Journal of Structural Geology 171 (2023) 104868

Contents lists available at ScienceDirect

Journal of Structural Geology

journal homepage: www.elsevier.com/locate/jsg

Structural geologic modeling and restoration using ensemble Kalman inversion

David Oakley [a,*], Nestor Cardozo [a], Ariel Almendral Vazquez [b], Per Røe [b]

[a] University of Stavanger, Department of Energy Resources, 4036 Stavanger, Norway
[b] Norwegian Computing Center, P.O.Box 114, Blindern, N-0314, Oslo, Norway

ARTICLE INFO

Keywords:
Ensemble Kalman inversion
Kinematic models
Geological modeling
Data inversion
Uncertainty quantification

ABSTRACT

We demonstrate the use of Ensemble Kalman Inversion (EKI) for building three-dimensional, multi-fault, kinematically restorable structural geologic models, by means of a workflow in which fault geometry, the distribution of slip on a fault, and the geometry of folded horizons are all modeled. The models are constrained by observations of faults and horizons in the present deformed state together with the expectation that horizons should restore flat. Two modeling approaches are tested: restoration from the deformed state, and forward modeling from the restored state. We first test these methods on a synthetic model involving a single fault, and then apply them to a real-field example involving five faults. Models are prone to ensemble collapse, which results in underestimation of uncertainty at small ensemble sizes, but localization and covariance inflation can mitigate this issue. With these methods, EKI can recover the true parameter values in the synthetic case and produce a solution consistent with the data in the real case, as well as quantify uncertainty in both cases. EKI, therefore, shows promise as a tool for building complex, restorable structural geologic models, and it holds the potential for integration of fault kinematics with other ensemble-based subsurface modelling workflows.

1. Introduction

Models of subsurface geologic structures are of critical importance to many applications in the geosciences and related industries, including those related to resource exploration, natural hazards, and scientific understanding of fundamental geological processes (Brandes and Tanner, 2014; Wellmann and Caumon, 2018). Static three-dimensional geological models can be built by interpolation between data points, using geostatistical methods such as Kriging (see Wellmann and Caumon, 2018; for a recent review of this kind of modeling). Such models represent the present-day (deformed) structure, but not the processes of deformation that produced them (Caumon et al., 2013; Laurent et al., 2013).

To better ensure geological realism, a model should not only represent accurately the observed deformed-state structure, but also be restorable to a pre-deformational state in which sedimentary strata are approximately horizontal or back to their regional dip. Restoration, incorporating knowledge of deformational processes such as fault-related folding, provides a means to test the validity of a geological model and to make necessary modifications. This approach is most associated with the concept of balanced cross sections (Dahlstrom, 1969; Elliott, 1983; Groshong et al., 2012) and with kinematic methods such as fault-bend folding (Suppe, 1983) and trishear (Erslev, 1991), but three-dimensional (e.g. Cardozo, 2008; Georgsen et al., 2012) and geomechanical restoration (e.g. Moretti et al., 2006) methods also exist. Recent work by Grose et al. (2021) has also incorporated fault kinematics into interpolation-based 3D modeling workflows.

Whether models are created to be restorable or through static interpolation, they are non-unique solutions and are subject to uncertainty (Bond, 2015; Wellmann and Caumon, 2018; Cardozo and Oakley, 2019). Data inversion methods, such as Markov chain Monte Carlo (MCMC), can be used to find an optimal solution and to estimate uncertainty in model parameters (de la Varga and Wellmann, 2016; Cardozo and Oakley, 2019). However, as the complexity of the model increases, such as by moving from two to three dimensions, allowing parameters such as fault displacement to vary spatially, and including multiple faults, the number of parameters that an inversion must fit for becomes very large. Such high dimensional parameter spaces challenge many of the standard algorithms, including MCMC, but can be addressed through alternative methods such as Ensemble Kalman Inversion (EKI) that are based on Gaussian approximations (Iglesias et al., 2013a, 2013b). In this paper, we apply EKI to structural geologic modelling,

* Corresponding author. BRGM, 45100 Orléans, France.
E-mail addresses: d.oakley@brgm.fr (D. Oakley), nestor.cardozo@uis.no (N. Cardozo), ariel@nr.no (A. Almendral Vazquez), perroe@nr.no (P. Røe).

https://doi.org/10.1016/j.jsg.2023.104868
Received 30 September 2022; Received in revised form 16 April 2023; Accepted 17 April 2023
Available online 18 April 2023

showing that it can be used to build complex, three-dimensional geological models that are kinematically restorable. We also find that the method can be used to estimate uncertainty in model parameters, but to do so, it is necessary to detect and account for the phenomenon of ensemble collapse. EKI has the potential to be a valuable tool in structural geologic modeling, allowing data inversion to be applied to complex or computationally intensive models that are difficult to fit with other methods.

2. Background

2.1. Data inversion in structural geologic modeling

Data inversion, i.e., using computational algorithms to fit a model to data, provides an alternative to manually building a geological model. In the use of fault-related fold kinematic methods for structural restoration, data inversion has been mostly associated with the trishear model (Erslev, 1991; Allmendinger, 1998). Data inversion algorithms that have been applied to trishear include simple grid searches (Allmendinger, 1998), local (Cardozo and Aanonsen, 2009) and global (Cardozo et al., 2011) optimization, randomized maximum likelihood (RML) (Cardozo and Aanonsen, 2009), Monte Carlo simulations (Regalla et al., 2010), and Markov chain Monte Carlo methods (Oakley and Fisher, 2015). A discussion about the advantages and limitations of these methods is given by Cardozo and Oakley (2019). These methods have been applied to relatively simple models involving a single fault. Faults are either assumed to be planar with constant slip or represented by simple circular listric geometries (Cardozo and Brandenburg, 2014). They have also mostly been applied in two-dimensions, with the few uses of three-dimensional trishear being limited to simple linear variations of model parameters along strike (Cardozo and Aanonsen, 2009).

In three-dimensional geological modeling, as in fold kinematics, there is also increasing interest in quantifying uncertainty rather than producing a single model (Røe et al., 2014; Cherpeau and Caumon, 2015; Wellmann and Caumon, 2018), and stochastic data inversion methods have been applied in this field as well. Cherpeau et al. (2012), de la Varga and Wellmann (2016), Aydin and Caers (2017), Grose et al. (2018), and Grose et al. (2019) have all applied Markov chain Monte Carlo methods to the problem of building three-dimensional geological models. In these studies, the goal is to build a static model, possibly incorporating geological knowledge in the process (Grose et al., 2019), but not to produce one that is kinematically restorable. In addition, due to the limitations of the algorithms used with respect to high-dimensional parameter spaces, they have either inverted for simplified parameterizations of three-dimensional fault or fold geometries (Cherpeau et al., 2012; Grose et al., 2018, 2019), or required highly informative priors (de la Varga and Wellmann, 2016).

2.2. Ensemble Kalman inversion and the ensemble Kalman Filter

Since its introduction by Evensen (1994), the ensemble Kalman Filter (EnKF) has become a widely used tool in several fields of Earth science, such as weather forecasting (Houtekamer and Mitchell, 2001), oceanography (Evensen, 1994; Evensen and van Leeuwen, 1996), and reservoir modeling (Aanonsen et al., 2009), which all require matching complex models to data. Although the EnKF was originally developed as a method for data assimilation in time-varying systems, EnKF-style updates to an ensemble of models can also be applied repeatedly using the same dataset as a means of solving inverse problems. Methods of this type include the Ensemble Kalman Inversion (EKI) of Iglesias et al. (2013b), the iterative ensemble smoother of Chen and Oliver (2012, 2013), and the ensemble smoother with multiple data assimilation of Emerick and Reynolds (2012, 2013). In this paper, we adopt the EKI approach, following the work of Iglesias et al. (2013b, 2016), Chada et al. (2018), and Iglesias and Yang (2021), but the other iterative methods are similar.

Ensemble Kalman methods have several benefits for solving inverse problems. There is no need to calculate derivatives of the forward model as in gradient-based inverse methods (Evensen, 2009a). They are well suited to inversions with high-dimensional parameter spaces (Evensen, 2009a), where the need for a large number of model evaluations may make other inversion methods such as MCMC impractical (Iglesias et al., 2013a). Because the calculation of the forward model for each ensemble member is done independently, the algorithm can easily take advantage of parallel computing (Muir and Tsai, 2020). By using an ensemble of model realizations, Ensemble Kalman methods can be used to estimate not only the best-fitting model parameters, as in optimization methods, but also uncertainty in the parameters (Evensen, 2009a). However, compared to MCMC methods that also allow uncertainty estimates, the EnKF, EKI, and related methods require far fewer evaluations of the forward model to converge to a solution (Iglesias et al., 2013a).

Ensemble Kalman methods do have some limitations, however. They converge to an exact sample of the posterior distribution only in the case of a linear model (Evensen, 2018). For a non-linear model, they provide only an approximation. Ensemble Kalman methods also rely on the assumption that parameter and data distributions are Gaussian, and they may be inexact if this assumption is not true (Evensen, 2018). There is thus a tradeoff in accuracy and speed when comparing EKI to MCMC methods, which can, with sufficient iterations, sample from the true posterior distribution of a non-linear model and from non-Gaussian distributions. Another drawback is the risk of ensemble collapse, in which the variance of model parameters is underestimated. This phenomenon results from the finite size of the ensemble, which can produce spurious correlations between data and model parameters or other data (Aanonsen et al., 2009; Evenson, 2009; Lacerda et al., 2019). Despite these limitations, EnKF, EKI, and similar methods are sufficiently accurate for many problems, including non-linear ones, as evidenced by their widespread application. In this paper, we seek to solve a problem with a high dimensional parameter space and with a somewhat costly to evaluate model, and for those reasons we adopt the EKI methodology.

2.3. Value of uncertainty quantification in structural geology

While structural geologic modeling and restoration provide valuable and widely-used tools, uncertainty in these models is often not quantified. Uncertainty arises from multiple sources. Subsurface data are often sparse, such as from boreholes, or have gaps in critical areas, such as around poorly imaged fault zones in seismic reflection data, so that the structure in these gaps is uncertain. Filling in these unknown areas is a major motivation for the use of structural balancing techniques (Groshong et al., 2012). Further, even where data are available, they are always subject to some uncertainty. In the case of seismic imaging, uncertainties of 10s or even 100s of meters may exist in the interpretation of horizon depths, fault geometries, and fault throw (Schaaf and Bond, 2019). Balancing can help to validate interpretations, but is itself subject to uncertainty (Judge and Allmendinger, 2011; Woodward, 2012). Uncertainty in structural models is of particular interest to the petroleum industry, but is similarly relevant to any application that requires an understanding of subsurface structure (Bond, 2015), such as mining, subsurface storage of fuels or wastes, and seismic hazard analysis. Even in the petroleum industry, the focus of uncertainty quantification with the EnKF has been on reservoir engineering variables such as porosity and permeability (Aanonsen et al., 2009) with the geology of the models assumed constant. A small number of studies have considered uncertainty in horizon or fault geometry within EnKF-based reservoir history matching (Seiler et al., 2010a, 2010b; Irving et al., 2014), but the focus there is still on matching fluid flow data, not on the types of data and methods typically used to build balanced geological models. Skauvold and Eidsvik (2018) also used the EnKF to quantify uncertainty in a model of sedimentation and stratigraphy, but this did not include faulting. As data volumes increase and computational methods improve, uncertainty quantification is of growing interest

throughout the earth sciences, including in structural geology as shown by an increasing number of papers on the subject published in the Journal of Structural Geology (Bond, 2015). We propose that Ensemble Kalman methods, already widely used in reservoir engineering, can be a valuable tool for structural geologists to meet the needs of uncertainty quantification in three-dimensional modeling and kinematic restoration.

3. Methods

3.1. Ensemble Kalman inversion

In an inverse problem, we seek to infer model parameters ($\boldsymbol{u}$) from observations ($\boldsymbol{d}$), given a known forward model (G), which relates model parameters to the observations. In real-world cases, the observed data are subject to noise so they will take the form:

$$\boldsymbol{d} = G(\boldsymbol{u}) + \boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{\eta}$ is the noise. We assume that $\boldsymbol{\eta}$ is multivariate Gaussian with a known covariance matrix, $\boldsymbol{\Gamma}$, and zero mean.

