchmark case presented here, an ensemble size of $n = 300$ is used. In the discussion part, ensemble sizes of $n = 150$ and $n = 600$ are also studied for comparison.

The support of the prior ensemble when mapped to the log-normal domain, is presented in Fig. 3. This displays the ensemble median and the span of 90% empirical coverage of the prior ensemble.

5.3 Results

The gather data were limited to the time interval 0.6 s–4 s. By running the algorithm this interval was partitioned into 5 blocks, as displayed in Fig. 4. The time lengths of these

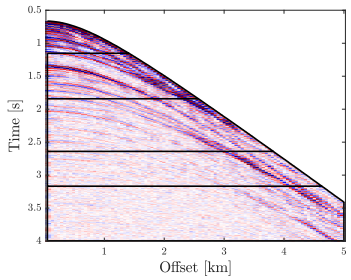


Fig. 4: Partition into DAWs of seismic gather data.

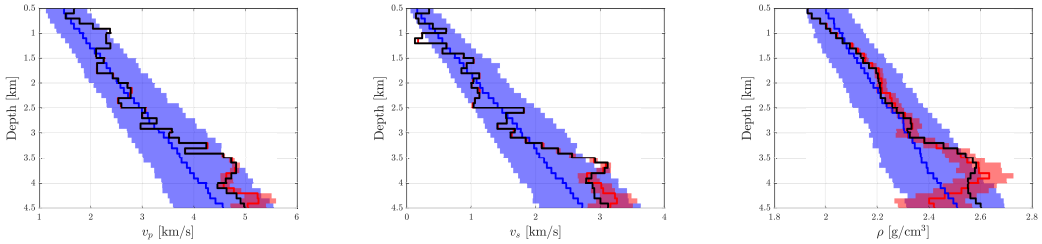
windows were 556, 690, 798, 528 and 828 ms, with corresponding number of data points 4247, 13 635, 25 304, 22 125 and 40 809 .

The resulting posterior ensemble is displayed in Fig. 3, along with the truth and the prior ensembles. This shows that acoustic and shear velocities are estimated well down to around 4 km depth, whereas density is only estimated well down to 3.5 km. Generally, the density estimate is less accurate than that of velocities, consistent with the expected reduced sensitivity of the waveform data misfit to perturbations in density.

The assimilation statistics are presented further in Fig. 5. Here, the estimation bias $\left|\widehat{\mathbf{m}} - \mathbf{m}^{\mathrm{t}}\right|$ is shown. The physical state estimate $\widehat{\mathbf{m}} = \mathrm{median}\left(\exp(\mathbf{E}^{\mathrm{a}})\right)$ is seen to correlate well with the (marginal) standard deviation from the analysis ensemble. For the shear velocity in Fig. 5b, estimation results stand out in two areas at shallower depths (at 0.6 km–0.8 km and 1 km–1.2 km depth). In these areas, the bias deviates largely and this can also be seen in the standard deviation where there is a local increase. Both cases are associated with very low values of true shear velocity. The estimation is there seen to be more difficult, possibly associated with challenging parts in the forward model. The consistent correlation between the (unknown) estimation bias and the ensemble spread shows that an indicative quantification of the estimation uncertainty can indeed be extracted from the ensemble solution.

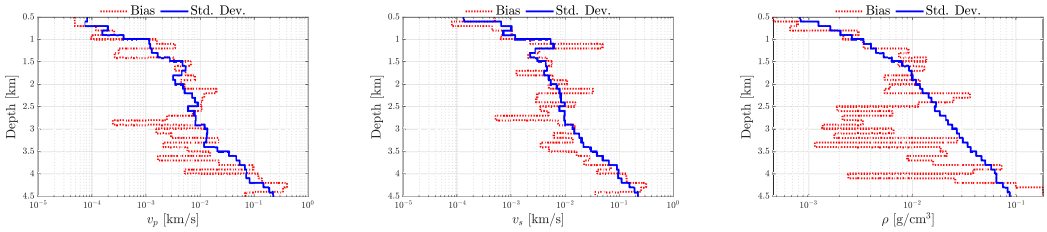
Figure 6 shows how the sequential estimation method proceeds, by displaying the ensemble (marginal) standard deviation for the three parameter types over the course of assimilation cycles. The sensitivity to the three parameter types is also somehow visible from these plots. Each analysis update reduces *P*-wave uncertainty to slightly larger depths than *S*-wave velocity and density. Hence, given a window of data, the estimation of acoustic velocity occurs slightly further in depth than for the other two parameters. This is to be expected due to the higher acoustic wave speed. Not visible from these plots is the order of estimation during iteration, where the acoustic velocity parameter also gets estimated first followed by shear wave velocity and density.

Figure 7 shows the iteration history of the objective function, the mutual information measure used as stopping criteria, and norms of control vector and its update, for each of the five DAWs. The objective function consistently reaches a stationary level faster than the other measures. The objective is dominated by the data misfit, and it flattens out when continued iteration does not update the ensemble mean $\mathbf{x}_j$ in the observed region. Still, the changes below the observed region could be substantial, where the analysis mean ideally should not be far from the prior mean. Hence, much of the later effort of iterating does not contribute to reduce the data misfit, but rather to focusing the analysis update to the relevant parameters in the observed region. This is reflected in the continued change in magnitude of control variable $\mathbf{w}$ and its update, which continues long after a stationary level of data misfit is observed. The MI measure also reaches a



(a) Acoustic velocity (b) Shear velocity (c) Density

Fig. 3: Ensemble and truth; blue is prior, red is posterior, black is truth.



(a) Acoustic velocity (b) Shear velocity (c) Density

Fig. 5: Estimation statistics.

stationary level later than the objective function, but the (iteration) onset of this flattening out correlates much better with $\|\mathbf{w}_j\|$ converging, than with data misfit. Which support the choice of using this measure as stopping criterion. Not shown here is the evolution of the degrees of freedom for signal measure d_s, mentioned in Section 4.1. Similarly to mutual information, it is based on the eigenvalues λ_i^2 and its iterative evolution is very similar to that of mutual information. The d_s indicates the proportion of $\lambda_i^2 \geq 1$ and the evolution of the information matrix spectrum during iterations. Similarly to mutual information, a stationary level means further iterations will no significant update to them.

Another apparent feature is the large dependency of relative difference between initial and final level of objective value, on the overall signal-to-noise ratio within the DAW. Seismograms will always have decaying amplitude with traveltime and the measurement error might not decay in a similar manner, so this issue will generally be present. In Fig. 7a the first DAW has much larger difference than the others as measurement noise is so relatively low. In constrast, the data of fifth assimilation cycle is so contaminated with noise that the data misfit has very little reduction, but nevertheless contributes to better estimation of parameters in the 3.5–4 km depth range.

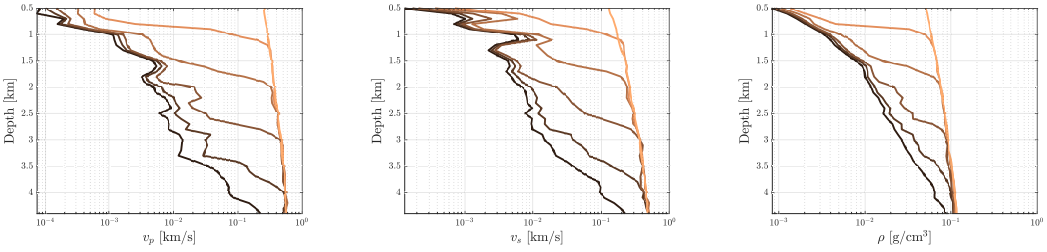
6 Discussion

The challenges of non-unique solutions to the inverse problem are discussed first, continuing the numerical example using the weight criterion rather than the norm criterion. Then a replicate study is presented for both weight and norm criteria, comparing stability, block sizes and the number of forward evaluations

6.1 Challenge of non-uniqueness

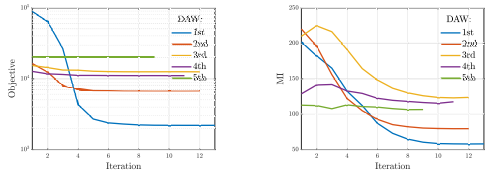
The example presented previously used the norm component approach to finding the DAW lengths. Using the same sample of prior ensemble and measurement noise, the inversion was performed using the weight curve strategy instead, with a ratio parameter $\beta = 1$. This approach generally choses larger windows thus making it more exposed to entering another mode of the posterior distribution.

Figure 8 shows the results after the third assimilation cycle, which indicates divergence and exemplifies a case of misestimation. The observable region starts at around 2.5 km depth, and the top of this region is well estimated. Down the observed region around 3 km, the divergence starts. The data



(a) Acoustic velocity (b) Shear velocity (c) Density

Fig. 6: Ensemble standard deviation over analysis cycles. Order is from lightest (initial ensemble) to darkest (final analysis).



(a) $J(\mathbf{w}_j)$ (b) MI_j

(c) $\|\mathbf{w}_j\|$ (d) $\|\Delta\mathbf{w}_j\|$

Fig. 7: Iterative history of (a) objective function, (b) mutual information and Euclidean norm of (c) $\mathbf{w}$ and (d) $\Delta\mathbf{w}$, over the 5 DAWs.

time window covers 2.326–3.184 s with 34052 data points. This DAW is larger than both the third and fourth DAW in Section 5.3.

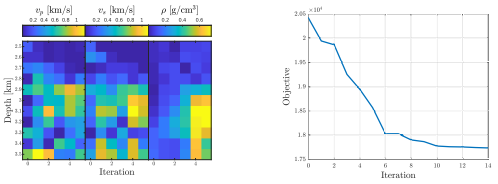
Below the depth where a wrong local mode is found, the mean is highly spurious. The mechanism driving the misestimation in this case is the ensemble linearization of the density gradient/tangent linear model. The cross-covariances between waveform data and densities are much more susceptible to being poorly estimated, i.e. "spurious correlations", than for the velocities. The density effect on the reflected waveform amplitudes is more obscure, and acts in combination with the velocities. Ambiguity in the estimation of density sensitivity can lead to an update direction $\Delta\mathbf{w}$ that points towards a local and erroneous mode. And the chance of this occuring increases with the DAW length. Once a local mode is discovered through the control vector $\mathbf{w}$, the mean $\mathbf{x}$ below that local misestimation will diverge. A false mode has $\widehat{\rho} > \rho^{\mathrm{t}}$ and velocities $\widehat{v}_{p,s} < v^{\mathrm{t}}_{p,s}$ or vice versa in some localised depth region. In which "direction" the false mode is estimated is seemingly a question of the position of the prior mean $\bar{\mathbf{x}}^{\mathrm{f}}$ of the assimilation cycle.

To illustrate this, a closer look at the course of iterations is displayed in Fig. 9. Focusing on the depth range 2.5 km–3.5 km, Fig. 9a shows the estimation bias for each of the elastic properties over the course of the first 5 iterations. The top 2-3 layers are seen to be well estimated within the first few iterations, whereas divergence clearly takes place in layer 4 or 5 and downwards. The density bias indicates that the onset of misestimation is from the very first iteration, starting at depth around 3.2 km and that it indeed is density that drives the estimation divergence. The objective function, Fig. 9b, shows that the data misfit is reduced while the mean is updated towards an erroneous local solution. So the minimization problem is converging in the sense of reducing the data misfit, just at the wrong solution.

The norm criterion seeks to limit the DAW so that $\|\Delta\mathbf{w}_{x,j}\|$ and $\|\Delta\mathbf{w}_{x,j}\|, j = 2$ are of comparable size in the hope that this is a good start for stable iterations. In comparison, the weight criterion has no notion of this. Figure 10 shows the norm components for this divergent 3rd DAW, along with the corresponding components from the 4th assimilation cycle in Section 5.3. It is not a fair comparison, as the previous section's 4th cycle is shorter with fewer data points, but it highlights an observation that is fairly consistent across encountered examples of divergence and exemplifies a characteristic of convergent versus divergent solutions. The main difference is the continued dominance of $\|\Delta\mathbf{w}_{y,j}\|$ when the estimation is diverging. While the norm criterion starts out with a slightly lower $\|\Delta\mathbf{w}_{x,2}\|$ than $\|\Delta\mathbf{w}_{y,2}\|$, the following iterations has a largest prior component magnitude until they equalize, which is the point where $\|\mathbf{w}\|$ reaches a stationary level (Fig. 7c). Contrary for the divergent weight criterion, the likelihood component keeps dominating until they equalize. From our experience, the pattern is that the relation between these vector norms in the first iterations, say 2 to 4 or 5, determines whether the estimation is converging or not.