The ensemble Kalman filter uses an ensemble of realizations of both the model parameters and the observed data. With N as the total number of ensemble members, the initial ensemble of parameter vectors, $\{\boldsymbol{u}_0^{(j)}\}_{j=1}^{N}$, is drawn from the prior distribution of $\boldsymbol{u}$. Similarly, an ensemble of data perturbations, $\{\boldsymbol{\xi}^{(j)}\}_{j=1}^{N}$, is drawn from the covariance $\boldsymbol{\Gamma}$, so that $\boldsymbol{y}^{(j)} = \boldsymbol{d} + \boldsymbol{\xi}^{(j)}$ is the data realization for ensemble member j.

At each iteration of the EKI, we update the values of $\mathbf{u}$ for all ensemble members using the formula (Iglesias and Yang, 2021, their Eq. (4))

$$\boldsymbol{u}_{n+1}^{(j)} = \boldsymbol{u}_n^{(j)} + \mathbf{C}_n^{\mathbf{uG}}\left(\mathbf{C}_n^{\mathbf{GG}} + \alpha_n \boldsymbol{\Gamma}\right)^{-1}\left(\boldsymbol{d} + \sqrt{\alpha_n}\,\boldsymbol{\xi}^{(j)} - G\left(\boldsymbol{u}_n^{(j)}\right)\right), \tag{2}$$

where n is the iteration number, and $\mathbf{C}_n^{\mathbf{uG}}$ and $\mathbf{C}_n^{\mathbf{GG}}$ are empirical covariance matrices, as given in Eqs. A1 and A2 in Appendix A, and α_n is a regularization parameter inspired by the regularizing Levenberg-Marquardt algorithm (Hanke, 1997). Eq. (2) is applied repeatedly to the same set of data to update the model parameters until a convergence criterion is reached. We choose α_n according to the method of Iglesias and Yang (2021), which they call EKI-DMC, for "data-misfit controller." Details are given in Appendix A.

3.2. The problem

In this paper, we aim to apply the EKI-DMC method to the problem of building a geological model that honors observed structural data and is kinematically restorable to an original horizontal or sub-horizontal, planar configuration. This requires building a model of the fault geometry and displacement on the fault surface, building a model of the horizon geometry in either the deformed or restored state, and using a kinematic model for fault-related deformation to transform the horizon model from the deformed to the restored state or vice versa. Depending on how much information is available a priori, it may not be necessary to consider all parts of the model as unknowns. In this paper, however, we fit for all parts of the model (fault geometry, fault displacement, and horizon geometry) to test the case of maximum uncertainty. The workflow is summarized in Fig. 1.

3.3. Fault model

We parametrize the fault in terms of a local coordinate system (u,v,w) relative to a dipping reference plane (Hoffman and Neave, 2007; Georgsen et al., 2012; Røe et al., 2014), with the u direction being along strike, v up-dip, and w normal to the plane towards the hanging wall. The origin of the reference plane is fixed, but its strike (φ) and dip (ψ) are allowed to vary. The fault surface is defined by a function $f(u,v)$, which is the signed distance from the reference plane and is parameterized as:

$$f(u,v) = c_f + \varepsilon_f(u,v), \tag{3}$$

where c_f is a constant offset from the reference plane and $\varepsilon_f(u,v)$ is an additional offset at position (u,v). For each fault, we then fit for a single value of each φ, ψ, and c_f (which together provide the three values necessary to define a plane) and for $\varepsilon_f(u,v)$ at each grid cell of the reference plane.

The three-dimensional fault kinematic model we use is that of Georgsen et al. (2012). This kinematic model is defined by a displacement field, $d(u,v)$, with a value at each grid point of the reference plane, an asymmetry parameter, γ, which defines the ratio of hanging wall to footwall displacement, and a reverse drag radius, R, which determines the extent of deformation away from the fault. The displacement profile is expected to be approximately elliptical (Walsh and Watterson, 1987). We therefore fit for an elliptical trend and a perturbation to it, defining d as:

$$d(u,v) = d_{ellipse}(u,v) + \varepsilon_d(u,v), \tag{4}$$

where $d_{ellipse}(u,v)$ is the elliptical trend, which depends on the along-strike and along-dip semi-axis lengths of the ellipse (l_u and l_v), the center of the ellipse (u_0, v_0), and the maximum displacement (d_{max}) as given in Appendix B, Eqs. B1-2. $\varepsilon_d(u,v)$ is the perturbation and is solved for at each point of the same grid used for $f(u,v)$.Displacement decreases away from the fault surface according to the curve of Cardozo et al. (2008). The three-dimensional displacement field (Georgsen et al., 2012; Røe et al., 2021) is given by Eqs. B3-5 in the Appendix.

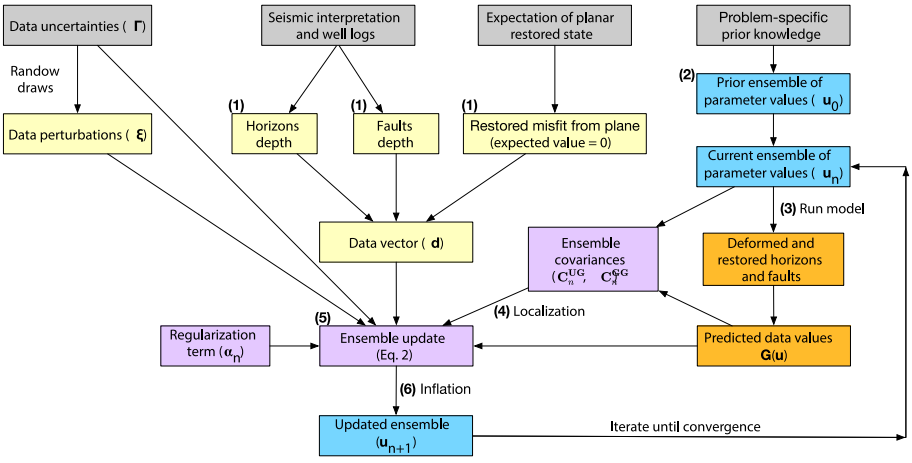

Fig. 1. Flow chart illustrating the EKI workflow. The creation of the prior ensemble will be problem-specific and may involve drawing random values within ranges considered reasonable or may involve the use of a pre-existing ensemble of models, perhaps derived from geostatistics or from interpretations by multiple people. The "Run Model" step (label 3) includes the interpretation of the fault and fold geometry from model parameters (sections 3.3 and 3.4) as well as restoration or forward modeling using the kinematic model. Meaning of colors: Grey: Data sources and prior knowledge, Yellow: Data, Blue: Model parameters, Orange: Model outputs, Purple: EKI terms and equations. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.4. Horizon model

Horizon geometry is defined by depth at points on a horizontal grid in the global Cartesian coordinate system. We follow the approach used in the Havana and RMS software, in which the model domain is divided into fault blocks defined by whether they are in the footwall or hanging wall of each fault (Hoffman and Neave, 2007). Within each fault block, only the points that fall on the correct side of the fault are used to make the modeled horizon that will be compared to data, while the rest of the points are considered inactive. This approach allows fault geometry to change easily and allows for sharp changes in horizon depth across a fault.

We parameterize the depth of the top horizon in terms of a planar trend plus a residual, which is fit for at each grid square. Since we are working with conformable horizons in the restored state (pre-growth strata), we parameterize the depth of horizons below the top horizon by distance below the overlying horizon. The full equation for the depth of each horizon is:

$$z_{i,j}(x,y) = \begin{cases} ax + by + c + d_j + \varepsilon_{z,i,j}(x,y), i = 1 \\ z_{i-1,j}(x,y) + T_i + \varepsilon_{z,i,j}(x,y), i > 1, \end{cases} \quad (5)$$

where the subscript i indexes the horizon, and j indexes the fault block. The coefficients a, b, and c define the plane that the top horizon should restore to, T_i is the restored-state mean thickness from horizon i to the overlying horizon, d_j is a vertical offset term for fault block j, and $\varepsilon_{zi,j}(x, y)$ is an additional vertical offset of horizon i in fault block j at location (x,y). In the inversion, we fit for each of the global parameters (a, b, c, d_j, and T_i) and for the local variables $\varepsilon_{z,i,j}(x,y)$ at each grid cell for each horizon and each fault block. In some cases, we define some fault blocks on smaller grids than the global model to reduce the number of parameters needed.

3.5. Restoration and forward modeling

We consider two approaches to fitting for horizon geometry: restoration and forward modeling. In the restoration method, we fit for the horizon geometry in the deformed state (as defined by the trend and perturbation terms in Eq. (5)). We measure misfits between these horizon depths and the horizon data. We then restore the model by running the kinematic model backwards and reinterpret the restored horizon depths onto the model grid, where we measure the misfit of each grid point from the expected restored state planar trend. In the forward method, in contrast, we fit for the horizon geometry in the restored state, again as a trend plus a perturbation. We take the perturbation term at each grid cell as the restored-state misfit, thus allowing the restored-state model to not be perfectly planar but penalizing it for deviations from a plane. We then run the kinematic model forward to calculate the deformed-state horizon geometry, and we measure the misfit from the horizon data. The parameterization of the horizon depths in this case is somewhat simpler, since the restored-state horizon is unfaulted, so only one fault block is needed. Therefore, d_j in Eq. (5) is 0, and there is only one ε_z parameter per horizon at each cell of the model grid, rather than one for each fault block.

3.6. Hierarchical parameterization

The results of the EnKF and EKI depend in part on the prior probability distribution of model parameters as represented by the initial ensemble. For the spatially varying parameters $\varepsilon_f(u,v)$, $\varepsilon_d(u,v)$, and $\varepsilon_z(x, y)$, a prior that incorporates spatial correlation is likely to be a good choice, but parameters such as the length-scale of correlation may be unknown. For this reason, hierarchical priors may be used (Chada et al., 2018), in which these parameters are unknowns to be fit during the inversion. In this paper, for all three spatially varying parameters, we use a hierarchical parameterization based on a spherical variogram. Given a variogram range, r, and standard deviation, σ, the covariance between two grid points is (Davis, 2002):

$$C_{ij} = \begin{cases} \sigma^2\left(1 + 0.5\left(h^3 - 3h\right)\right), h < 1 \\ 0, h \geq 1, \end{cases} \quad (6)$$

where:

$$h = \frac{l_{ij}}{r}, \quad (7)$$

and l_{ij} is the distance between the two grid points. Given the covariance matrix $\boldsymbol{C}$ and its lower triangular Cholesky decomposition $\mathbf{L}$, such that $\boldsymbol{C} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}$, a random realization of the spatially varying model parameter ($\boldsymbol{p}$) at all points on the grid is given by:

$$\boldsymbol{p} = \boldsymbol{L}\boldsymbol{s}, \quad (8)$$

where $\boldsymbol{s}$ is a vector of random numbers drawn from the standard normal distribution (mean of 0 and standard deviation of 1). Other methods of drawing realizations could also be used if needed for large grids (e.g. Abrahamsen et al., 2018), but this method is sufficient for our models. Using the hierarchical parameterization of $\boldsymbol{p}$ in terms of $\boldsymbol{s}$, r, and σ, we fit for these values as the model parameters rather than fitting for $\boldsymbol{p}$ directly. This parameterization in terms of a spherical variogram assumes a zero mean; we have accounted for this fact by including the necessary trend terms as separate parameters in Eqs. (3)–(5) so that the assumption of zero mean is reasonable for the ε terms. Since we use the hierarchical parameterization for multiple properties, we denote r and σ with subscripts f, d, and z for the fault surface, displacement, and horizon depths respectively and with numbers in the case of multiple faults or horizons (see Table 1 in the manuscript and S1 in the supplementary material).

3.7. Data

We fit our models to three types of data (Fig. 1, label 1). Two of these are straightforward: points on the fault surface and points on the deformed-state horizons, which typically come from interpretations of seismic reflection data or well logs, and we try to minimize the misfit of the model to each of these data types in the deformed state. To include kinematic restoration in our models, it is necessary to include a third type of data that can test the quality of the restoration. For this, we use the restored-state depths of each horizon, measured at each grid point. Since the expected depth of the restored-state plane is defined by the unknown parameters a, b, c, and T_i, we cannot create realizations of the restored state depths directly, so we create realizations of the differences between the restored horizons and the plane defined by these parameters, which we compare to the observed differences in each model. These realizations have a mean value of 0 and a user-specified covariance.

3.8. Priors

The initial ensemble (Fig. 1, label 2) should be drawn from the prior probability distribution of the model parameters. Since the gridded model parameters $\varepsilon_f(u,v)$, $\varepsilon_d(u,v)$, and $\varepsilon_z(x,y)$ are defined hierarchically, the priors for the hierarchical parameters ($\boldsymbol{s}$ in Eq. (8)) are standard normal distributions. The remaining non-local parameters are restricted to a specified range of reasonable values, which must be chosen for each specific problem (see Table 1 for an example). Since the EKI does not enforce bounds on parameter values, we use a logarithmic transformation to convert the non-local parameters to an unbounded domain. For a variable x with limits $x_{\min}$ and $x_{\max}$, the transformation is:

$$x_{transformed} = \ln\left(\frac{x - x_{min}}{x_{max} - x}\right). \quad (9)$$

Initial values for these transformed variables are all drawn from a normal distribution with standard deviation of 1.8, which gives a

Table 1
Allowed ranges for the non-local parameters in the synthetic model. The local parameters $\varepsilon_f(u,v)$, $\varepsilon_d(u,v)$, and $\varepsilon_z(x,y)$, and the corresponding **s** variables in the hierarchical representation are not bounded.

Description	Variable	Minimum	Maximum	True Value[a]
Fault surface variogram range	r_f	500 m	5000 m	N/A
Fault surface variogram st. dev.	σ_f	50 m	500 m	N/A
Fault strike	φ	60°	80°	70°
Fault dip	ψ	40°	65°	55°
Fault surface offset	c_f	−200 m	200 m	0 m
Displacement variogram range	r_d	50 m	2500 m	600 m
Displacement variogram st. dev.	σ_d	5 m	100 m	20 m
Maximum displacement	$d_{\max}$	50 m	500 m	250 m
Displacement horizontal semi-axis	l_u	250 m	1500 m	1000 m
Displacement vertical semi-axis	l_v	250 m	1500 m	500 m
Displacement ellipse center u	u_0	−200 m	200 m	0 m
Displacement ellipse center v	v_0	−200 m	200 m	0 m
Displacement asymmetry	γ	0	1	0.6 m
Reverse drag radius	R	100 m	1500 m	700 m
Horizon A variogram range	r_{z1}	50 m	2000 m	N/A
Horizon B variogram range	r_{z2}	50 m	2000 m	N/A
Horizon A variogram st. dev.	σ_{z1}	1 m	100 m	N/A
Horizon B variogram st. dev.	σ_{z2}	1 m	60 m	N/A
Regional slope in x direction	a	-tan (2.5°)	tan(2.5°)	0°
Regional slope in y direction	b	-tan (2.5°)	tan(2.5°)	0°
Horizon A mean depth	c	1000 m	1200 m	1100 m
Block 1 offset[b]	d_1	−100 m	100 m	N/A
Block 2 offset[b]	d_2	−100 m	100 m	N/A
Horizon B mean depth below A	T_2	100 m	400 m	300 m

[a] N/A indicates values that were not used in making the synthetic forward model, because the horizons were initially perfectly horizontal and the fault surface curvature was not done with a variogram.
[b] Restoration method only.

distribution that is close to uniform when transformed back to the bounded domain. For the prior ensemble and the data realizations, we use Latin hypercube sampling (McKay et al., 1979; Stein, 1987) to ensure an approximately even representation of the distribution.

3.9. Detecting and mitigating ensemble collapse

A simple way to detect ensemble collapse, which we adopt in this work, is to add a dummy parameter to the model with known prior mean and variance and no real correlation with the data (Evensen, 2009b; Røe et al., 2016). Any reduction of the posterior variance of the dummy parameter is therefore a sign of ensemble collapse and an indication of its magnitude. Two widely used methods for mitigating ensemble collapse are localization and covariance inflation (Evenson, 2009), and we consider both in this work. Localization (Fig. 1, label 4) attempts to reduce or eliminate spurious correlations that can appear in the covariance terms of Eq. (2) (Fig. 1, label 5), while covariance inflation (Fig. 1, label 6) attempts to correct the variance of model parameters in the ensemble after the update step (Eq. (2)) is completed.

A variety of localization methods exist (Lacerda et al., 2019), but we use the bootstrap method of Zhang and Oliver (2010)., which has the benefit that it does not rely on distance and can therefore be applied to the non-local parameters in our model, such as R and γ in the fault kinematic model and the various hierarchical parameters. For covariance inflation, we use the method of Evensen (2009a,b), which uses dummy parameters to estimate the amount of variance loss, and which was found to work well by Lacerda et al. (2019). While the dummy parameters for inflation can be regenerated at each iteration of the inversion (as described by Evensen, 2009b), we found by comparison to an independent dummy parameter, that ensemble collapse was best averted if the same set of dummy parameters is carried through all iterations, and we do that here. The equations for the localization and inflation methods are given in Appendix C.

4. Results

4.1. Synthetic model

To test the proposed method, we first created a synthetic model (Fig. 2). This model consists of a single elliptical fault and two horizons. The fault has a strike of 070° and dip of 55° with its center at 1200 m depth. The displacement ellipse has dimensions $l_x = 1000$ m and $l_y = 500$ m, with asymmetry $\gamma = 0.6$, reverse drag radius $R = 700$ m, and maximum displacement $d_{\max} = 250$ m. To make a more challenging model to fit (i.e. one in which $f(x,y)$ and $\varepsilon_d(x,y)$ are not 0 at all grid points), we alter this basic model. The fault surface is curved as shown in Fig. 2 and the displacement field is perturbed by addition of a Gaussian random field drawn from a spherical variogram with range of 600 m and standard deviation of 20 m. The two horizons in the model were horizontal before deformation, with depths of 1400 m and 1100 m. To create synthetic data points for the horizons and fault surface, we take their elevations at all points where they are present on a 2 km × 2 km horizontal grid at 100 m spacing between grid points. This dense dataset is roughly analogous to data from a 3D seismic survey. All data are perturbed with a 10 m standard deviation, so that the data uncertainty is known for this case. Horizon data points within 25 m of the fault surface are removed (Yielding and Freeman, 2016). For the inversion, horizons and faults are gridded with 100 m spacing. Allowed parameter ranges for the inversion are as shown in Table 1. We draw data realizations from normal distributions with a standard deviation of 10 m for all three data types, to match the uncertainty in the synthetic data.

To determine the ability of the algorithms to find the correct solutions without ensemble collapse, we test the EKI-DMC alone, with localization, with covariance inflation, and with both localization and covariance inflation. For each case, we test ensemble sizes of 100, 200, 500, 1000, and 2000 members, and we test both the restoration and forward modeling methods.

The standard deviation of the dummy parameter (Fig. 3, top row) shows significant ensemble collapse at low ensemble sizes with the EKI-DMC algorithm alone. Ensemble standard deviation tends to approach the true value (1.0) more accurately as ensemble size increases, but ensembles of as many as 2000 members may be prohibitively time-consuming for large models. The collapse is worse for the restoration method, but it is substantial for forward modeling as well. Inflation shows the most significant effect in reducing ensemble collapse, while localization has more modest effects. The combination of both localization and inflation leads to dummy parameter standard deviations close to 1 even at small ensemble sizes such as 100 members.

In addition to the changes in the standard deviation of the dummy variable, the mean also shows changes away from zero (Fig. 3, bottom row). The significance of these changes is unclear. They are both positive and negative and do not show clear trends, suggesting that they are random. We also find that they vary from one run to the next if the inversion is rerun with the same model sizes but different starting ensembles. They may therefore be random fluctuations that are able to grow because there is no actual constraint on the dummy variable from the data as there would be for a physically meaningful variable.

In the remainder of our analysis, we focus on the N = 200 ensemble size with both localization and inflation, as this size and combination of methods shows only slight ensemble collapse and less change of the dummy mean than in the N = 100 case.

If EKI is to be a useful tool for structural modeling, then it must be able to retrieve the true values of parameters for the synthetic model with reasonable accuracy. Model parameters come in two types: those

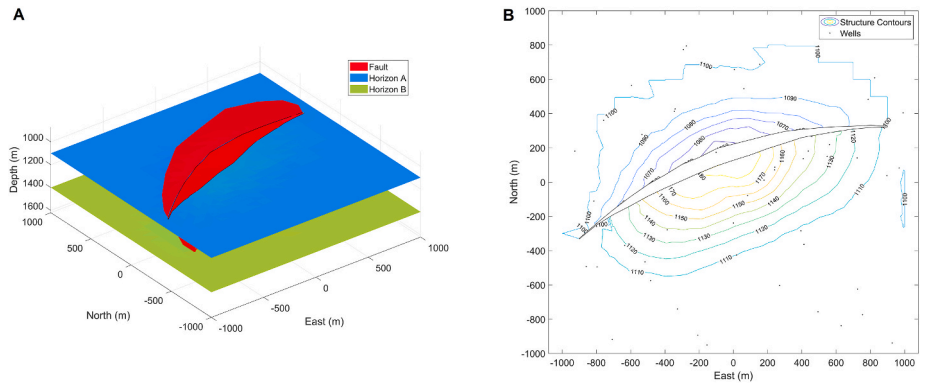

Fig. 2. Geometry of the synthetic model. (a) Three-dimensional view. Blue lines on fault are the upper horizon A cutoffs. (b) Structure contour map of Horizon A. Contours show depth in meters. Black dots show locations of simulated wells used in the sparse data case (see section 4.2). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

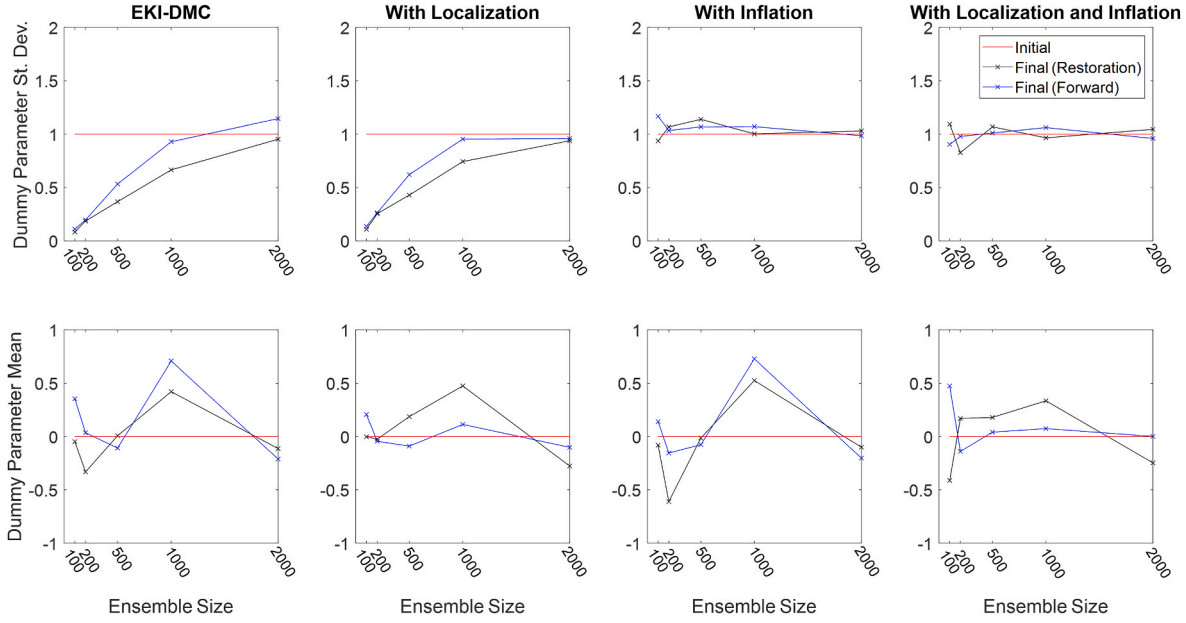

Fig. 3. Comparison of different methods and ensemble sizes. Top row: Ensemble standard deviations for the dummy parameter. Bottom row: Ensemble means for the dummy parameter. Columns from left to right show the results using only the EKI-DMC algorithm, the EKI-DMC algorithm with localization, the EKI-DMC algorithm with covariance inflation, and the EKI-DMC algorithm with both localization and inflation. In each case, black lines show the restoration method, and blue lines the forward modeling method, while red lines show the initial ensemble. Since the dummy parameter does not affect the model, its final ensemble mean and standard deviation should ideally be equal to those of the initial ensemble. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