We speculate whether monitoring the course of the update vector's components' magnitude could effectively be used as a running diagnostic. A diagnostic that indicates divergence with the potential to stop iterating and restart the assimilation cycle with a shorter data window. The monitoring does not come for free though, as the left singular vector $\mathbf{u}_i$ must be available for calculating $\Delta\mathbf{w}_y$ while not strictly necessary for various component of IEnKS as such. Yet, this could be outweighed by the possible robustness added to an inversion routine

As a final remark on the issue of non-uniqueness and misestimation of a gradient towards a false local mode. As the general observed picture is as seen in this case, where the top of the observed region is well estimated but the mean updates towards a local mode further down, we thought that dampening the update could help by the mean moving less (at deeper depths), giving an opportunity to re-evaluate the sensitivities when the top of the observed region had been accounted for. To dampen the update step, the principle of Multiple Data Assimilation (MDA) was applied. A sequence of MDA iterations with its inflation of the observation error covariance matrix, was used for a fixed number of initial iterations. The sequence choice of inflation factors was based on a geometric serie. The MDA error inflation changes both the update direction and dampen its magnitude, and it was thought that it could downplay the contributions from observations to the control vector that locked the mean state in a



(a) $|\widehat{\mathbf{m}}_j - \mathbf{m}^t|$ (b) $J(\mathbf{w}_j)$

Fig. 9: Over iterations: (a) the estimation bias $|\widehat{\mathbf{m}}_j - \mathbf{m}^t|$ for the first 5 iteration for each elastic attribute and $j = 0$ is prior estimate, and (b) the objective function.

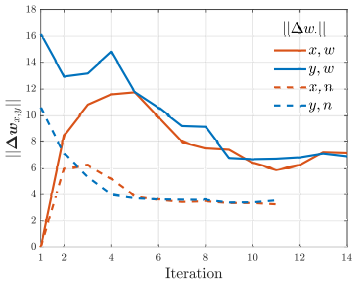


Fig. 10: Norm over iterations of vector components $\Delta\mathbf{w}_{x,j}$ and $\Delta\mathbf{w}_{y,j}$. In legend, 'w' (weight) refers to 3rd DAW in Fig. 8, while 'n' (norm) refers to 4th DAW of Fig. 3.

wrong mode. But the results showed no effective improvement as it did not guarantee against misestimation. This might be due that MDA inflates the error for all data points equally, whereas it probably would be benefial to have inflated them differently, in order to downplay data at later time points within the window. This could be archived though covariance (R-) localization, by upscaling the observation error at later time instances and reducing this inflation gradually over the initial iterations. How the upscaling should be distributed across time and offsets is the question and a complicated one, and this approach has not been pursued any further here.

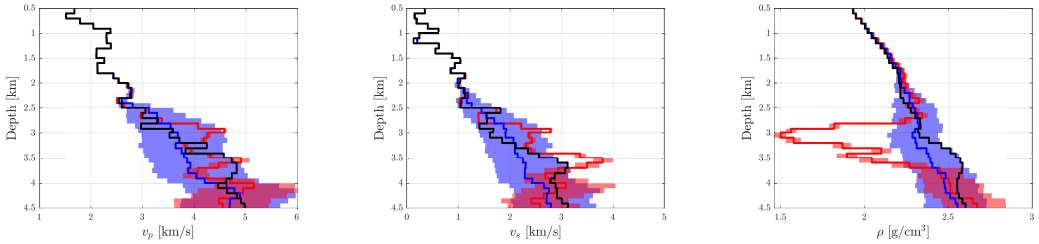
6.2 Replicate study

The example presented so far was for a particular sample of prior ensemble and measurement noise. If these were resampled and the inversions done again, another outcome is obtained. In order to examine the general robustness of the strategies, the parameter β and dependency on the ensemble size, repeated estimation trials are performed. Independent replicate trials that randomize the initial prior ensemble and the additive noise in the synthetic measurement data, are used to evaluate estimation performance. The statistical model and forward model configuration are kept fixed, so the results are in light of those. For each of the ensemble sizes used, a batch of 20 samples are used across the β parameter variation and the strategies. For each strategy, the configurations are combinations of three ensemble sizes $n = (150, 300, 600)$ and $\beta = (3/2, 1, 3/4)$. The larger $\beta = 3/2$ is a slightly more conservative choice with shorter time span of windows and smaller β increases the window lengths.

Each replicated estimation is accepted or rejected. If the solver was not able to compute with the given model input, the replicate is naturally rejected. Otherwise, to classify a posterior as an acceptable estimation, the state subset $\mathbf{z} = \log \mathbf{v}_P|\text{depth} \leq 3.5\,\text{km} \in \mathbb{R}^{35}$ is used as a proxy. This is so because the estimation of acoustic velocity will generally be better than for the other two elastic properties, especially for smaller ensemble sizes. As estimation measure we use the Mahalanobis distance of the true $\mathbf{z}^t$ subset to the distribution represented by the posterior ensemble:

$$\text{MD}_t = \left((\mathbf{z}^t - \widehat{\mathbf{z}})^T \mathbf{C}^{-1} (\mathbf{z}^t - \widehat{\mathbf{z}})\right)^{1/2}, \tag{15}$$

where $\widehat{\mathbf{z}}$ is the ensemble mean. The covariance matrix $\mathbf{C}$ is the ensemble sample covariance, but with an important



(a) Acoustic velocity (b) Shear velocity (c) Density

Fig. 8: Prior and analysis ensemble of the 3rd DAW; blue is prior, red is posterior, black is truth.

modification as it will use a truncated eigenbasis that retains only 75% of the total variance. The reason for doing this truncation is to make the distance measure more robust. The eigen-components with the smallest 25% of total variance are associated with the shallower layers and smaller estimation error at these layers between different ensembles, makes the measure more volatile and less useful for this purpose. To set a (per replicate) threshold for the accept/reject classification, each ensemble member is distance-measured as MD_i against the same $(\widehat{\mathbf{z}},\mathbf{C})$, which gives a level of within-ensemble distance. If $\mathrm{MD}_\mathrm{t} \leq \overline{\mathrm{MD}_i} + 4 \times \mathrm{std}(\mathrm{MD}_i)$, it is accepted as a satisfactory solution. Thus each replicate has its own threshold value. No false positive, i.e. an accepted divergent solution, was confirmed by visual inspection. On the contrary, especially for the smallest ensemble size considered, some cases could have been judged acceptable but did not pass the classification rule.

According to this rule, the number of accepted runs' amongst the 20 replicates is listed in Table 1 and the average number of resulting windows $\overline{K}$ is listed in Table 2. In the header of these tables, the strategy 'w' and 'n' refers to the weight and norm criteria respectively.

n	150		300		600	
β\Criteria	w	n	w	n	w	n
3/2	11	17	9	20	15	20
1	8	19	8	17	12	17
3/4	4	12	9	13	5	15

Table 1: Number of accepted inversion runs out of 20 replicates.

The general pattern is that the norm strategy is much more robust than the weight strategy, and that estimation performance decays with increasing DAW lengths. A deviation of from are the cases $\beta = 3/2$ versus $\beta = 1$ for $n = 150$, but the cause is more related to ensemble size than to DAW length.

From these results, the emphasis on balancing the update vector contributions from prior and likelihood, is definitely more influential on stable updates than is the weight curve argument. The latter, with its focus on the likelihood vector itself, is not addressing the mechanism that controls the potential spurious update. As the norm approach utilizes information from the actual errors and hence is more adaptive, this increased robustness was pictured.

The expected correlation between shorter windows and estimation stability is present for all ensemble sizes. The norm approach is generally more conservative than its alternative, resulting in a higher number of windows, as Table 2 shows. Where perturbing the β-parameter for the weight approach seemingly does not do much for the number of windows, the difference lies in their distribution, where for $\beta = 3/4$ the last window just becomes shorter (up to the gather end time). The norm criterion on the other hand, is more sensitive to the ratio β setting along with a stronger dependency on ensemble size.

n	150		300		600	
β\Criteria	w	n	w	n	w	n
3/2	7	18	5	8	3	5
1	6	9	4	6	3	4
3/4	6	7	4	5	3	3

Table 2: Average number $\overline{K}$, rounded to nearest integer, of DAWs.

The number of accepted estimation (for n-criterion) are comparable for $\beta \geq 1$ and $n = 300$ and 600, and one could get the impression that there is no benefit of using the largest ensemble size. Especially when considering the total number of forward model evaluations, as listed in Table 3. But the estimation results (not shown here) show that $n = 600$ performs much better than $n = 300$, at estimating density generally and all elastic properties at the bottom 1 km depth. So size does matter for sufficiently resolving the gradients when data has a high level of noise.

The numbers in Table 3 are quite high and this implies and demands parallel ensemble evaluation. The number of evaluation decreases with longer windows, so it is not the case that shorter data windows results in an earlier termination of iterating, sufficient to counterbalance the larger number of windows. In terms of efficiency, the $\beta = 1$ case is preferable, while not as consistent in accepted estimations as $\beta = 3/2$ (for $n \geq 300$). If combined with a monitoring and handling of divergence as described in previous section, the intuitive parameter value of 1 seems as a good choice.

7 Summary and conclusions

In this paper an ensemble-based sequential method for seismic inversion is presented. The iterative ensemble Kalman smoother is the core method that uses the ensemble to evaluate sensitivites, thus no tangent linear model is needed and suitable for black-box forward models.

The approach for assimilating the high-dimensional seismic data builds on a strategy of partitioning the data in windows of traveltime, and the inversion is stable if these windows are selected wisely.

A method for automatically selecting an appropriate data window when entering an assimilation cycle is introduced. The method is based on an analysis of the iterative update to the control variable of the variational problem, and on an interpretation of how this update is influenced by the prior and

n	150		300		600	
β\Criteria	w	n	w	n	w	n
3/2	9122 (327)	15908 (1146)	16766 (1054)	19230 (967)	29120 (903)	32640 (2548)
1	8643 (375)	10318 (762)	15450 (578)	17311 (1111)	25900 (668)	29188 (819)
3/4	8812 (495)	9200 (390)	14866 (400)	16223 (1063)	24000 (1200)	25840 (1350)

Table 3: Total number of forward model evaluation. The table entries is the average over accepted replicates, while in parenthesis is the standard deviation. Numbers are rounded to nearest integer.

likelihood. Two alternative angles of interpretation are presented and their performance evaluated through a repeated trials simulation study. Only one of the alternatives showed robust with respect to estimation performance, the approach based on norms of prior and likelihood vector. This aspect was highlighted and discussed in a comparison of a converging and a divergent estimation.

Even though the motivation for this work is reliable nonlinear elastic inversion, the need to partition the data set and assimilate these sequentially is expected to be present in other types of parameter estimation problems. In the geoscience domain there is for instance potential for similar inversion methods for electromagnetic data, gravimetric data, fiber optical data and ground penetrating radar data, which all involves large-size data, complex physical forward models and static state parameters. As a consequence, the observations from this study might be applicable to and of use in other domains.

Acknowledgements We thank partners of the Uncertainty in Reservoir Evaluation (URE) project at NTNU, and BP for data.