such as fault strike or displacement asymmetry that are not tied to a specific location, and those that are defined at a specific grid point on a fault or horizon. We used kernel density estimation (KDE), as implemented in Matlab's ksdensity function, to estimate probability distributions for parameters from the final ensembles. Results for 9 non-local parameters with N = 200 (Fig. 4) show means that are close to the truth, and probability distributions that are sufficiently wide to encompass the true value within uncertainty. Some parameters are more accurately and precisely fit than others, however. For instance, the restored depth of horizon A is fit within a few meters of the true value of 1100 m (1099.5 $\pm$ 2.3 m and 1101.0 $\pm$ 3.5 m for the forward and restoration methods, respectively) (Fig. 5c). Given that Fig. 2 shows a large area unaffected by faulting and still at this original depth, it is likely that the data from this area strongly constrain this model parameter. On the other hand, the along-dip semiaxis of the displacement ellipse (l_y) is much more uncertain (Fig. 4i). Since displacement is constrained primarily by the magnitude of deformation of horizons and the two deformed horizons used in the model do not cover the full dip extent of the fault (Fig. 2), it is likely that there is simply not sufficient data to constrain this value with high precision. For most parameters, the forward and restoration methods give similar results. Fault dip, with a true value of 55° is a notable exception, with the forward mean being too low (51.3° $\pm$ 4.5°) and the restoration mean too high (59.0° $\pm$ 5.1°), but even in that case there is substantial overlap of the two distributions, and both means are within one standard deviation of the truth (Fig. 4b).

To see how well the Ensemble Kalman inversion can find the correct values for and quantify uncertainty in the gridded displacement field, fault surface geometry, and horizon geometry, we plot ensemble means and standard deviations of these values and compare them to the true values (Figs. 5 and 6). These results are for forward modeling, but restoration results are similar (Figs. S1 and S2 in supplementary material). The horizon geometry in both the restored and deformed states is fit very well (Fig. 5), with both the restored and deformed state ensemble means being nearly indistinguishable from the truth (Fig. 5a-b

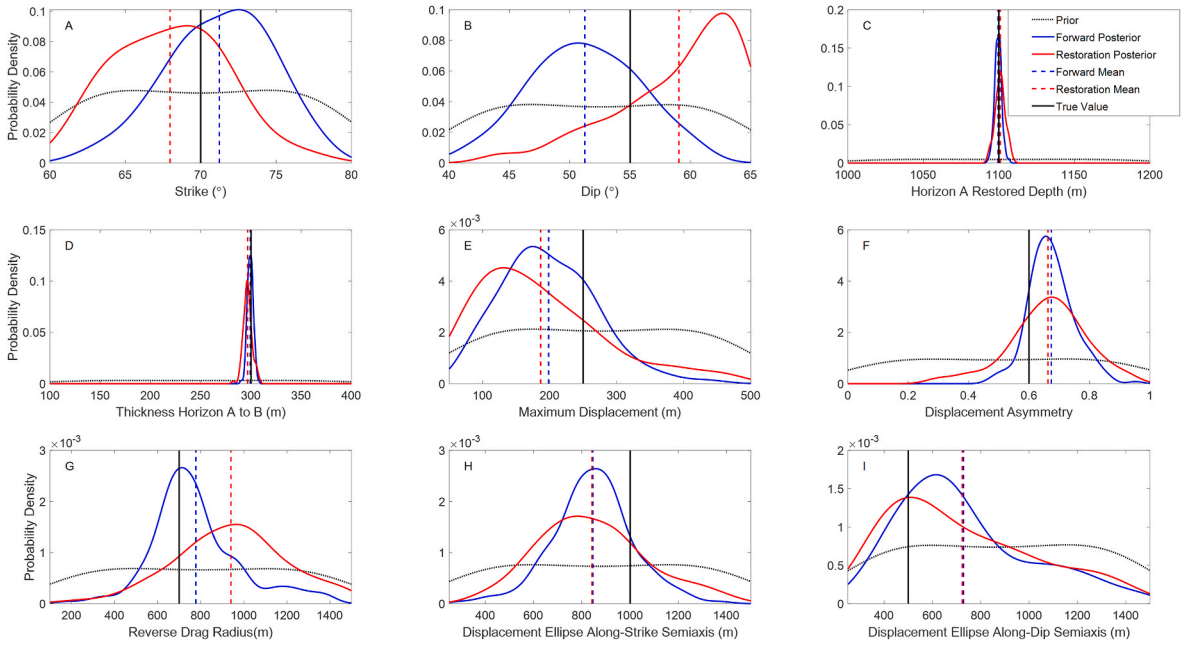

Fig. 4. KDE probability densities for several non-local model parameters, N = 200 case, and forward modeling (blue) or restoration (red) methods. The prior density (black dotted line) corresponds to the initial ensemble and is the same for both methods, while the posterior density corresponds to the final ensembles. Ensemble means are indicated by vertical dashed lines and the true values by a black solid line. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

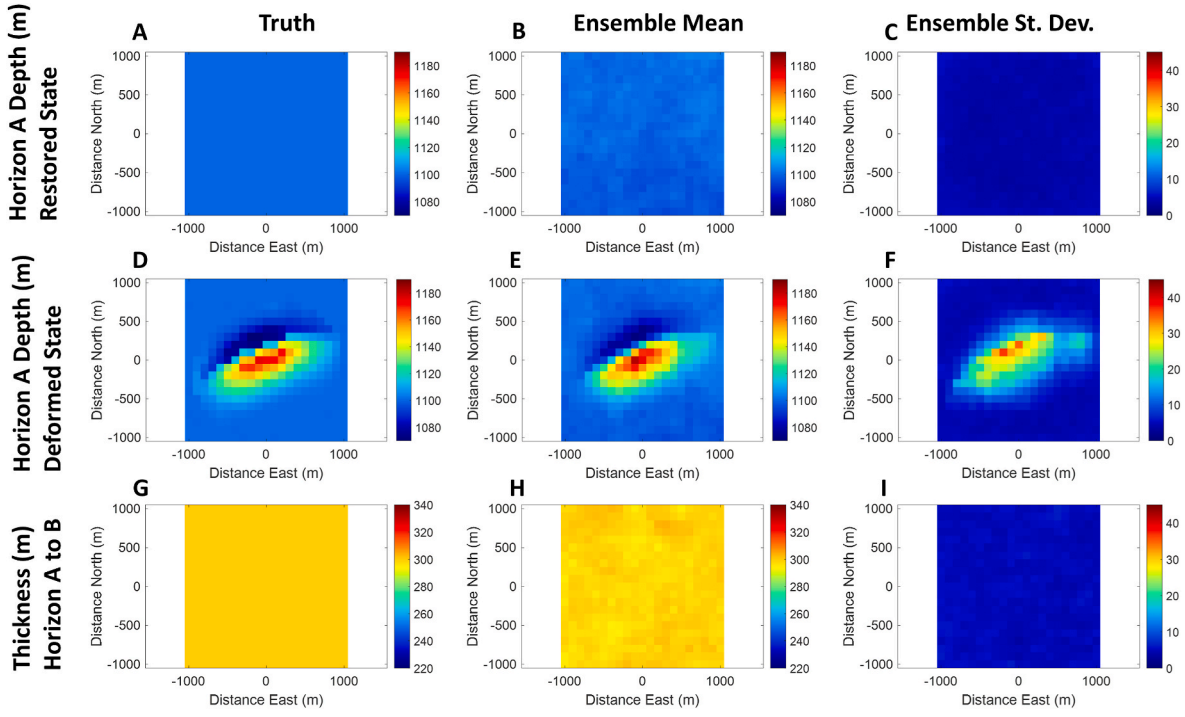

Fig. 5. Results for gridded horizon depths, $z(x,y)$ and thicknesses (vertical in the restored state) for the synthetic model. True values are shown in the left column, ensemble means in the middle column, and ensemble standard deviations in the right column. The results shown here are for forward models with N = 200, using both localization and covariance inflation with the EKI-DMC algorithm.

and d-e). Standard deviations in the depth to the top horizon are highest near the fault (up to about 37 m), but are less than 10 m in other areas and in the restored state (Fig. 5c and f), while the thickness of the unit between the two horizons (Fig. 5g-i) in the restored state is well-constrained to within about 10 m of its true value of 300 m.

The fault surface geometry and displacement are not quite as well fit as the horizon geometry, in that some differences from the truth can easily be distinguished visually (Fig. 6). Nonetheless, the major features are accurately reproduced. Since the fault is curved (Fig. 2), the fault surface function, $f(u,v)$, is negative in the along-strike middle of the fault and positive towards its edges (Fig. 6a), and this is seen in the ensemble mean (Fig. 6b), with uncertainty lowest in the center and higher towards

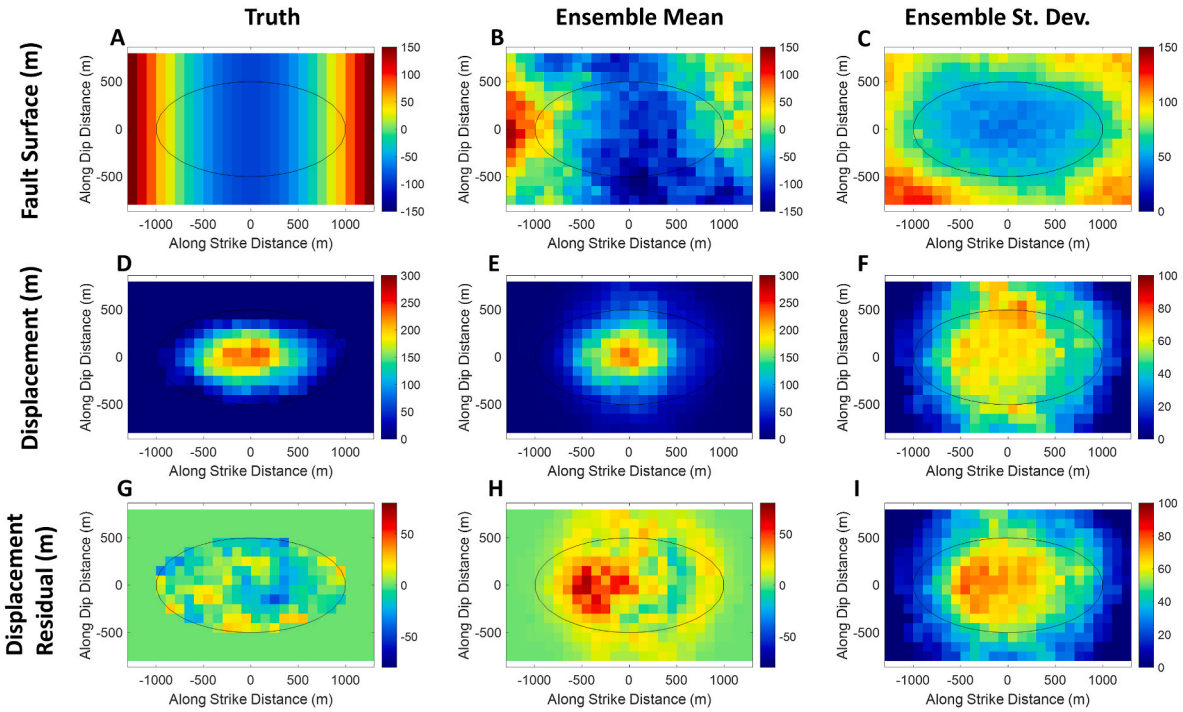

Fig. 6. Results for gridded fault surface geometry, $f(u,v)$, and displacement, $d(u,v)$, for the synthetic model. True values are shown in the left column, ensemble means in the middle column, and ensemble standard deviations in the right column. The results shown here are for forward models with N = 200, using both localization and covariance inflation with the EKI-DMC algorithm. The black line shows the fault tip line in the true model. Values outside of this tip line in the true model (A) do not affect the model data, so the results in (B) are not required to match this region.

the edges (Fig. 6c). Fault displacement (Fig. 6 d-f) has approximately the correct elliptical shape and maximum central displacement (see also Fig. 4). Uncertainty in displacement is highest near the upper edge of the fault, likely reflecting the lack of horizons in that area to constrain it. Displacement is both a function of an elliptical trend, the parameters of which were shown in Fig. 4, and a gridded residual (ε_d). Much of the displacement field that is shown in Fig. 6 (middle row) is due to this trend. To separate the contributions of the residual, we plot just the residual in Fig. 6g-i. Some features of the residual, such as the presence of a negative region on the right side of the fault are captured in the ensemble mean, but in general, the ensemble mean shows lower residuals that are within one standard deviation of zero.

4.2. Sparse data

In the synthetic model analysis so far, we have used a data set consisting of points on a regularly spaced grid. In many real-world cases, however, only a smaller number of scattered data points are available, such as from wells or outcrops. Further, in such sparse data cases there may be greater reliance on models to fill in the structure where data are not available. To test this case, we created a sparse data version of the synthetic model. Fifty simulated wells were created at scattered locations (Fig. 2b), and their intersections with the two horizons and the fault were taken as data, providing 14 fault data points and 97 horizon data points. Since the restored state misfit from a plane is measured on

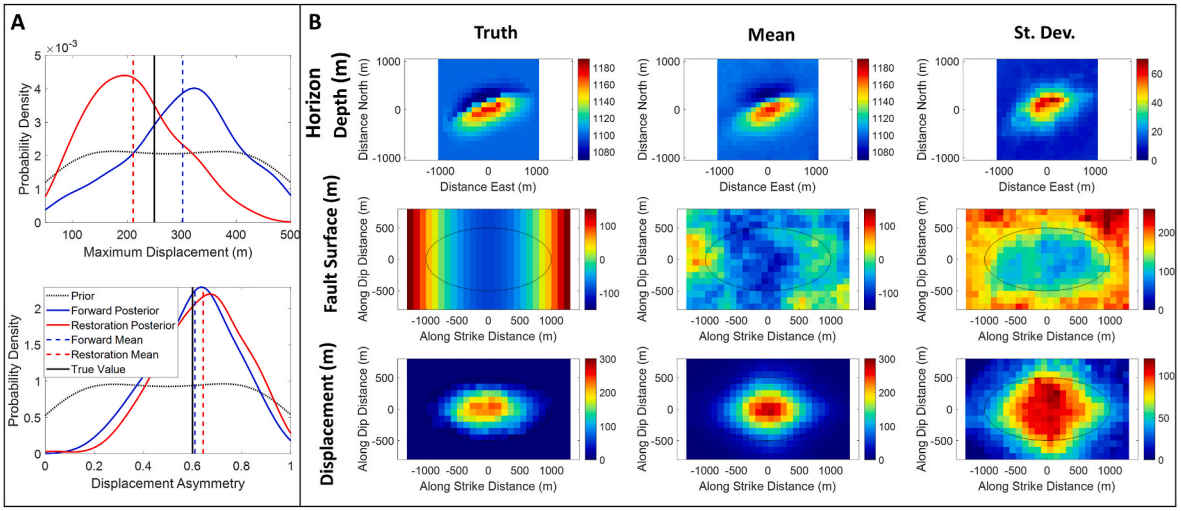

Fig. 7. Some results for the sparse data case (wells in Fig. 1b): (A) KDE probability densities of fault displacement and displacement asymmetry (compare to Fig. 4), and (B) comparisons of the true model, ensemble mean, and ensemble standard deviation for the top horizon depth (compare to Fig. 5) and the fault surface function, $f(u,v)$ (compare to Fig. 6). The results in (B) are for forward models with N = 200, using both localization and covariance inflation with the EKI-DMC algorithm.

the grid (see section 3.5), we still use all 441 grid squares for each horizon as data points for this constraint.

Results for this inversion using forward modeling (Fig. 7) and restoration (Fig. S3 in supplementary material) show that the ensemble mean is still able to capture the major features of the true model, but with greater uncertainty than in the dense data case. For example, with sparse data the maximum displacement and the asymmetry (Fig. 7) are 302 ± 101 m and 0.61 ± 0.17, respectively, with forward modeling and 211 ± 83 m and 0.64 ± 0.17 with restoration. In contrast, with dense data, they are 198 ± 71 m and 0.67 ± 0.07 with forward modeling and 187 ± 97 m and 0.66 ± 0.13 with restoration. All of these values are within error of the true values of 250 m and 0.6, but the uncertainty is generally greater with sparse data. Similarly, the mean upper horizon depth matches the truth well in Fig. 7, as in Fig. 5, but the standard deviation is higher and the area of high standard deviations around the fault is larger. The fault, as well, has the basic curved shape and elliptical displacement of the true model but with greater uncertainties in Fig. 7 than in Fig. 6.

4.3. Emerald Field

For a real-world case study, we use the Emerald Field model that is provided as a tutorial example in the RMS modeling software and that has been used in previous studies such as Cardozo et al. (2008), Georgsen et al. (2012), and Røe et al. (2010; 2014). We consider a portion of the model encompassing five normal faults (F1 to F5), and we use data from two horizons (A and C, Fig. 8). Parameter ranges are given in Table S1 (supplementary material). These were chosen based on general knowledge from the data and the interpolation-based model in Fig. 8 (e.g. number of faults and their approximate locations and reasonable horizon depths) but with sufficient uncertainty to not pre-determine the solution. We use an ensemble size of N = 200 with covariance localization and inflation. We test both the forward modeling and restoration methods, although we focus on the forward modeling results here since they give a slightly better fit to the data. (Results for the restoration method are given in the supplementary material in Figs. S4–S6 and Text S1.) The data consist of points on the horizons interpreted from a 3D seismic survey (5323 total points), with points less than 50 m from a fault removed from the dataset, and points comprising fault sticks and midlines (5424 total points). The horizontal grid used to model the horizons and assess their fit to a plane in the restored state has

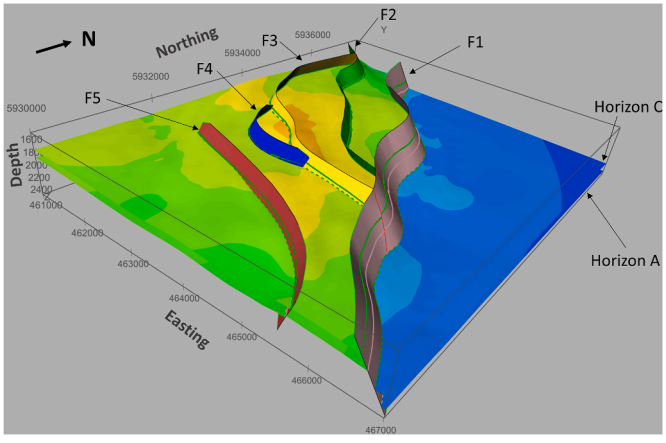

Fig. 8. The portion of the Emerald Field model used in our inversion, as seen in RMS. There are two horizons in this model, horizon A which is the lower horizon, and horizon C which is the upper horizon. F1 to F5 are faults. Green and pink lines on the faults show the horizon cutoffs. Continuous lines are footwall cutoffs, and dashed lines are hanging wall cutoffs. This is a model as made by interpolation in RMS, rather than by our method, and thus provides a basis for comparison. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

100 m spacing with a total of 4697 grid points per horizon (or 9394 total for the two horizons). In total, there are 20,139 data points, while the number of model parameters is 21,401. The uncertainty in the data is not well known, but it is to be expected that the horizons are better imaged and thus less uncertain than the faults. Further, the misfit from a planar restored state may be affected by the imperfect nature of the kinematic model (section 3.3), so this also is expected to be less certain than the directly observed data. To reflect this state of knowledge, we estimate uncertainties (1 standard deviation) of 20 m for the horizon data and 50 m for the fault data and restored state misfits. A dummy parameter is also included in the model, with an initial Gaussian distribution of 0.0 ± 1.0, and its final distribution of 1.0 ± 1.0 shows little ensemble collapse but does indicate a shift of the mean similar to those observed in Fig. 3.

In the work so far, we have used minimally informative prior distributions in which the non-local model parameters are limited to a specified range but have no other constraints or correlations, while the spatially varying values are constrained only by the use of a spherical variogram prior. However, the geometries of faults and horizons, and perhaps even the distribution of displacement on the fault, may be approximately known, especially in dense data cases. In this case, it makes sense from a Bayesian perspective to make use of such prior knowledge. Therefore, we try the Emerald Field case with two different prior ensembles. In one case, we randomly generate parameters within the allowed range as described above. In the other case, for a more informative prior, we create an ensemble of 200 initial realizations using the software packages RMS and Havana, both of which can create multiple geostatistical realizations of a model. Havana is used to simulate uncertainty in the fault surfaces and displacement fields, while RMS is used to simulate uncertainty in the fault positions and orientations and in the horizons. In order to use the forward modeling method and hierarchical parameterization, we restore each realization after export from RMS, and we fit variograms to gridded parameters to convert them to a hierarchical form.

Fig. 9 shows horizon C, which is the top horizon in the model, in the deformed state. Using the minimally informative prior as described above, the prior ensemble mean shows only the general position and strike of fault F1 and little hint of the other faults (Fig. 9a), and uncertainty is high throughout (Fig. 9e). The final ensemble mean resolves the shape of F1 and the offset across it well and shows the other faults to some degree, such as by the presence of a higher block between F1 and F2 and a lower one between F2 and F3 (Fig. 9b). Uncertainty is highest in the vicinity of the faults, and especially in complex areas such as the truncation of F2 and F3 by F1 in the SE part of the model (Fig. 9f). When instead the prior generated by RMS and Havana is used, even the prior ensemble mean shows the different fault blocks well (Fig. 9c), and the final ensemble mean is not very different (Fig. 9d). What does change substantially is the ensemble standard deviation (Fig. 9g–h), which is greatly decreased in the final ensemble.

Looking at the misfit between the ensemble mean models and the data points (Fig. 10) shows that they are mostly fit well but with some outliers. When using the minimally informative prior, the RMS error is 8.7 m in the restored state, with a maximum value of 30.4 m, and 37.2 m in the deformed state, with a maximum of 109.7 m, showing that the restored state is better fit, at least when using forward modeling. When using the more informative prior, the RMS misfit is 13.4 m in the restored state, with a maximum of 70.4 m, and 13.9 m in the deformed state, with a maximum of 50.0 m. When using the restoration method instead of the forward modeling method, the results are reasonably similar but with somewhat higher restored state misfits, slightly lower deformed state misfits, and significantly higher maximum misfits (Text S1 in supplementary material).