References

1. Asch M, Bocquet M, Nodet M (2016) Data Assimilation: Methods, Algorithms, and Applications. Society for Industrial and Applied Mathematics, Philadelphia, PA, DOI 10.1137/1.9781611974546
2. Bocquet M, Sakov P (2014) An iterative ensemble Kalman smoother. Quarterly Journal of the Royal Meteorological Society 140(682):1521–1535, DOI 10.1002/qj.2236
3. Cardinali C, Pezzulli S, Andersson E (2004) Influence-matrix diagnostic of a data assimilation system. Quarterly Journal of the Royal Meteorological Society 130(603):2767–2786, DOI 10.1256/qj.03.205
4. Carrassi A, Bocquet M, Bertino L, Evensen G (2018) Data assimilation in the geosciences: An overview of methods, issues, and perspectives. WIREs Climate Change 9(5):e535, DOI 10.1002/wcc.535
5. Evensen G (2009) Data Assimilation. Springer, DOI 10.1007/978-3-642-03711-5
6. Fowler A, van Leeuwen PJ (2012) Measures of observation impact in non-gaussian data assimilation. Tellus A: Dynamic Meteorology and Oceanography 64(1):17192, DOI 10.3402/tellusa.v64i0.17192
7. Gineste M, Eidsvik J, Zheng Y (2020) Ensemble-based seismic inversion for a stratified medium. GEOPHYSICS 85(1):R29–R39, DOI 10.1190/geo2019-0017.1
8. Kennett B (2005) ERZSOL3. http://www.spice-rtn.org/library/software/ERZSOL3.html, accessed on 2017-09-01
9. Kennett B (2011) Seismic Wave Propagation in Stratified Media. ANU Press
10. Leeuwen PJV, Cheng Y, Reich S (2015) Nonlinear Data Assimilation, Frontiers in Applied Dynamical Systems: Reviews and Tutorials, vol 2. Springer International Publishing, DOI 10.1007/978-3-319-18347-3
11. Liu J, Kalnay E, Miyoshi T, Cardinali C (2009) Analysis sensitivity calculation in an ensemble kalman filter. Quarterly Journal of the Royal Meteorological Society 135(644):1842–1851, DOI 10.1002/qj.511
12. Raanes PN, Stordal AS, Evensen G (2019) Revising the stochastic iterative ensemble smoother. Nonlinear Processes in Geophysics 26(3):325–338, DOI 10.5194/npg-26-325-2019
13. Rodgers CD (2000) Inverse Methods for Atmospheric Sounding. World Scientific Publishing, DOI 10.1142/3171
14. Sakov P, Oliver DS, Bertino L (2012) An iterative EnKF for strongly nonlinear systems. Monthly Weather Review 140(6):1988–2004, DOI 10.1175/mwr-d-11-00176.1
15. Sheriff RE, Geldart LP (1995) Exploration Seismology, 2nd edn. Cambridge University Press, DOI 10.1017/CBO9781139168359
16. Shirangi MG, Emerick AA (2016) An improved tsvd-based levenberg–marquardt algorithm for history matching and comparison with gauss–newton. Journal of Petroleum Science and Engineering 143:258 – 271, DOI https://doi.org/10.1016/j.petrol.2016.02.026
17. Thurin J, Brossier R, Métivier L (2019) Ensemble-based uncertainty estimation in full waveform inversion. Geophysical Journal International 219(3):1613–1635, DOI 10.1093/gji/ggz384
18. Zupanski D, Hou AY, Zhang SQ, Zupanski M, Kummerow CD, Cheung SH (2007) Applications of information theory in ensemble data assimilation. Quarterly Journal of the Royal Meteorological Society 133(627):1533–1545, DOI 10.1002/qj.123

Paper III

Examples and computer code for seismic inversion using the iterative ensemble Kalman smoother

Michael Gineste & Jo Eidsvik

Internal report, 2020

Examples and computer code for seismic inversion using the iterative ensemble Kalman smoother

Michael Gineste and Jo Eidsvik

August 2020

Abstract

Demonstrations of code for doing seismic inversion with the iterative ensemble Kalman smoother is provided. This kind of inversion is relying on sequential data assimilation over blocks of seismic data. The setup is applied to three types of seismic inversion of increasing complexity: linear seismic traveltime inversion, seismic amplitude versus offset inversion, and seismic common midpoint gather inversion. In each example of inversion, the subsurface model parameters of interest are represented by a layered earth model. Documentation of the main steps of the iterative ensemble Kalman smoother is presented. The code is implemented in MATLAB. The presentation is tutorial in its kind, with pedagogical demos for each case and comparison with other methods for doing seismic inversion.

1 Introduction

The iterative ensemble Kalman smoother (IEnKS) (see e.g. Bocquet and Sakov, 2014; Asch et al., 2016) is a method for sequential data assimilation, that easily can be applied to estimation problems with static model parameters. Iterative ensemble smoothers have been successfully used in various application domains such as history matching (Chen et al., 2014), fluid contacts estimation (Wang et al., 2010), and wind resources evaluation (Defforge et al., 2019). Ensemble Kalman smoothers has recently caught some interest in seismic data inversion (Gineste and Eidsvik, 2015; Liu and Grana, 2018; Thurin et al., 2019; Gineste et al., 2020), where the challenges include massive data dimensions and non-linear forward models linking the elastic subsurface model parameters with the seismic data. In mathematical terms, the inverse problem is addressed in a Bayesian framework. The focus is on assimilating the seismic data reliably, starting with prior knowledge, in order to assess the posterior distribution of the subsurface parameters of interest.

The core of the IEnKS algorithm is only few lines of code, but it often requires some adjustment for each practical case. In this document case studies for seismic inversion using the IEnKS are highlighted, with code examples implemented in MATLAB.

The purpose of this work is to demonstrate the applicability of the IEnKS method for Bayesian ensemble-based inversion. It can be tested on various kinds of seismic data without too much change in the implementation setup, and also to other data inversion problems by some modifications. Even though the method is exemplified with MATLAB, the elementary coding lines and the pedagogical presentation can hopefully make it relatively easy to transform the scripts to another programming language.

Seismic data are characterized by waveforms that are generated by a source and with measured responses at receivers. The source can be man-made in a controlled setting, say by an air-gun. It can also be less controlled, say by (vibrational) energy generated from cars or a train. The seismic signal could also be induced naturally for instance by a landslide or earthquake. There is hence a rich class of problems which can benefit from the availability of seismic data, and the specific type of seismic data that is acquired depends very much on the situation. Our cases are largely linked with traditional petroleum applications, where the seismic data acquisition is conducted according to an accurately planned source and receiver setup.

In addition to having much variety in the acquisition procedures of seismic data, there are also myriads of processing methods (Sheriff and Geldart, 1995). This means that there are various levels or versions of seismic data even with the same original raw seismic dataset. Higher degree of processing usually indicates more approximations, in the sense that some physical relations are simplified or some parameters are kept fixed to ease the interpretation and analysis of data. With a simpler physical model, the inverse problem tends to be less difficult because the forward model is more tractable or faster to compute, but the solutions could be biased because of processing artefacts in the simplified geophysical modeling assumptions.

Here, the implementation and testing of IEnKS for seismic inversion is summarized for three common levels of seismic data processing:

1. The simplest case represents **seismic traveltime (TT)** data. In the processing this often entails some manual picking of seismic waveform data, where the traveltimes are extracted from the larger amplitudes. In this way, the processed traveltimes capture the largest signals in the seismic data, and the following analysis does not care about the additional amplitude information or the information content in the subsequent waveforms.

2. The next case is that of **seismic amplitude versus offset (AVO)** data which entails processing amplitude data to a common traveltime in gathers of similar angles. Processed data consists of series of amplitude data over traveltimes, for each lateral seismic data gathering location. One often process and analyze three gather stacks (near, mid and far angles), giving tri-variate time series.

3. The third case has less processing as it only aligns the source and receiver data to a **reflection (Refl)** common mid point (CMP) dataset. Processed amplitude data then has traveltime on the first axis and receiver-source offset on the second axis. This method has much less alignment or averaging associated with it than the seismic AVO analysis.

This document is structured as follows: Section 2 provides the background on Bayesian inversion and presents IEnKS as a sequential data assimilation method, with implementation details for the considered demo cases of seismic inversion. Section 3 outlines the case on seismic traveltime data. Section 4 outlines the case on seismic AVO data. Section 5 outlines the case on seismic reflection data.

2 Seismic inversion by sequential assimilation

The variable of interest is denoted by state vector $\mathbf{x} \in \mathbb{R}^n$. In all examples, n is directly proportional to the number of depth layers (multiplied by 3 for the elastic case where interest lies in (acoustic) P-wave velocity, (shear) S-wave velocity and density in each layer). The prior probability density function (pdf) for these variables is defined by $p(\mathbf{x})$ and represented by a Gaussian distribution denoted by $p(\mathbf{x}) = \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ represents the prior mean and $\boldsymbol{\Sigma}$ is the prior covariance matrix.

The likelihood of data $\mathbf{y} \in \mathbb{R}^p$ is defined by a forward model and zero-mean independent additive noise terms with a Gaussian error structure. This model is denoted by $p(\mathbf{y}|\mathbf{x}) = \text{Normal}(h(\mathbf{x}), \mathbf{R})$, where the forward model h specifies the mean of the seismic response and $\mathbf{R}$ is the noise covariance matrix. An assumption of diagonal or block diagonal $\mathbf{R}$ is made in the following.

The posterior model for the variables is computed by Bayes' rule:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x}) \ . \tag{1}$$

If the forward model $h(\mathbf{x})$ is linear, this posterior is Gaussian with known mean and covariance matrix. If the forward model in non-linear, there is no closed form solution to the posterior.

Data can be split in n_k disjoint subsets or data assimilation windows (DAWs), and this gives $\mathbf{y}_k$, $k = 1, \ldots, n_k$, such that $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_{n_k})$. The data in block k has size p_k, so that $\mathbf{y}_k = (y_{k,1}, \ldots, y_{k,p_k})$. The likelihood for block subsets of data is denoted by $p(\mathbf{y}_k|\mathbf{x}) = \text{Normal}(h_k(\mathbf{x}), \mathbf{R}_k)$, where the forward model h_k specifies the mean seismic response for the k-th DAW, and $\mathbf{R}_k$ is the noise covariance matrix in this DAW.

The posterior model for the variables is computed by sequentially assimilating data in the DAWs $k = 1, \ldots, n_k$. From Bayes' rule:

$$p(\mathbf{x}|\mathbf{y}_1, \ldots, \mathbf{y}_k) \propto p(\mathbf{y}_k|\mathbf{x})\, p(\mathbf{x}|\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}) \ , \tag{2}$$

for $k = 2, \ldots, n_k$. In this expression, the assumption of block diagonal covariance matrix for the forward model noise terms is used, so the data in the disjoint blocks are conditionally independent. At step $k = n_k$, the posterior distribution is achieved, with conditioning on all seismic data.

2.1 Iterative ensemble Kalman smoother

The IEnKS is a hybrid method that combines a variational approach and an ensemble approach. The variational aspect is that the state estimate, the mean, is found as the

state minimizing a functional whose solution is the maximum a posteriori (MAP) state. The ensemble is used to evaluate sensitivities of this functional, and the solution can be found iteratively using some gradient-based optimization scheme. Here, the Gauss-Newton update is used. At the optimal solution, the Hessian provides information on how to transform the forecast ensemble deviations around its mean, to analysis deviations. Combined with the newly found analysis mean, this form the analysis ensemble.

The IEnKS is in the current context used for the joint subsurface model and with blocks of seismic data. For non-linear problems, it tends to be more stable to include subsets of the data in a sequential routine, because it improves the ensemble approximation of the tangent linear model (Gineste et al., 2020).

The ensemble-based approximations of the pdfs are adjusted in the sequential assimilation scheme, using data in subsequent DAWs. This setting is illustrated in Figure 1. At step k, the ensemble is conditioned on observed data $\mathbf{y}_k^{\text{o}}$. This means that the forecast ensemble is changed to an analysis ensemble. This is then used as forecast for step $k+1$, and the recursion continues for all DAWs $k = 1, \ldots, n_k$.

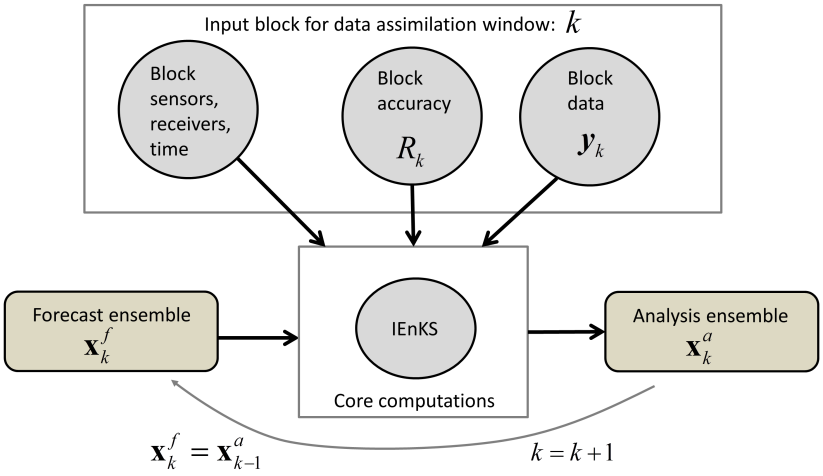


Figure 1: Illustration of blocking scheme, and forecast and analysis update.

Formally, the members (columns) of a forecast ensemble $\mathbf{E}^{\text{f}}$ are denoted $\mathbf{x}_i^{\text{f}}$, $i = 1, \ldots, n_e$, where n_e is the ensemble size. Initially, they are sampled from the prior model

$p(\mathbf{x})$. The state estimate is the ensemble mean:

$$\bar{\mathbf{x}}^{\mathrm{f}} = \frac{1}{n_e}\sum_{i=1}^{n_e}\mathbf{x}_i^{\mathrm{f}} \,. \tag{3}$$

An associated $n \times n_e$ anomaly matrix $\mathbf{X}_{\mathrm{f}}$ has ith column

$$[\mathbf{X}_{\mathrm{f}}]_i = (n_e-1)^{-1/2}\left(\mathbf{x}_i^{\mathrm{f}} - \bar{\mathbf{x}}^{\mathrm{f}}\right) \,. \tag{4}$$

This normalized anomaly matrix is the square root of the forecast error covariance estimate, which becomes

$$\mathbf{P}^{\mathrm{f}} = \mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\mathrm{T}} \,. \tag{5}$$

The ensemble mean and anomaly matrix re-assembles the forecast ensemble as

$$\mathbf{E}^{\mathrm{f}} = \bar{\mathbf{x}}^{\mathrm{f}}\mathbf{1}^{\mathrm{T}} + (n_e-1)^{1/2}\mathbf{X}_{\mathrm{f}} \,. \tag{6}$$

Predictions of data are generated from the forecast parameter ensemble. Denote the size p_k synthetic data by $\mathbf{y}_i^{\mathrm{f}} = h_k(\mathbf{x}_i^{\mathrm{f}})$, $i = 1,\ldots,n_e$, with mean $\bar{\mathbf{y}}^{\mathrm{f}} = \frac{1}{n_e}\sum_{i=1}^{n_e}\mathbf{y}_i^{\mathrm{f}}$. A square root of the forecast covariance in observation space is defined as the $p_k \times n_e$ observation anomaly matrix $\mathbf{Y}_{\mathrm{f}}$:

$$[\mathbf{Y}_{\mathrm{f}}]_i = (n_e-1)^{-1/2}\left(\mathbf{y}_i^{\mathrm{f}} - \bar{\mathbf{y}}^{\mathrm{f}}\right) \,.$$

The ensemble updating, the analysis step, transforms the forecast anomalies and centers these around a new mean that is found by an optimization procedure. A transform that fulfills the Kalman covariance update equation can be derived as follows. This covariance update equation with the sample (ensemble) covariance matrices inserted can, using the matrix inversion lemma (Petersen and Pedersen, 2012), be formulated as

$$\mathbf{X}_{\mathrm{a}}\mathbf{X}_{\mathrm{a}}^{\mathrm{T}} = \mathbf{X}_{\mathrm{f}}\left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_{\mathrm{f}}\right)^{-1}\mathbf{X}_{\mathrm{f}} = \mathbf{X}_{\mathrm{f}}\mathbf{T}\left(\mathbf{X}_{\mathrm{f}}\mathbf{T}\right)^{\mathrm{T}} , \tag{7}$$

so by selecting the ensemble transform matrix (ETM) as $\mathbf{T} = \left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_{\mathrm{f}}\right)^{-1/2}$ the analysis square-root covariance can be updated as $\mathbf{X}_{\mathrm{a}} = \mathbf{X}_{\mathrm{f}}\mathbf{T}$. As the square root matrix, the symmetric square root is used (Sakov and Oke, 2008). Using the eigendecomposition $\left(\mathbf{I} + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_{\mathrm{f}}\right) = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$, the unique (inverse) symmetric square root is $\mathbf{T} = \mathbf{V}\mathbf{\Lambda}^{-1/2}\mathbf{V}^{\mathrm{T}}$.

The analysis state is found as a linear combination within the subspace spanned by the forecast anomalies, $\mathbf{x}^{\mathrm{a}} \in \left\{\bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w} \,|\, \mathbf{w} \in \mathbb{R}^{n_e}\right\}$, where the weight vector $\mathbf{w}$ is optimized in a variational problem. The cost function to find the optimal weights is defined by

$$J(\mathbf{w}) = \tfrac{1}{2}\left[\mathbf{y}_k^{\mathrm{o}} - h_k\left(\bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w}\right)\right]^{\mathrm{T}}\mathbf{R}_k^{-1}\left[\mathbf{y}_k^{\mathrm{o}} - h_k\left(\bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w}\right)\right] + \tfrac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}, \tag{8}$$

which is minimized using an iterative method where the gradient ∇J and approximative Hessian $\mathbb{H}$ are computed from the ensemble. With the optimal $\mathbf{w}^{\mathrm{a}}$ and $\mathbf{T}_{\mathrm{a}} = \mathbb{H}^{-1/2}\big|_{\mathbf{w}^{\mathrm{a}}}$, the analysis ensemble is then

$$\mathbf{E}^{\mathrm{a}} = \mathbf{x}^{\mathrm{a}}\mathbf{1}^{\mathrm{T}} + (n_e-1)^{1/2}\mathbf{X}_{\mathrm{a}} = \bar{\mathbf{x}}^{\mathrm{f}}\mathbf{1}^{\mathrm{T}} + \mathbf{X}_{\mathrm{f}}\left(\mathbf{w}^{\mathrm{a}}\mathbf{1}^{\mathrm{T}} + (n_e-1)^{1/2}\mathbf{T}_{\mathrm{a}}\right) \,. \tag{9}$$

78

The structure of an IEnKS analysis cycle is summarized in Algorithm 1.

Algorithm 1 Iterative Ensemble Kalman Smoother

Require: Prior ensemble $\mathbf{E}^{\mathrm{f}} = \mathbf{E}^{\mathrm{f}}_k$
  $j = 0$, $\mathbf{w}_j = \mathbf{0}$, $\mathbf{T}_j = \mathbf{I}_n$
  $\bar{\mathbf{x}}^{\mathrm{f}} = \mathbf{E}^{\mathrm{f}}\mathbf{1}/n$
  $\mathbf{X}_{\mathrm{f}} = \left(\mathbf{E}^{\mathrm{f}} - \bar{\mathbf{x}}^{\mathrm{f}}\mathbf{1}^{\mathrm{T}}\right)/\sqrt{n-1}$
  repeat
    $\mathbf{x}_j = \bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w}_j$
    $\mathbf{E}_j = \mathbf{x}_j\mathbf{1}^{\mathrm{T}} + \sqrt{n-1}\mathbf{X}_{\mathrm{f}}\mathbf{T}_j$
    $\bar{\mathbf{y}} = h_k(\mathbf{E}_j)\mathbf{1}/n$
    $\mathbf{Y} = \left[h_k(\mathbf{E}_j) - \bar{\mathbf{y}}\mathbf{1}^{\mathrm{T}}\right]\mathbf{T}_j^{-1}/\sqrt{n-1}$
    $\nabla J = \mathbf{w}_j - \mathbf{Y}^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathbf{y}^{\mathrm{o}}_k - \bar{\mathbf{y}})$
    $\mathbb{H} = \mathbf{I} + \mathbf{Y}^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{Y}$
    $\mathbf{w}_{j+1} = \mathbf{w}_j - \mathbb{H}^{-1}\nabla J$
    $\mathbf{T}_{j+1} = \mathbb{H}^{-1/2}$
    $j = j + 1$
  until termination criteria met
  $\mathbf{E}^{\mathrm{a}}_k = \bar{\mathbf{x}}^{\mathrm{f}}\mathbf{1}^{\mathrm{T}} + \mathbf{X}_{\mathrm{f}}\left(\mathbf{w}_{j-1}\mathbf{1}^{\mathrm{T}} + \sqrt{n-1}\mathbf{T}_j\right)$

This algorithm is the core for all examples, but the setup and splitting of data differs somewhat for each type of seismic inversion, and this is explained in the relevant sections for seismic traveltime, AVO and reflection inversion that follow. This algorithm is tailored to the current purpose of ensemble-based seismic inversion. A general version of the method with detailed discussion is provided by Asch et al. (2016).

2.2 Description of code

Accompanying this report is a MATLAB code base to provide practical examples of seismic inversion using the IEnKS method. The code folder has the structure as described in Table 1, where the generic IEnKS functionality resides in the `ienks` folder and the different seismic examples in the `seis-model` folder.

Folder	Description
`ienks/`	main IEnKS function, helper and plot functions.
`seis-model/`	has the three subfolders:
`TT/`	files for traveltime inversion.
`AVO/`	files for AVO inversion.
`Refl/`	files for CMP inversion using the reflectivity method.

Table 1: Folder structure

Filenames within each subfolder have folder name prefix, so that files in `ienks/` folder are prefixed with `ienks_` and similarly for the seismic models. The IEnKS files

are listed in Table 2. The main function used by the examples is `ienks_cycle` which performs an assimilation cycle. This function, which implement the Algorithm 1, is listed in Appendix A.

Filename	Description
`ienks_cycle`	main function, performs an assimilation cycle
`ienks_make_matrices`	helper function, forms the transform matrices and inverse Hessian.
`ienks_zeropad`	helper function.
`ienks_plot_ensemble`	plot ensemble during iterations.
`ienks_plot_iter`	plot cost function and norms during iterations.
`ienks_plot_spectra`	plot eigenspectra during iterations.

Table 2: Content of IEnKS folder.

The seismic models share a common structure and file naming, as listed in Table 3. The examples are provided by their respective `demo` script and the remaining functions are the forward models and functionality to extract the data block subsets. Additional setup for each module is provided in the respective sections that follow.

Filename	Description
`demo`	main demo script and runs the example.
`forward`	evaluates the (ensemble) forward model.
`make_data_block`	makes linear indices used to extract the data for an assimilation window.
`make_prior`	samples a prior ensemble.

Table 3: Common file naming across seismic models.

Download

The package can be downloaded from `https://folk.ntnu.no/michaegi/seismic-ienks/seismic-ienks.zip`.

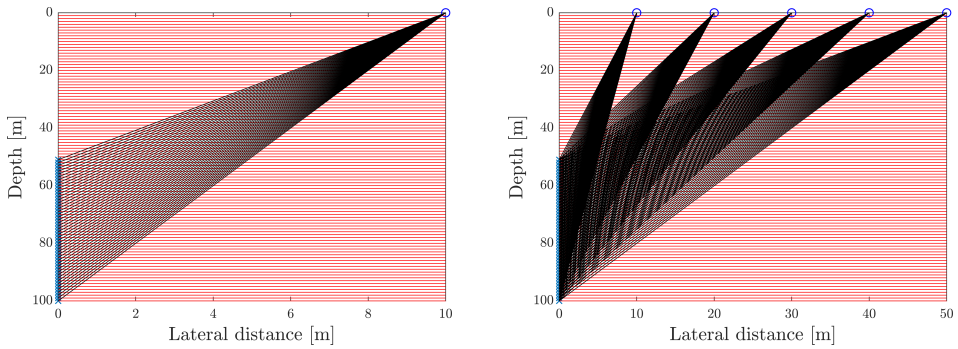
3 Traveltime tomography

3.1 Case description

The problem described here is motivated by tomographic imaging of the subsurface using seismic traveltime data. The situation that is tested resembles that of vertical seismic profiling, where the source(s) is on the surface while the receiver(s) is in a borehole.