The throw on fault F1, as measured by the offset of the two horizons A and C, also shows the effects of the EKI inversion with the two priors (Fig. 11 and S6 in supplementary material). With the minimally informative prior, the initial ensemble has a very large uncertainty (blue

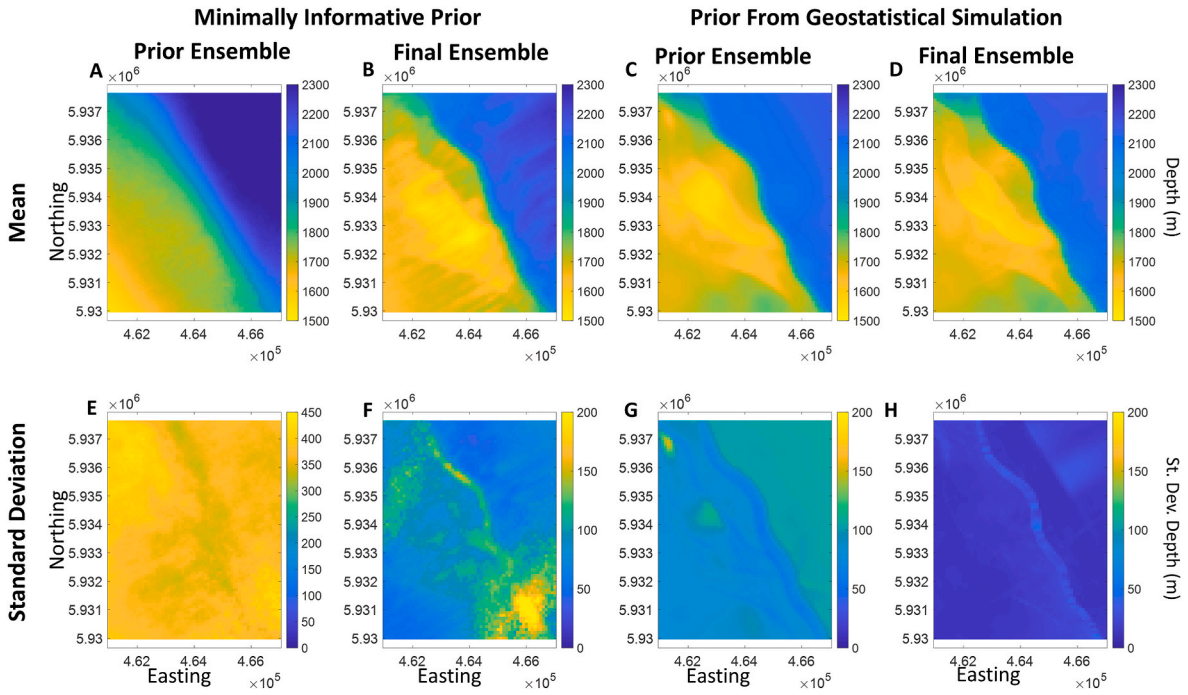

Fig. 9. Prior and final ensemble means (top row) and standard deviations (bottom row) for the deformed state depth to the upper horizon (Horizon C) in the Emerald Field model with the forward modeling method. (A, B, E, F): With a minimally informative prior ensemble drawn randomly from the allowed parameter space. (C, D, G, H): With a prior ensemble generated using geostatistical methods. Note that E uses a different color scale than F-H for the standard deviation of depth due to the much higher values in that case. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

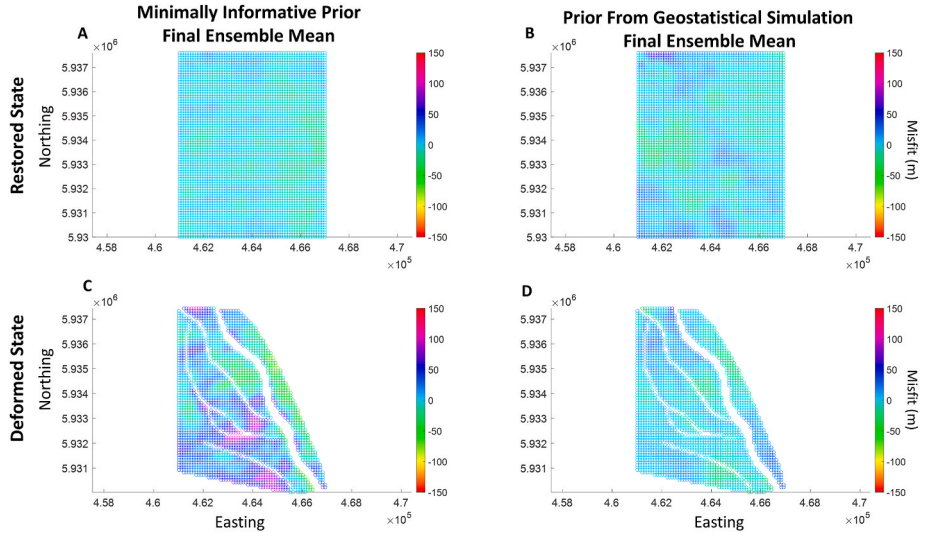

Fig. 10. Misfit between the final ensemble mean model and data points in the restored state (top row) and deformed state (bottom row) depth to the upper horizon (C) when using a minimally informative prior (left column) and a prior from geostatistical simulation by RMS and Havana (right column).

dashed lines), which is successfully narrowed down by the EKI (red dashed lines). With the more informative prior, there is less uncertainty, but it is still narrowed somewhat, especially when using the restoration method. The mean throw also changes from the prior to the final ensemble, in particular for the forward modeling method where it decreases on the SE side and increases on the NW, although those changes are still mostly within error of the prior mean. In both cases, the throw in the interpolation-based model shown in Fig. 8 is similar, but not identical, to the ensemble mean resulting from EKI.

The throw on fault F1 also shows important differences between the forward modeling (Fig. 11a–b) and restoration (Fig. 11c–d) methods. For the uninformative prior, the throw is always positive (normal sense), while in the restoration method it can be negative. This occurs because, in the forward method the horizon is continuous in the restored state and the fault displacement is always positive, so the resulting throw is positive, while with the restoration method, the throw is determined directly from the deformed horizon geometry, which is discontinuous across faults and can therefore have a negative throw if there is no prior information to constrain it. With the informative priors, this issue is avoided, but another difference appears: the forward modeling final ensemble mean shifts down in the SE and up in the NW, while the restoration final mean does not show this pattern and is very close to Fig. 8 interpolation-based model along much of its length. This difference is likely related to the differences in the deformed and restored

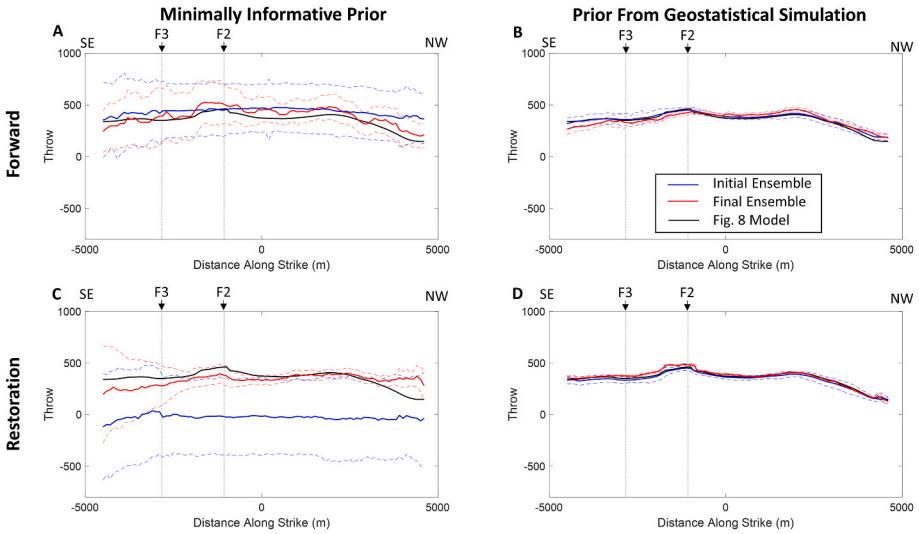

Fig. 11. Along-strike throw profiles on fault F1 based on offset of the top horizon (Horizon C). Solid lines show ensemble means for the initial ensemble (blue), final ensemble (red), and the model from Fig. 8 (black). Dashed blue and red lines show one standard deviation limits around the ensemble means. F2 and F3 mark the locations where faults F2 and F3 intersect fault F1 in Fig. 8. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

state RMS misfits from the data. The restoration method fit the deformed state data better, and it is thus similar to Fig. 8 model, which is based on interpolation in the deformed state without consideration of the restored state. The forward method produces a model that better matches a plane in the restored state but does not match the deformed state data quite as well. It is hard to say that either result is more correct, but it is important to consider how the modeling method may affect whether some data are better fit than others.

5. Discussion

5.1. Use of EKI in structural geologic modeling

Results show that EKI can be used for building structural models that incorporate kinematic constraints. The method is well suited to large models involving multiple faults and horizons and spatially varying parameters. We have been able to model a realistic system involving five different faults with spatially varying fault geometry, horizon geometry, and fault displacement distribution, in which we invert simultaneously for thousands of parameters.

Tests with a synthetic model show that the ensemble average model is generally able to achieve a good approximation of the true model. The ensemble also provides a measure of uncertainty in the parameters, which can be measured through statistics such as the standard deviation, but the usefulness of this is undermined by the phenomenon of ensemble collapse. Covariance inflation can mitigate ensemble collapse, while localization provides a lesser benefit. Large ensemble sizes are also able to avoid ensemble collapse if they are practical. The use of a dummy variable provides a simple way to identify ensemble collapse and therefore to determine whether uncertainty estimates from the method can be trusted.

Some model parameters are, however, better fit than others. In the synthetic model, the horizon geometry (Fig. 5) was better fit than the fault geometry and displacement (Fig. 6). This may be due to a greater number of data points constraining the horizons, but it may also be due to more complex or difficult to fit non-local parameters. For instance, fault strike and dip showed considerable uncertainty (Fig. 4), and gridded fault surface or displacement parameters will likely have to change to achieve a good fit to data as the orientation of the reference plane changes. By contrast, horizons were originally horizontal in the synthetic model, and their depths were precisely fit (Fig. 4). In the case of fault displacement, the use of an elliptical trend meant a more complex trend function than for horizons or the fault surface. The results, as shown in Fig. 6, suggest that this trend alone was enough to capture much of the information available about the displacement, with the residual field being within error of zero. In the case of a fault that adheres less well to the assumption of an elliptical displacement field (as might be true in a real-world case such as the Emerald Field), however, the residual would likely be more important.

With the Emerald Field case, we have illustrated how our method applies to a particular real-world case. The analysis of fault throw in Fig. 11 is particularly indicative of how this method can help to provide best estimates and quantified uncertainties for a variable of interest. Fault throw and displacement and their variations along strike are useful in understanding fault scaling laws (Kim and Sanderson, 2005) and fault growth history (Jackson, et al., 2017) and interpreting geologic structures (Barnett et al., 1987; Freeman et al., 1990). In petroleum geology, fault throw is of critical interest due to its effects on reservoir volumes and fault sealing (Røe et al., 2014; Yielding, 2015). Similar concerns apply to CO_2 storage (e.g. Michie et al., 2021; Wu et al., 2021). Seismic hazard analysis is another application in which quantifying throw or slip (or their rates) on a fault is of primary importance, and in which along-strike variations in fault throw affect predictions (e.g. Faure Walker et al., 2019). It is also a quantity for which direct observations are frequently unavailable. In seismic imaging, horizon cutoffs along the fault are frequently poorly resolved, so horizon geometries at some distance from the fault together with a model of how displacement decreases away from the fault must be used to infer throw (Godefroy et al., 2018). That is what we have done here, since our dataset included horizon depths in the vicinity of the fault, but not any horizon cutoffs on the fault surface. Our results (Fig. 11 and S6 in supplementary material) show that with this information we were able to recover details of the fault throw. Our ensemble mean is similar to a purely interpolation-based solution (black lines), although whether it is better or worse cannot be said with certainty since the absolute truth is not known. Perhaps more importantly, our results show a reduction of uncertainty in the fault throw over much of its length when compared to an informative, but interpolation-based prior (Fig. 11b and d).

The Emerald Field case also illustrates the fact that the prior information is important, as in any Bayesian method, and must be considered when creating the initial ensemble. We found that we can fit the major features of even a large model with a minimally informative prior, but if a more informative prior is available, it will improve the results (Figs. 9–11). We also found that the EKI significantly reduced the standard deviation of the ensemble in both cases, showing its ability to find the best-fitting, kinematically balanced models. When used with an informative prior, the result will depend in part on how much uncertainty is in the prior and on the uncertainty assigned to data in the EKI,

but it does suggest that if an initially wide range of models is considered, the EKI can help to narrow down the possibilities. We also see in Fig. 10 that the forward modeling method results in lower misfit in the restored state than in the deformed state, with the difference being largest for the minimally informative prior. This result likely reflects the fact that the minimally informative prior for the restored state is created by zero-mean perturbations about a plane with no guarantee of being close to the deformed-state data. With the restoration method (Fig. S5 in supplementary material) the misfit is instead somewhat higher in the restored state than the deformed state, suggesting that each method favors fitting the data in the state in which the model is parameterized.