A layered earth model with n subsurface layers of 1m thickness is considered. The measurements consist of traveltimes from the source locations to the receiver locations.

There are n_r receivers located at the $n - n_r + 1, \ldots, n$ bottom layers in the subsurface model. The case is illustrated in Figure 2 with $n = 100$ layers and receivers from layer 51 to 100. In the left display there is $n_s = 1$ source at 10m. In the right display there are $n_s = 5$ sources at 10, 20, 30, 40 and 50 m. In this application an assumption of



(a) One source (b) Five sources

Figure 2: The layered earth model (red) is observed through seismic traveltime data. The geometric acquisition design has a source (blue circle) at the surface and receivers in a borehole (blue crosses). The measurements are traveltimes [msec] from the source to receivers assuming straight ray-paths (black). (a) One source location. (b) Five source locations.

straight line ray-paths is used. These are sketched in black in Figure 2. This involves an approximation because the paths would bend according to Snell's law giving more complex ray-paths (Slawinski et al., 2000). For modest gradients in the subsurface velocity properties, this assumption is still reasonable for many applications.

The goal of seismic traveltime inversion for this problem, is to learn the subsurface properties from the traveltime data. For this setting, the seismic wave slowness (the inverse of velocity) is often used as the variable of interest. The slowness in layer $j = 1, \ldots, n$ is denoted by x_j, and the length vector of slownesses is $\mathbf{x} = (x_1, \ldots, x_n)$, sorted from shallow to deep. The traveltime from source s to receiver r is then described by the following forward model

$$\text{traveltime}_{rs} = h_{rs}(\mathbf{x}) = \frac{\sum_{j=1}^{d_r} x_j}{\cos(\theta_{rs})}, \tag{10}$$

where d_r is the known depth layer of receiver r and θ_{rs} is the angle between this receiver and source location s. The total data size is $p = n_r n_s$. Notably, the forward model in (10) is linear in the slownesses, and hence the characterization of these subsurface parameters from the processed traveltime data represents a relatively simple inversion problem.

The prior model for slownesses is Gaussian with mean $\mu_j = E(x_j) = 0.5 - 0.001j$, $j = 1, \ldots, n$, standard deviation $\sigma = 0.05$ and correlation $\text{Corr}(x_j, x_{j'}) = (1+\eta h_{j,j'}) \exp(-\eta h_{j,j'})$,

$\eta = 0.1$, for distance $h_{j,j'} = |j - j'|$. These expressions define the joint prior model as $p(\mathbf{x}) = \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

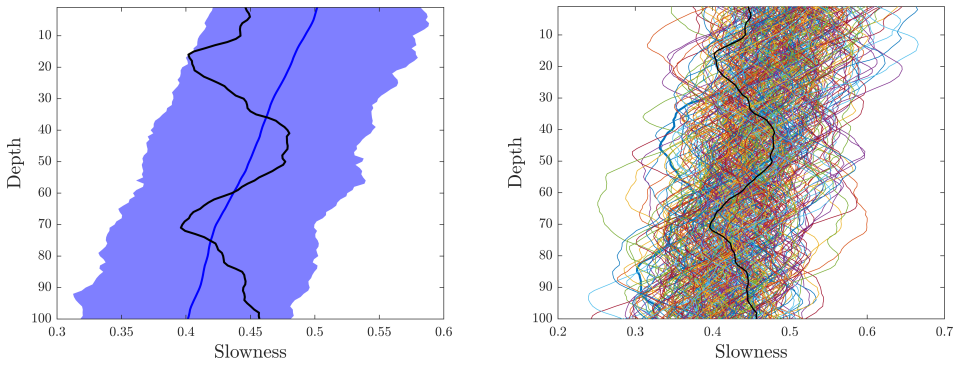
Extending the forward model in (10) to a statistical model, the traveltime data y_{rs}, $r = 1, \ldots, n_r$, $s = 1, \ldots, n_s$, are described by the following model:

$$y_{rs} = h_{rs}(\mathbf{x}) + \epsilon_{rs}. \tag{11}$$

Here, $\epsilon_{rs} \sim N(0, \tau^2)$ are assumed independent and $\tau = 0.5$.

3.2 Simulation examples

For the ensemble-based approach, $n_e = 300$ independent realizations from the prior model for slowness is generated. They are denoted $\mathbf{x}_i^{\text{f}}$, $i = 1, \ldots, n_e$. For this purpose a Cholesky factorization $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$ is used in the generation of correlated variables, and then adding the mean $\boldsymbol{\mu}$: $\mathbf{x}_i^{\text{f}} = \boldsymbol{\mu} + \mathbf{L}\mathbf{z}_i$, where $\mathbf{z}_i \sim N(0, \mathbf{I})$. Figure 3b shows the sampled prior ensemble members. The true synthetic slowness profile is illustrated as well (black).



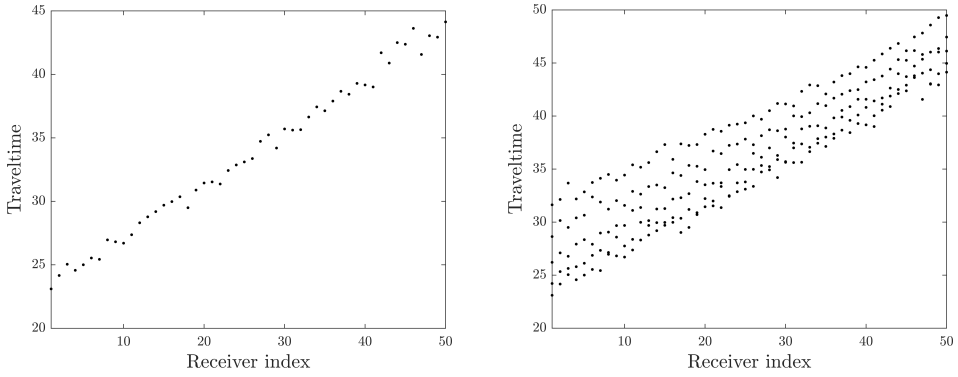
(a) Ensemble coverage (b) Ensemble members

Figure 3: Prior ensemble with truth in black; (a) shows the ensemble mean as blue line and its 90 % coverage displayed as shaded area, while (b) shows the ensemble members.

Figure 3a shows the marginal 90 percentiles made up of the prior ensemble. Hence, this display illustrates plotting of all ensembles or via the empirical 5 and 95 percentiles of slowness in each layer, obtained by sorting the ensemble values.

Figure 4 shows traveltime data computed from the true synthetic model. This is for the seismic wave going from source to each of the 50 receivers in the borehole. Here, noise is added to the forward model, according to the likelihood specification. The traveltimes of course increase as the lateral distance to source increases and with the depth of the receivers.

The sequential algorithm goes recursively through all data in the n_k blocks, each of size $p_k = n_{rk} n_s$ (n_s sources and n_{rk} receivers in each block k). The blocking of data for the case with five sources is illustrated in Figure 5. Here, ten blocks are used, and



(a) One source (b) Five sources

Figure 4: Traveltime data from source to receivers using the source-receiver designs in Figure 2. (a) One source at the surface. (b) Five sources at the surface.

in each block there are then five receivers with traveltime data from each of the five sources, giving $p_k = 25$.

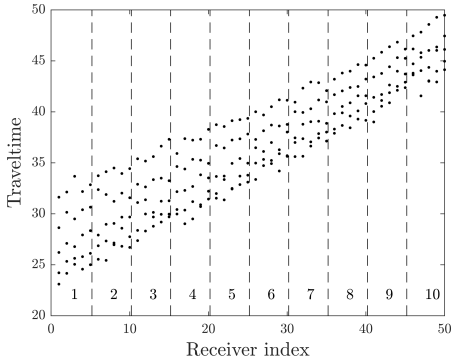
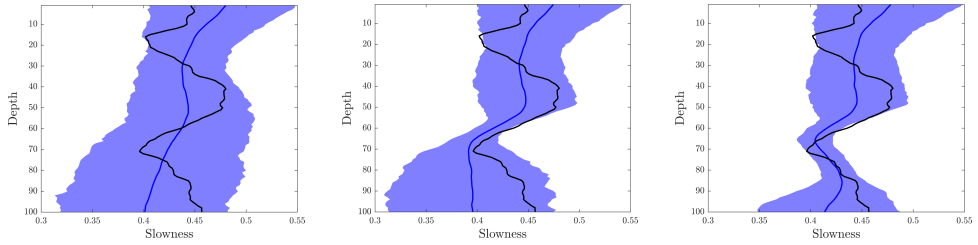


Figure 5: Data and blocking strategy for the situation with five sources and ten blocks.

Next, the IEnKS algorithm for the sequential data assimilation is demonstrated for this problem. In Figure 6, the ensemble representation is visualized at a couple of intermediate steps of the sequential analysis. The displays demonstrate the top-down nature of the analysis due the block ordering. When the first few blocks have been assimilated, the slowness in layers 50-70 is rather accurate, while there is still much uncertainty for the deeper layers. The ensemble-based solution covers the truth rather well. When all data have been assimilated, there is much more uncertainty for shallow depths where the information only contains the average properties.

For this linear forward model, Gaussian prior and measurement noise assumptions, the posterior solution for slowness is Gaussian. This can hence be computed directly, or



(a) 1. cycle (b) 4. cycle (c) 8. cycle

Figure 6: IEnKS solution after 1, 4, and 8 data assimilation cycles.

via a recursive formulation as in a Kalman filter. Here, this exact solution is compared with the ensemble-based approach. For a Kalman filter solution, starting with $E(\mathbf{x}) = \boldsymbol{\mu} = \boldsymbol{\mu}_0$ and $\mathrm{Var}(\mathbf{x}) = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, update the conditional mean $\boldsymbol{\mu}_k = E(\mathbf{x}|\mathbf{y}_1, \ldots, \mathbf{y}_k)$ and covariance $\boldsymbol{\Sigma}_k = \mathrm{Var}(\mathbf{x}|\mathbf{y}_1, \ldots, \mathbf{y}_k)$ as follows:

- For each receiver and each source, form a length n vector of active slowness variables and use angle θ_{rs} between receiver and source to adjust the distance in the layer, i.e. $\mathbf{g}_{rs} = (1, \ldots, 1, 0, \ldots, 0)/\cos(\theta_{rs})$. Stack these rows on top of each other for the relevant block, to get a $p_k \times n$ matrix $\mathbf{G}_k$.
- Kalman gain $\mathbf{K}_k = \boldsymbol{\Sigma}_{k-1}\mathbf{G}_k^T/(\mathbf{G}_k\boldsymbol{\Sigma}_{k-1}\mathbf{G}_k^T + \tau^2)$ (size $n \times p_k$ vector).
- Update mean $\boldsymbol{\mu}_k = \boldsymbol{\mu}_{k-1} + \mathbf{K}_k(\mathbf{y}_k^o - \mathbf{G}_k\boldsymbol{\mu}_{k-1})$.
- Update covariance matrix $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_{k-1} - \mathbf{K}_k\mathbf{g}_k\boldsymbol{\Sigma}_{k-1}$.

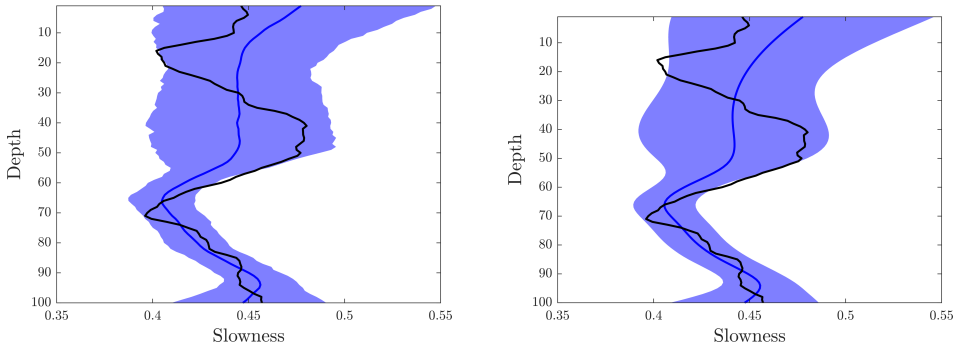
The resulting exact distribution can then be compared at each DAW of the IEnKS approach. The solution is shown only at the last step. This is done by plotting the mean and the marginal 90 percent uncertainty intervals in Figure 7. In this display the ensemble-based solutions (7a) is very similar to the exact one (7b).

Further, a comparison of the marginal cumulative distribution functions of the exact and approximate solutions is carried out, using the energy score as follows: In a layer j, denote the exact Gaussian posterior cumulative distribution function by $F(x) = \Phi(x; \mu_{j|y}, \sigma^2_{j|y})$, while the ensemble-based solution is denoted $\hat{F}_j(x) = \frac{1}{n_e}\sum_{i=1}^{n_e} I(x_{i,j} < x)$. The energy score between the two distributions is then

$$E_j = \int (F_j(x) - \hat{F}_j(x))^2 dx,$$

and in our setting is summed over all layers $j = 1, \ldots, n$ and over 2000 replicate runs of different ensembles and synthetic data.

The results are summarized in Table 4. The ensemble-based approximation is clearly more accurate when the ensemble size increases. When it goes to infinity, it is exact in this case. This holds for both data gathering schemes ($n_s = 1$ and $n_s = 5$). The



(a) Ensemble Kalman filter solution (b) Exact Kalman filter solution

Figure 7: Traveltime solution by ensemble approach and exact.

	$n_s = 1$			$n_s = 5$		
n_e	20	100	500	20	100	500
$n_k = 1$	0.160	0.022	0.004	0.169	0.017	0.003
$n_k = 10$	0.158	0.022	0.004	0.165	0.017	0.003

Table 4: Energy score results over 2000 replicate data sets and ensembles for comparing the IEnKS and the exact Gaussian solution for different source numbers (n_s), ensemble sizes (n_e) and number of blocks (n_k).

uncertainty in the posterior distribution is smaller for $n_s = 5$, so the larger discrepancy for small ensemble sizes ($n_e = 20$) means that the ensemble representation is slightly less accurate when there is more data. In this linear example, there is only a very minor effect of the blocking scheme, visible for the case with small ensemble size.

4 Seismic AVO data inversion

4.1 Case description

The case presented next is common in the setting with seismic AVO inversion. It is assumed that the raw seismic data has been processed to seismic AVO data for a set of angle gathers and in the traveltime domain. This is usually available as traces for every inline and crossline in the seismic acquisition grid design. In this presentation only one trace is considered. It is common to have three angles as is also considered here: near, mid and far angles.

The inverse problem is then to predict the elastic properties as a function of traveltime (depth), conditional on the seismic AVO data. With the three elastic properties in each layer, the size of the variable of interest $\mathbf{x} = (x_1, \ldots, x_n)$ is then $n = 3n_t$, where n_t is the number of traveltimes considered. A Gaussian prior model is assigned to the logarithm of the elastic properties. This is defined via a separable model assuming constant trivariate

marginal distribution for the three elastic properties at all traveltimes, and a correlation function for modeling the dependence at various depths. This means that the prior model has a constant mean (μ_1, μ_2, μ_3) and constant 3×3 covariance matrix $\mathbf{\Sigma}_0$, for the three elastic properties at all traveltimes. The joint pdf for all n elements is then constructed via Kronecker products such that.

$$p(\mathbf{x}) = \text{Normal}((\mu_1, \mu_2, \mu_3) \otimes 1_{n_t}, \mathbf{\Sigma}_0 \otimes \mathbf{\Gamma}), \tag{12}$$

where 1_{n_t} is a length n_t vector of 1 entries and the matrix elements $\Gamma_{j,j'}$ are defined by a parametric correlation function over the traveltime distances $|j - j'|$, see e.g. Buland and Omre (2003). The specific parametric form of the covariance entries is similar to what is presented in that paper.

Next, the forward model associated with seismic AVO inversion is described. The amplitude data at an interface varies as a function of angle of incidence and as a function of the changes in elastic properties in the depth profile. In addition, it is coupled by a wavelet inherited from the seismic source signature.

The established model for seismic AVO data at an interface is due to Zoeppritz equations, with an accurate approximation for weak contrasts and moderate angles as developed by Aki and Richards (2002). In this formula, the seismic response at an interface is affected by the contrast in the medium properties. Here, let α_{j-1} and α_j denote the logarithm of the pressure wave velocity at the top and bottom layer of an interface. Similarly, β_{j-1} and β_j are the logarithm of the shear wave velocity, and ρ_{j-1} and ρ_j are the logarithm of the density in these layers. At the interface between layer $j-1$ and j, and for angle stack θ (near, mid or far), Aki-Richards formula says that the seismic AVO reflectivity (r) is given by the following forward model:

$$\text{r}_{j,\theta}(\mathbf{x}) = a_{\alpha,j,\theta}(\alpha_j - \alpha_{j-1}) + a_{\beta,j,\theta}(\beta_j - \beta_{j-1}) + a_{\rho,j,\theta}(\rho_j - \rho_{j-1}), \tag{13}$$

$a_{\alpha,j,\theta} = \frac{1}{2}(1 + \tan(\theta)^2)$, $a_{\beta,j,\theta} = -4\zeta^2 \sin(\theta)^2$, $a_{\rho,j,\theta} = \frac{1}{2}(1 - 4\zeta^2 \sin(\theta)^2)$, and the non-linearity enters via the velocity ratio

$$\zeta = \bar{V}_s/\bar{V}_P\,, \quad \bar{V}_P = \frac{1}{2}\left[\exp(\alpha_j) + \exp(\alpha_{j-1})\right]\,, \quad \bar{V}_S = \frac{1}{2}\left[\exp(\beta_j) + \exp(\beta_{j-1})\right].$$

The seismic AVO measurements are modeled as a convolution of these interface reflectivities by a wavelet operator. The wavelet, denoted here by $w_{j'',\theta}$, for angle θ and convolution window times $j'' = -c, -c+1, \ldots, c$ in a length $2c+1$ window around location j, can usually be specified from well data combined with seismic data at a key geological interface. It is often assigned a parametric form, for instance a Ricker (Mexican hat) wavelet, which is used in our implementation. The forward model for the seismic AVO data at depth j and angle θ is then defined as

$$\text{AVO}_{j,\theta} = h_{j,\theta}(\mathbf{x}) = \sum_{j'=j-c}^{j+c} w_{j-j',\theta}\text{r}_{j,\theta}(\mathbf{x}), \tag{14}$$

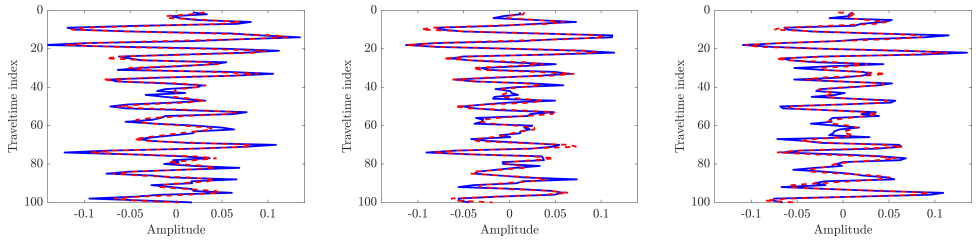
where the wavelet weights $w_{j'',\theta}$ are normalized to sum to 0, for each angle θ. The measurements are

$$d_{j,\theta} = h_{j,\theta}(\mathbf{x}) + \epsilon_{j,\theta}, \tag{15}$$

where $\epsilon_{j,\theta}$ are Gaussian distributed noise terms, and (15) then defines the likelihood model in this setting. Edge problems of the convolution model are handled by wrapping the trace around on a circle (torus).

4.2 Simulation examples

Simulated seismic AVO data for three angles are shown in Figure 8. Here, the Ricker



(a) Near (b) Mid (c) Far

Figure 8: Seismic AVO data for three angles. In blue the (synthetic) measurement data, in red the forward model evaluation of posterior estimate.

wavelet is set to have a bandwidth defined by $c = 10$. The data indicates large amplitude data at the locations with large contrasts in elastic properties.

The IEnKS is run with 5 iterations at each assimilation step. The convergence is plotted in Figure 9. The objective function drops very fast and indicates that the problem is almost linear since there is only minuscule decline in the curves after the first iteration.

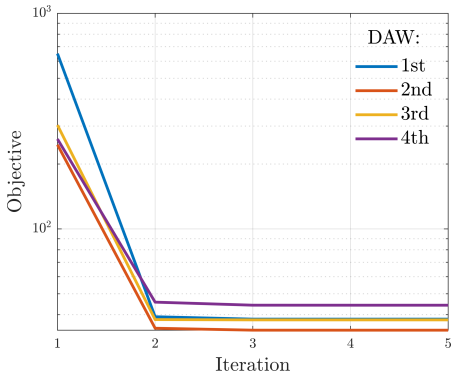


Figure 9: Convergence of cost function $J(\mathbf{w}_j)$ over the 4 DAWs.

Figure 10 shows an IEnKS solution for each of the log elastic parameters after assimilation of data blocks 2 and 4, along with the prior (top). The results clearly show

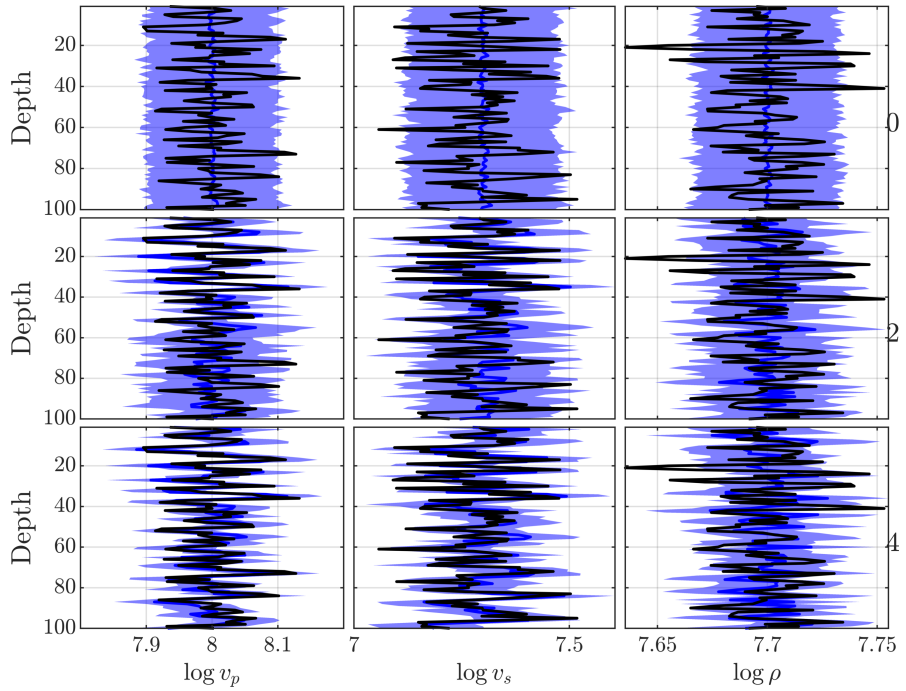


Figure 10: Ensemble solution at different assimilation cycles. Black line is truth, blue line is ensemble mean and shaded area is ensemble 90% percentile. Top row is prior ensemble, middle row after 2 assimilation cycles and bottom row is final posterior.

the top-down idea of the sequential data assimilation procedure, where data from the top influence only the shallower variables.