In the examples presented here, we have jointly fit for horizon geometry, fault geometry, and fault kinematics. It is possible, of course, to fit for only some of these parameters. For instance, the method can be used to fit horizon and fault geometry in the deformed state without considering fault kinematics. In that case, the methods presented here provide an alternative to interpolation-based static modeling methods, and may be especially useful for sparse data cases where the interpolation is more uncertain. By using a hierarchical parameterization, we are also able to incorporate uncertainty in the variogram range and sill rather than using single prescribed or best-estimate values. In the field of static three-dimensional geological modeling, there is considerable interest in quantifying uncertainty (Schweizer et al., 2017; Wellmann and Caumon, 2018; Yang et al., 2019; Pirot et al., 2022), and the methods that we present here, or similar ones based on EKI, could provide a path towards accomplishing these goals.

Alternatively, an approach often used with kinematic models such as trishear is to consider horizon geometry as known and to fit only for fault kinematic parameters, displacement, and possibly fault geometry. That can easily be done with EKI, in which case EKI provides an alternative to methods such as global optimization and Markov chain Monte Carlo that have been previously used for these problems. It requires significantly fewer model evaluations than MCMC methods, which is important for computationally demanding models, but it still provides an estimate of uncertainty unlike methods aiming at a single best-fit solution. It is somewhat like the RML method used by Cardozo and Aanonsen (2009) for trishear, but with the difference that the ensemble members are used together to update their parameters rather than each ensemble member being fit individually.

5.2. Restoration and forward modeling methods

Incorporating fault kinematics into the model requires transforming the model between a restored state and a deformed state. Constraints are applied to horizon geometry in both states: It should be as flat as possible in the restored state and should match the observed data in the deformed state. When horizon geometry is one of the unknowns to be fit for in the inversion, it is possible to parameterize it in either the deformed or restored state, with the fault kinematic model then used to find the horizon geometry in the other state. We have experimented with both methods: a restoration-based approach in which horizon geometry is fit in the deformed state and then restored, and a forward modeling approach in which horizon geometry is fit in the restored state and is then deformed. Based on the results of our synthetic model, both methods work well.

Restoration is widely used in the field of fault kinematics. It provides a simple method to test the validity of a model constructed from observed present-day data. In such an approach, there is an iterative process of constructing the present-day model and then restoring it. EKI replaces such iteration on a single model with an automated iterative process using an ensemble of possible models. One advantage of the restoration approach is that it is easier to use informative priors since prior information is more available for the deformed state, but in the forward modeling approach, we were able to get around this limitation by kinematically restoring an ensemble of deformed-state prior models to create an informative restored-state prior.

The forward modeling approach provides the advantage of greater simplicity. It requires fewer model parameters because the restored-state horizon geometry is continuous without the need to parameterize multiple fault blocks. In addition, the hierarchical parameterization of the horizons in terms of a Gaussian random field means that the prior mean for each horizon is a flat surface. Since the deformed state surface is more irregular than the restored state surface, the inversion does not have to move as far from the prior mean when modeling horizons in the restored state as when doing so in the deformed state. Another advantage of forward modeling, which is specific to the kinematic model of Georgsen et al. (2012), is that the equations (Eq. (4) and B1-B5) do not have an analytical inverse, and thus the computation of the forward deformation is computationally simpler. A potential drawback of the forward modeling approach that we observed with the Emerald Field is that it is more likely to better fit the expected planar restored state at the expense of the directly observed deformed-state data, although giving a lower uncertainty to the deformed-state data can reduce this issue. When trying to simultaneously solve for fault kinematics and produce a model that interpolates between observed data, our results suggest that forward modeling-based approaches as well as restoration may be useful, and they are likely to be somewhat simpler.

5.3. Limitations of the proposed method

We have considered three types of data to constrain our models: deformed-state horizon geometries, deformed-state fault geometries, and restored-state horizon misfits from an originally planar geometry. Deformed-state horizon and fault geometry data can be derived from seismic reflection images or well logs and are subject to uncertainty in the interpretation of these datasets (Bond, 2015). The requirement for a planar restored state is based on the assumption of originally horizontal stratigraphy and may be subject to uncertainties due to inaccuracies of the kinematic model used for restoration or to the presence of a more complicated regional trend than the planar one we have used. Non-conformable beds in the restored state (e.g. growth strata) may be present, and we have not considered that. In all cases, values for the data uncertainties must be prescribed. The α term in Eq. (2) can help to mitigate the effect of potentially underestimated uncertainties, but the ratios between the uncertainties of different data types remain unchanged. The need to estimate at least the relative uncertainty of the different data types is thus a limitation of the proposed method. In particular, when a model cannot both exactly fit present-day data and restore perfectly flat, there is a tradeoff between those two constraints. We have also not considered correlation of the data, which may be present, for instance due to effects of time-to-depth conversion.

Some aspects of model uncertainty are also not considered in our current approach. For instance, the number of faults in the model, the order in which they truncate each other, and the order in which they slip, are all prescribed. Such problems are an area of ongoing work in other forms of geological modeling as well (Godefroy et al., 2019), and developing ways to consider uncertainty in these issues could be a target of future work. EnKF and EKI require a fixed-dimension parameter space, which makes changing the number of faults difficult. One option would be to give each fault an existence parameter, corresponding to the probability of its inclusion in the model, as in Cherpeau et al. (2012). Alternatively, the reversible jump MCMC method could be used, as has been done for stratigraphic modeling (Charvin et al., 2009), although as with other MCMC methods, this will likely require more model evaluations than does EKI.

5.4. Future directions

The methods that we have developed can be straightforwardly applied to other kinematic models, besides the Georgsen et al. (2012) model used here. Three-dimensional trishear models (Cristallini and Allmendinger, 2001; Cristallini et al., 2004; Cardozo, 2008), for

instance, require specifying how the fault slip, propagation rate, and trishear angle vary along strike, and EKI would provide a way to invert for those spatially variable parameters without requiring simplistic assumptions such as a linear gradient. Trishear parameters may also change over time (Allmendinger, 1998; Allmendinger et al., 2004), but such parameters are usually held constant in inversions, or at most a single change is inverted for (Oakley et al., 2018). With EKI being suited to many-parameter models, complex variations of trishear parameters with time, as well as in space, could be inverted for.

As noted in section 2.3, the EnKF is already widely used in reservoir simulations and history matching, but uncertainty in the structural geology is often not considered. In studies such as Seiler et al. (2010a, b) and Irving et al. (2014) that have considered this uncertainty, an ensemble of structural models is first created statistically, and then used as the prior for history matching using the EnKF, in which the structural model is updated as dynamic fluid flow data are assimilated. These methods do not use the EnKF to match fault and horizon geometry data or ensure that the structural models are kinematically restorable. A possible future direction is, therefore, to combine our methods with EnKF-based history-matching methods, either by using the results of the structural restoration EKI as the prior ensemble for a reservoir modeling EnKF, or by simultaneously fitting structural data and reservoir production data in a single large inversion. Such a methodology could also be combined with the use of software such as RMS and Havana to create prior geostatistical realizations, which are then improved through kinematic restoration, and then further used to model fluid flow with consideration of the uncertainty in structural features such as fault geometry and across-fault juxtapositions. Seiler et al. (2010a, 2010b) have shown that joint updating of the flow and structural models using flow data can improve both, and including structural data and structural balancing constraints would likely offer further improvements. Understanding fault throw is particularly important in history matching since it controls unit juxtaposition and flow across faults (Seiler et al., 2010b), and as shown in Fig. 11, our proposed methods can help to quantify throw and its uncertainty.

Another potential use for the EKI is with more computationally intensive methods such as geomechanical restoration (e.g., Stockmeyer et al., 2018). In the examples that we have shown, the number of evaluations of the forward model is in the 100s to low 1000s, which is much less than MCMC requires for even simple kinematic models (e.g., Oakley and Fisher, 2015). EKI has already been applied to geophysical inversions (e.g. Muir and Tsai, 2020; Tso et al., 2021), and in structural geology it could likely also be applied to physics-based modeling methods as an alternative to kinematic ones.

A final possibility for future work is to incorporate uncertainty in time-to-depth conversions into the modeling workflow presented here. For instance, a stochastic depth conversion method could be used to create the data realizations, rather than assigning an uncertainty to already depth-converted data, which would help to account for correlations and spatial variation in the uncertainty. Further, depth conversion could be incorporated into the forward model with interval velocities as parameters for the EKI to fit for. Different velocity models could be validated through kinematic restoration or by area-depth-strain analysis (Totake et al., 2017), with the EKI providing estimates of the most likely velocity model and its uncertainty.

6. Conclusions

Ensemble Kalman Inversion is a promising tool for stochastic modeling in structural geology. Using this tool, we were able to simultaneously invert for horizon geometries, fault geometries, displacement on faults, and parameters of a kinematic model in order to create an ensemble of kinematically restorable models. The method can be applied to three-dimensional models, rather than just cross sections, and to cases involving multiple faults. It is thus applicable to models of much greater complexity than were possible with previous attempts to apply data inversion to fault kinematic modeling. EKI also allows uncertainty in model parameters to be estimated, but it is necessary to mitigate ensemble collapse using methods such as covariance inflation for the uncertainty estimates to be reliable. In addition to the applications demonstrated here, possible future directions for the use of EKI in structural modeling include applications to other kinematic models, allowing the use of more physically realistic but computationally intensive mechanical methods, and linking kinematic modeling workflows with others such as history matching and depth conversion.

Author statment

David Oakley: Conceptualization, Methodology, Software, Writing – original draft. **Nestor Cardozo:** Conceptualization, Supervision, Writing – review & editing. **Ariel Almendral Vazquez:** Methodology, Writing – review & editing. **Per Røe:** Methodology, Writing – review & editing.

Code availability

Matlab codes for the methods described in this paper can be downloaded at: https://github.com/dosoakley/EKIFaultFold. This repository includes the files necessary to run the inversions and recreate the figures for the synthetic model (Fig. 2) shown in this paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

A link to the code and the data for the synthetic model is in the paper. We do not have permission to share the data for the Emerald Field example.

Acknowledgements

This research was funded by the Norwegian Ministry of Education and Research (Kunskapsdepartamentet) who funded a 2 years postdoc for D. Oakley at UiS. We thank Emerson for use of the RMS software and the Norwegian Computing Center (Norsk Regnesentral) for use of the Havana software. We thank two anonymous reviewers for their comments, which have improved this manuscript.

Appendix D. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsg.2023.104868.

Appendix A. The EKI algorithm

The EKI algorithm is similar to the EnKF, but not identical. Traditional formulations of the EnKF, such as that given by Aanonsen et al. (2009), involve a matrix $\boldsymbol{H}$, which linearly relates measurements to model parameters and often makes use of an augmented state vector that combines $\boldsymbol{u}$ and G

($\boldsymbol{u}$). They are also developed for time-varying systems, in which the state vector includes both time-varying state variables and other unknown parameters (Aanonsen et al., 2009), in place of our vector $\boldsymbol{u}$. For inversion, rather than data assimilation, Iglesias et al. (2013b) derive a simpler formulation for updating $\boldsymbol{u}$ directly, which with the addition of α_n (Iglesias, 2016) is given in Eq. (2). Eq. (2) is basically the same formula as Ma and Bi (2019, their Eq. 63), and it is closely related to the ensemble smoother with multiple data assimilation (Emerick and Reynolds, 2013).

The empirical covariance matrices in Eq. (2) can be calculated by the formulas

$$\mathbf{C}_n^{\mathbf{uG}} = \frac{1}{N-1}\sum_{j=1}^{N}\left(\boldsymbol{u}_n^{(j)} - \overline{\boldsymbol{u}}_n\right) \otimes \left(\boldsymbol{G}_n^{(j)} - \overline{\boldsymbol{G}}_n\right), \tag{A1}$$

and

$$\mathbf{C}_n^{\mathbf{GG}} = \frac{1}{N-1}\sum_{j=1}^{N}\left(\boldsymbol{G}_n^{(j)} - \overline{\boldsymbol{G}}_n\right) \otimes \left(\boldsymbol{G}_n^{(j)} - \overline{\boldsymbol{G}}_n\right), \tag{A2}$$

where $\boldsymbol{G}_n^{(j)}$ is the vector of forward model results $G(\boldsymbol{u}_n^{(j)})$, $\overline{\boldsymbol{u}}_n$ and $\overline{\boldsymbol{G}}_n$ denote averages over all ensemble members, and $\otimes$ denotes the Kronecker product.