Next, a comparison is done with other approaches for seismic AVO inversion. A linearized approach Buland and Omre (2003) is established by setting the P-S wave velocity ratio constant. In this approach, the ζ parameter is here fixed at 0.22, and the AVO inversion becomes a linear problem. A comparison with Markov chain Monte Carlo (MCMC) sampler is also done. For this seismic AVO example such comparison has been done previously by Eidsvik and Tjelmeland (2006) and Rabben et al. (2008). Here, a Metropolis–Hastings version of MCMC sampling is implemented. A hybrid version between a Random-Walk proposal and an independent proposal from the linearized approximation is used. A proposed state is accepted or rejected according to the required acceptance probability for each proposal mechanism.

Figure 11 shows the comparison of IEnKS with the linearized solution and the MCMC sampler. The solutions are very similar for this problem because it is so close to linear. The linearized solution tends to have near constant standard-deviations at all depths, which cause slight overestimation of variability at some larger contrasts, e.g. for shear velocity variable at depth 80-90, compared with the non-linear methods.

Another notable difference in Figure 11 is for the density estimates, where the IEnKS solution shows a slightly smaller uncertainty than the other two solutions. This is attributed to spurious correlations between density and the seismic traces. In Figure 8 the predicted seismic data for the posterior state estimate is shown in red, which is seen to correspond very well to the measurement data (in blue). The AVO seismic is generally rather insensititive to the density parameter, thus the larger bias in density estimation has no significant influence on the predicted seismic.

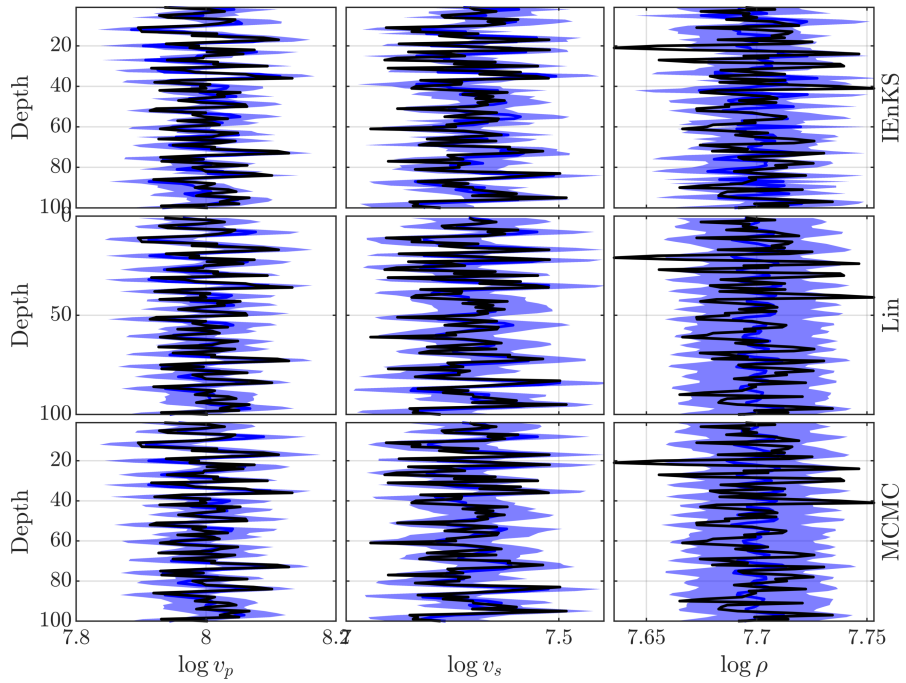


Figure 11: Ensemble-based solution compared with a linearized solution and an MCMC sampler.

5 Seismic reflectivity data inversion

5.1 Case description

The reflectivity method is an approach for solving the elastic wave equation, based on the simplifying assumption of a layered medium, see e.g. Kennett (2011). Under this medium assumption and a plane wave approximation, the response from the stacked layers, i.e. the reflected signal, can be derived analytically in the form of a wavenumber integral. The elastic wave model is thus transformed to the wavenumber-slowness domain, solved therein, and the solution transformed back to time-domain seismograms.

The method builds up a kind of response function for each wavenumber component, a computational procedure that scales with number of layers and the range and discretization of the slowness domain.

Thus, the computing time of the reflectivity method scales with the number of layers so this is kept low in the current setup, to avoid very large computer times. Below a top layer of 500 m thickness with known and fixed parameters, a total depth of 2 km is parameterized into 20 layers, with uniform thickness of 100 m. Each layer has elastic properties of acoustic wave velocity v_p in $\mathrm{km\,s^{-1}}$, shear wave velocity v_s in $\mathrm{km\,s^{-1}}$, and density ρ in $\mathrm{g\,cm^{-3}}$.

The subsurface model $\mathbf{x}$ thus has dimension 3×20 and holds the log-transformed elastic attributes of each layer, $\mathbf{x} = [\log \mathbf{v}_p, \log \mathbf{v}_s, \log \boldsymbol{\rho}]$. The log-transform is to ensure positive physical properties.

The gather collects seismograms at 40 offsets, ranging from 75 m to 3 km with a uniform spacing of 75 m. The source is positioned 5 m down into the first layer, while receiver locations are at 0 m depth.

The seismograms are waveform amplitude data, measurements of vertical displacement. These are synthesized without the direct wave and multiples. The waveforms are sampled at 2 ms and the total length of the seismograms is 8 s. This excess total length is to alleviate issues with time aliasing and to provide sufficient resolution for the frequency integration (Mallick and Frazer, 1987). The frequency bandwidth is limited to 2 Hz–20 Hz.

The prior is specified as a multivariate Gaussian $p(\mathbf{x}) = \mathrm{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Units in the log-domain are not directly intuitive, so linear means and scales in depth are set in the physical domain and mapped into log-domain mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$. Each of the elastic parameters has the same depth correlation structure $\boldsymbol{\Gamma}$, chosen to be a Matérn function of order 3/2, as was used in the AVO example. For this order the correlation function is

$$\boldsymbol{\Gamma}_{j,j'} = (1 + \sqrt{3}h_{j,j'}/\alpha) \exp\left(-\sqrt{3}h_{j,j'}/\alpha\right) ,$$

where $h_{j,j'}$ is distance between layers j and j', and a multiple of 100 m, and the range parameter α is set such that correlation is 5 % at 500 m.

Further, cross-correlations between velocities is included while density is kept independent of the two velocities. Combined, this gives the state prior covariance

$$\boldsymbol{\Sigma} = \mathrm{diag}\,(\boldsymbol{\sigma}) \left[\begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \otimes \boldsymbol{\Gamma} \right] \mathrm{diag}\,(\boldsymbol{\sigma}) ,$$

where $\otimes$ is the Kronecker product.

Code description

Since this example is slightly more complex than the seismic traveltime and AVO, some further implementation details are required. These are outlined next: The CMP exam-

ple uses the reflectivity method from the CREWES[1] MATLAB Toolbox, which can be downloaded from `https://www.crewes.org/ResearchLinks/FreeSoftware/`.
Only the `reflectivity/` folder of this toolbox is needed, and the path to this folder must be added the MATLAB path. The main function called is `reflectivity/reflectivity.m` that calculates the radial and vertical displacement seismograms, and depends on the remaining functions in `reflectivity/`.

Some additional helper functions are required for this seismic model, which are listed in Table 5.

Filename	Description
`refl_init_model`	specifies parameters for the forward model solver and offsets.
`refl_make_model`	formats a state vector into the matrix format expected by the forward model solver.
`refl_make_mute_nmo`	makes (time) indices for each offset that specifies end of mute region. Used by `refl_make_data_block`.

Table 5: Additional content of seismic model folder `Refl`.

5.2 Simulation examples

From the prior distribution, an ensemble of size $n_e = 200$ is sampled plus one extra sample used as truth. This prior ensemble, when exponentially mapped to the log-normal, physical domain, is shown in the top row of Fig. 13.

The synthetic gather data using the true state sample is added noise with constant variance $\mathbf{R} = \sigma_e^2\mathbf{I}$. The noise level $\sigma_e^2 = 10^{-8}$ is relatively benign. Gather data is visualized in Fig. 12. This is in the range 0.6 s to 2.4 s, as used for the inversion. This interval is partitioned into time windows of 0.6 s duration, resulting in 3 assimilation windows. This blocking is depicted with the seismic data in Fig. 12, which also displays a blanked out region which constitutes a mute region. This region specifies gather data that are excluded. This region is defined by a time value for each offset, and data before this time is part of the mute region. The time value is found by normal moveout in the top layer. The traveltime for the first reflection from source to the rth offset, t_r, is

$$t_r^2 = (2(d_1 - z_{\text{source}})/v_{p,1})^2 + (h_r/v_{p,1})^2 \ ,$$

where d_1 and $v_{p,1}$ are top layer thickness and acoustic velocity, z_{source} the source depth (assumed $< d_1$) and h_r the rth offset distance. While the first window is specified to start from 0.6 s, it really starts at ≈ 0.66 s as the muting excludes the beginning.

The resulting posterior ensembles are shown in Fig. 13, while the iterative history of the cost function and the norms of control variables and updates are displayed in Fig. 14. Each assimilation cycle takes 15 iterations, no other stopping criterion is used.

[1]Consortium for Research in Elastic Wave Exploration Seismology, Department of Geoscience, University of Calgary.

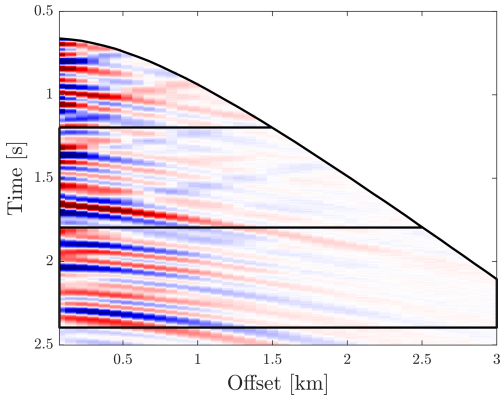


Figure 12: CMP gather with data partitioning.

Fig. 13 shows that the estimation of all three elastic attributes performs very well. The estimation of shear velocity is slightly delayed compared to acoustic velocity due its lower velocity, so within a data block shear waves arrival are from shallower depths compared to acoustic waves. Density estimation has more uncertainty associated with it as sensitivity to this parameter is lower than for the velocities.

Figure 14a shows that the data misfit converges after the first 4 to 6 iterations, so the max limit on iterations could likely be smaller for this example. On the other hand, the control variables continue to be updated after the data misfit has reached a stationary level, as visible from Fig. 14b and 14c. Hence the ensemble mean gets some minor updates at these later iterations.

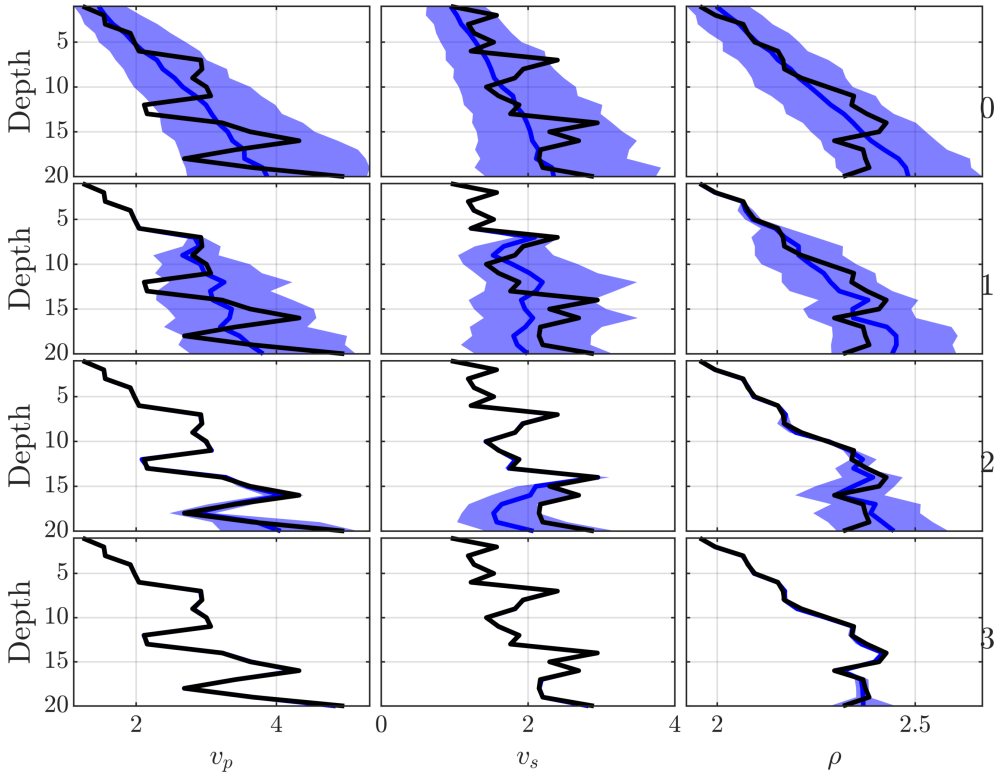
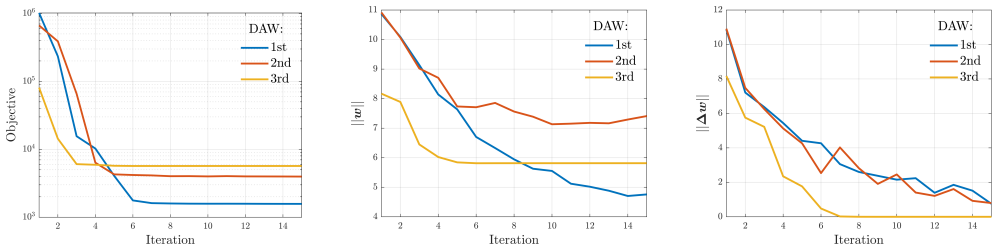


Figure 13: Sequential estimation results. Black is true parameter, blue line is ensemble median and shaded area is empirical 90% coverage. Top row is initial prior ensemble and subsequent rows are posterior over assimilation windows.



(a) Objective $J(\mathbf{w}_j)$ (b) Norm $||\mathbf{w}_j||$ (c) Norm $||\Delta\mathbf{w}_j||$

Figure 14: Iterative history of (a) objective function and Euclidean norm of (b) $\mathbf{w}$ and (c) $\Delta\mathbf{w}$.

Acknowledgments

We acknowledge support from the Norwegian Research Council and industry partners of the URE consortium and GAMES consortium at the Norwegian University of Science and Technology (NTNU) for the financial support (grant no. 294404).

References

Aki, K. and Richards, P. G. (2002). *Quantitative seismology*. University Science Books.

Asch, M., Bocquet, M., and Nodet, M. (2016). *Data Assimilation: Methods, Algorithms, and Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Bocquet, M. and Sakov, P. (2014). An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1521–1535.

Buland, A. and Omre, H. (2003). Bayesian linearized avo inversion. *Geophysics*, 68(1):185–198.

Chen, Y., Oliver, D. S., et al. (2014). History matching of the norne full-field model with an iterative ensemble smoother. *SPE Reservoir Evaluation & Engineering*, 17(02):244–256.

Defforge, C. L., Carissimo, B., Bocquet, M., Bresson, R., and Armand, P. (2019). Improving cfd atmospheric simulations at local scale for wind resource assessment using the iterative ensemble kalman smoother. *Journal of Wind Engineering and Industrial Aerodynamics*, 189:243–257.

Eidsvik, J. and Tjelmeland, H. (2006). On directional metropolis–hastings algorithms. *Statistics and Computing*, 16(1):93–106.

Gineste, M. and Eidsvik, J. (2015). Framework for seismic inversion of full waveform data using sequential filtering. In *Petroleum Geostatistics 2015*, pages cp–456. European Association of Geoscientists & Engineers.

Gineste, M., Eidsvik, J., and Zheng, Y. (2020). Ensemble-based seismic inversion for a stratified medium. *Geophysics*, 85(1):R29–R39.

Kennett, B. (2011). *Seismic Wave Propagation in Stratified Media*. ANU Press.

Liu, M. and Grana, D. (2018). Stochastic nonlinear inversion of seismic data for the estimation of petroelastic properties using the ensemble smoother and data reparameterization. *Geophysics*, 83(3):M25–M39.

Mallick, S. and Frazer, L. N. (1987). Practical aspects of reflectivity modeling. *GEOPHYSICS*, 52(10):1355–1364.

Petersen, K. and Pedersen, M. (2012). The matrix cookbook, version 20121115. *Technical Univ. Denmark, Kongens Lyngby, Denmark, Tech. Rep*, 3274.

Rabben, T. E., Tjelmeland, H., and Ursin, B. (2008). Non-linear bayesian joint inversion of seismic reflection coefficients. *Geophysical journal international*, 173(1):265–280.

Sakov, P. and Oke, P. R. (2008). Implications of the form of the ensemble transformation in the ensemble square root filters. *Monthly Weather Review*, 136(3):1042–1053.

Sheriff, R. E. and Geldart, L. P. (1995). *Exploration seismology*. Cambridge university press.

Slawinski, M. A., Slawinski, R. A., Brown, R. J., and Parkin, J. M. (2000). A generalized form of snell's law in anisotropic media. *Geophysics*, 65(2):632–637.

Thurin, J., Brossier, R., and Métivier, L. (2019). Ensemble-based uncertainty estimation in full waveform inversion. *Geophysical Journal International*, 219(3):1613–1635.

Wang, Y., Li, G., Reynolds, A. C., et al. (2010). Estimation of depths of fluid contacts by history matching using iterative ensemble-kalman smoothers. *SPE Journal*, 15(02):509–525.

A Core IEnKS functions

Listing 1: `ienks_cycle.m`

```
function [mEnsVar_A, varargout] = ienks_cycle(mEnsVar_F, hForward, vObs, mRis, varargin
    )
% IENKS_CYCLE - Performs an IEnKS analysis cycle
%
%
% Input:
% mEnsVar_F - matrix - [nVar, nEns] prior ensemble
% hForward  - handle - to forward model evaluation. Takes as input a [nVar, nEns]
    parameter ensemble and returns an [nObs, nEns] ensemble of observations.
% vObs      - vector - [nObs, 1] of measurements.
% mRis      - matrix - [nObs, nObs] inverse square root of observation error covariance
    .
%
% Output:
% mEnsVar_A - matrix - [nVar, nEns] analysis ensemble
  % The forecast parameter ensemble mEnsVar_F is of size nVar * nEns.
  [nVar, nEns] = size(mEnsVar_F);
  % Initialise the static (for a single analysis cycle) ensemble mean and anomaly
      matrix
  vX_0 = mean(mEnsVar_F,2);
  mX_0 = (mEnsVar_F - repmat(vX_0, 1, nEns)) ./ sqrt(nEns-1);
  % Initial transform matrix and its inverse
```

```
mT = eye(nEns);
mTinv = eye(nEns);
% Initial weights vector
vW = zeros(nEns, 1);
% Iteration counter
jj = 0;
% Flag can be used for other stopping criteria
bStop = 0;
while ((jj < MaxIter) && not(bStop))
  % current iteration mean
  vX = vX_0 + mX_0 * vW;
  % current iteration anomalies
  mE = repmat(vX, 1, nEns)  + sqrt(nEns-1) * mX_0 * mT;
  if bPlot(2)
    % Plot iteration ensemble
    ienks_plot_ensemble(hAxes{2}, jj, mE, vTrue, 0);
  end
  % The function handle hForward maps each ensemble (in columns) to an output of size
      nObs * 1, such that the forecast
  % observation ensemble is [nObs, nEns]
  mEnsObs_F = hForward(mE);
  % Predicted observation; mean of observations
  vObs_F = mean(mEnsObs_F, 2);
  % Observation anomaly matrix
  mY  = ( (mEnsObs_F - repmat(vObs_F, 1, nEns)) * mTinv ) ./ sqrt(nEns-1);
  % Normalized observation anomalies and innovations
  mRY = mRis * mY;
  vRy = mRis * (vObs - vObs_F);
  % Gradient vector
  vJ = vW - mRY' * (mRis * (vObs - vObs_F));
  % % Hessian matrix (NOT USED FOR ANYTHING)
  % mH = eye(nEns) + mRY' * mRY;
  % Objective function
  Jobj = 0.5 * (vRy' * vRy) + 0.5 * (vW' * vW);
  % If ClipLevel is set empty, no ETM clipping is performed.
  ClipLevel = [];
  [mG, mT, mTinv, vS, mU, ~, mV] = ienks_make_matrices(mRY, ClipLevel);
  [dW_X, dW_Y] = ienks_make_update_norms(vW, vRy, vS, mU, mV);
  % Update weight
  dW = -mG * vJ;
  if bPlot(1)
    ienks_plot_iter(hAxes{1}, jj, Jobj, vW, [dW, dW_X, dW_Y]);
  end
  if bPlot(3)
    ienks_plot_spectra(hAxes{3}, jj, vS);
  end
  % Update iteration counter
  jj = jj + 1;
  fprintf(' - iteration %02i/%02i\n', jj, MaxIter);
  % Evaluate other stopping criteria
  bStop = 0;
  if not((jj + 1) == MaxIter) || not(bStop)
```

```
      % Update for next iteration. If the loop is about to exit, this update should not
          be performed as the (w,T) pair used
      % for analysis should correspond.
      vW = vW + dW;
    end
  end % while
  % New analysis mean
  vX = vX_0 + mX_0 * vW;
  % New perturbations matrix
  mX = sqrt(nEns-1) * mX_0 * mT;
  % New analysis ensemble
  mEnsVar_A = repmat(vX, 1, nEns) + mX;
  return;
end % function ienks
```

Listing 2: `ienks_make_matrices.m`

```
function [mG, mT, mTi, vS, varargout] = ienks_make_matrices(mRY, Clip)
% IENKS_MAKE_MATRICES - Finds the inverse Hessian, ensemble transform and inverse
    transform matrices.
%
% Input:
% mRY  - matrix - [nObs, nEns] of normalized observation anomalies
% Clip - scalar - level below which to clip transform matrix egenvalues. Propagates to
    inverse transform as well.
%
% Output:
% mG   - matrix - [nEns, nEns] inverse Hessian
% mT   - matrix - [nEns, nEns] transform matrix
% mTi  - matrix - [nEns, nEns] inverse transform matrix
% vS   - vector - [nEns, 1] of information matrix eigenvalues
  nObs = size(mRY, 1);
  nEns = size(mRY, 2);
  [mU,mS,mV] = svd(mRY,0);
  % Vector of singular values squared, padded with zeros if nObs<nEns.
  vS = ienks_zeropad(diag(mS).^2, nEns);
  %% Vector of diagonal matrices
  % Inverse Hessian
  vG = 1./(1 + vS);
  % Transform matrix and its inverse. Clipping of eigenvalue only applied to ETM
      matrices.
  vT = 1./sqrt(1 + vS);
  vTi = sqrt(1 + vS);
  % The clipping functionality is optional and not applied if argument is empty
  if not(isempty(Clip))
    vIdx = find(vT < Clip);
    if not(isempty(vIdx))
      vT(vIdx) = Clip;
      vTi = 1./vT;
      fprintf(' - ienks_make_matrices: clipping %i eigenvalues\n', numel(vIdx));
    end
  end % if not
  %% Assembling matrices with diagonals and eigenvectors.
```

```
  % Inverse ETM
  mTi = mV * diag(vTi) * mV';
  % ETM
  mT = mV * diag(vT)  * mV';
  % Inverse Hessian
  mG = mV * diag(vG)  * mV';
  if (nargout == 7)
    varargout{1} = mU;
    varargout{2} = mS;
    varargout{3} = mV;
  end
end % function ienks_make_matrices
```