The term $\mathbf{C}_n^{\mathbf{uG}}\left(\mathbf{C}_n^{\mathbf{GG}} + \alpha_n\boldsymbol{\Gamma}\right)^{-1}$ in Eq. (2) is the Kalman-type gain matrix, $\boldsymbol{K}$ (Iglesias et al., 2013). To calculate $\boldsymbol{K}$, we make use of a subspace pseudo-inverse method described by Evensen (2004, 2009), which makes the calculation faster and is especially helpful when working with many model parameters and data.

To choose the regularizing term α_n, we use the EKI-DMC method of Iglesias and Yang (2021). In this method, one first calculates the least-squares misfit-functional, Φ_n, for each ensemble member at each iteration:

$$\Phi_n = \left\{\frac{1}{2}\left\|\boldsymbol{\Gamma}^{-\frac{1}{2}}\left(\boldsymbol{d} - G\left(\boldsymbol{u}_n^{(j)}\right)\right)\right\|^2\right\}_{j=1}^{N}. \tag{A3}$$

The value of α_n is then calculated as:

$$\alpha_n^{-1} = min\left\{max\left\{\frac{M}{2\langle\Phi\rangle_n}, \sqrt{\frac{M}{2\langle\Phi,\Phi\rangle_n}}\right\}, 1 - t_n\right\}, \tag{A4}$$

where M is the number of data points, $\langle\Phi\rangle_n$ and $\langle\Phi,\Phi\rangle_n$ are the mean and variance of Φ_n, respectively, and t_n is given by the following equation:

$$t_n = \begin{cases} \sum_{j=0}^{n-1}\alpha_j^{-1} & if\ n \geq 1 \\ 0 & if\ n = 0. \end{cases} \tag{A5}$$

The algorithm stops when $\alpha_n^{-1} = 1 - t_n$, in order to fulfill the requirement that the sum of α_n over all iterations equals 1.

Appendix B. Fault Displacement Model

The elliptical displacement model takes the form:

$$d_{ellipse}(u,v) = 2d_{max}(1 - r(u,v))\sqrt{\frac{(1 + r(u,v))^2}{4} - r(u,v)^2}, \tag{B1}$$

and:

$$r(u,v) = \sqrt{\left(\frac{u - u_0}{l_u}\right)^2 + \left(\frac{v - v_0}{l_v}\right)^2}. \tag{B2}$$

The contour of zero displacement μ0 = 0 (or r = 1) is used to define the faut tip line.

The equations for the three-dimensional displacement field are:

$$D_u(u,v,w) = 0 \tag{B3}$$

$$D_v(u,v,w) = \begin{cases} \gamma \bullet d(u,v) \bullet \left[1 - \frac{|w - f(u,v)|}{R}\right]^2, f(u,v) \leq w \leq f(u,v) + R \\ (\gamma - 1) \bullet d(u,v) \bullet \left[1 - \frac{|w - f(u,v)|}{R}\right]^2, f(u,v) - R \leq w < f(u,v) \\ 0, w < f(u,v) - R\ or\ w > f(u,v) + R \end{cases} \tag{B4}$$

$$D_w(u,v,w) = f(u, v + D_v(u,v,w)) - f(u,v). \tag{B5}$$

Thus, we assume no strike-slip displacement (D_u is zero everywhere). These equations produce deformation in both the hanging wall and footwall, with displacement decreasing away from the fault and towards the tip line. The structure contours of Fig. 2b illustrate this pattern of displacement by the offset of a horizon that was originally at 1100 m depth but is now deeper in the hanging wall and shallower in the footwall.

Appendix C. Equations for Localization and Covariance Inflation

In the localization method of Zhang and Oliver (2010), bootstrap resampling of the ensemble of model parameters, predictions, and observed data is used to create N_B realizations of the Kalman gain matrix $\boldsymbol{K}$. From these realizations, a bootstrap estimate of the variance in each element of $\boldsymbol{K}$ is calculated:

$$\left(\sigma^*_{K_{i,j}}\right)^2 = \frac{1}{N_B}\sum_{n=1}^{N_B}\left(K^*_{n,i,j} - K_{i,j}\right)^2, \tag{C1}$$

where $\boldsymbol{K}^*_{\boldsymbol{n}}$ is the nth bootstrapped $\boldsymbol{K}$, and *(i,j)* are the indices of matrix elements. From this equation, the terms of the localization, or screening matrix, $\boldsymbol{S}$, are calculated as:

$$S_{i,j} = \frac{1}{1 + V^2_{i,j}\left(1 + \frac{1}{\sigma_S^2}\right)}, \tag{C2}$$

where $\boldsymbol{V}$ is the squared coefficient of variation:

$$V^2_{i,j} = \frac{\left(\sigma^*_{K_{i,j}}\right)^2}{K^2_{i,j}}. \tag{C3}$$

$\boldsymbol{S}$ is then used to calculate a screened Kalman gain matrix ($\boldsymbol{K}^{\boldsymbol{S}}$) by the multiplication:

$$\boldsymbol{K}^{\boldsymbol{S}} = \boldsymbol{S} \circ \boldsymbol{K}, \tag{C4}$$

where $\circ$ denotes the Hadamard or Schur (i.e., elementwise) product. This method does include one user-chosen value, σ_S, which determines the strength of the localization. Zhang and Oliver (2010) suggest a value of 0.6, which we follow.

For localization by the method of Evensen (2009a,b), N_t vectors of N elements each are drawn from a standard normal distribution and then shifted and scaled if necessary, so that the mean and standard deviation of each vector are exactly 0 and 1, respectively. These dummy parameters are then added to the matrix of model parameter realizations and updated using Eq. (2). After updating, the standard deviation of each dummy parameter is calculated as:

$$\sigma_{t,i} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(t_{i,j} - \overline{t_i}\right)^2}, \tag{C5}$$

where i is the index for each of the N_t dummy parameters, j is the index for each of the N ensemble members, and $\overline{t_i}$ denotes the average of $t_{i,j}$ over all j values. The inflation factor is then:

$$\rho = \frac{1}{\frac{1}{N_t}\sum_{i=1}^{N_t}\sigma_{t,i}}. \tag{C6}$$

The updated ensemble of model parameters from Eq. (2) is inflated by changing its variance while keeping its mean constant through the update equation:

$$\boldsymbol{u}^{(j)}_{n+1} = \rho\left(\boldsymbol{u}^{(j)}_{n+1} - \overline{\boldsymbol{u}_{n+1}}\right) + \overline{\boldsymbol{u}_{n+1}} \tag{C7}$$

References

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

Abrahamsen, P., Kvernelv, V., Barker, D., 2018. Simulation of Gaussian random fields using the fast Fourier transform (fft). In: Conference Proceedings, ECMOR XVI - 16th European Conference on the Mathematics of Oil Recovery, Sep 2018, vol. 2018, pp. 1–14. https://doi.org/10.3997/2214-4609.201802134.

Allmendinger, R.W., 1998. Inverse and forward numerical modeling of trishear fault-propagation folds. Tectonics 17 (4), 640–656.

Allmendinger, R.W., Zapata, T., Manceda, R., Dzelalija, F., 2004. Trishear kinematic modeling of structures, with examples from the Neuquén Basin, Argentina. In: McClay, K.R. (Ed.), Thrust Tectonics and Hydrocarbon Systems, vol. 82. AAPG Memoir, pp. 356–371, 2004.

Aydin, O., Caers, J.K., 2017. Quantifying structural uncertainty on fault networks using a marked point process within a Bayesian framework. Tectonophysics 712–713, 101–124.

Brandes, C., Tanner, D.C., 2014. Fault-related folding: a review of kinematic models and their application. Earth Sci. Rev. 138, 352–370.

Bond, C.E., 2015. Uncertainty in structural interpretation: lessons to be learnt. J. Struct. Geol. 74, 185–200. https://doi.org/10.1016/j.jsg.2015.03.003.

Cardozo, N., Røe, P., Soleng, H.H., Fredman, N., Tveranger, J., Schueller, S., 2008. A methodology for efficiently populating faulted corner point grids with strain. Petrol. Geosci. 14 (2), 205–216.

Cardozo, N., 2008. Trishear in 3D. Algorithms, implementation, and limitations. J. Struct. Geol. 30, 327–340. https://doi.org/10.1016/j.jsg.2007.12.003.

Cardozo, N., Aanonsen, S., 2009. Optimized trishear inverse modeling. J. Struct. Geol. 31, 546–560. https://doi.org/10.1016/j.jsg.2009.03.003.

Cardozo, N., Jackson, C.A.-L., Whipp, P.S., 2011. Determining the uniqueness of best-fit trishear models. J. Struct. Geol. 33, 1063–1078. https://doi.org/10.1016/j.jsg.2011.04.001.

Cardozo, N., Brandenburg, J.P., 2014. Kinematic modeling of folding above listric propagating thrusts. J. Struct. Geol. 60, 1–12. https://doi.org/10.1016/j.jsg.2013.12.004.

Cardozo, N., Oakley, D., 2019. Inverse modeling for possible rather than unique solutions. J. Struct. Geol. 125, 285–295. https://doi.org/10.1016/j.jsg.2018.05.026.

Caumon, G., Laurent, G., Pellerin, J., Cherpeau, N., Lallier, F., Merland, R., Bonneau, F., 2013. Current Bottlenecks in Geomodeling Workflows and Ways Forward. Bulletin of Canadian Petroleum Geology (Geo- modeling Special issue - Gussow Conference.

Chada, N.K., Iglesias, M.A., Roininen, L., Stuart, A.M., 2018. Parameterizations for ensemble Kalman inversion. Inverse Probl. 34, 055009 https://doi.org/10.1088/1361-6420/aab6d9.
Charvin, K., Gallagher, K., Hampson, G.L., Labourdette, R., 2009. A Bayesian approach to inverse modelling of stratigraphy, part 1: method. Basin Res. 21, 5–25. https://doi.org/10.1111/j.1365-2117.2008.00369.x.
Chen, Y., Oliver, D.S., 2012. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math. Geosci. 44, 1–26. https://doi.org/10.1007/s11004-011-9376-z.
Chen, Y., Oliver, D.S., 2013. Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. Comput. Geosci. 17, 689–703. https://doi.org/10.1007/s10596-013-9351-5.
Cherpeau, N., Caumon, G., 2015. Stochastic structural modelling in sparse data situations. Petrol. Geosci. 21, 233–247. https://doi.org/10.1144/petgeo2013-030.
Cherpeau, N., Caumon, G., Caers, J., Lévy, B., 2012. Method for stochastic inversion modeling of fault geometry and connectivity using flow data. Math. Geosci. 44, 147–168. https://doi.org/10.1007/s11004-012-9389-2.
Cristallini, E.O., Allmendinger, R.W., 2001. Pseudo 3-D modeling of trishear fault-propagation folding. J. Struct. Geol. 23, 1883–1899.
Cristallini, E.O., Giambiagi, L., Allmendinger, R.W., 2004. True three-dimensional trishear: a kinematic model for strike-slip and oblique-slip deformation. GSA Bulletin 116 (7/8), 938–952. https://doi.org/10.1130/B25273.1.
Dahlstrom, C.D.A., 1969. Balanced cross sections. Can. J. Earth Sci. 6, 743–757.
Davis, J.C., 2002. Statistics and Data Analysis in Geology, third ed. John Wiley & Sons, New York, p. 638.
de la Varga, M., Wellmann, J.F., 2016. Structural geologic modeling as an inference problem: a Bayesian perspective. Interpretation 4 (3). https://doi.org/10.1190/INT-2015-0188.1. SM1–SM16.
Elliott, D., 1983. The construction of balanced cross sections. J. Struct. Geol. 5, 101.
Emerick, A.A., Reynolds, A.C., 2012. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. Comput. Geosci. 16, 639–659. https://doi.org/10.1007/s10596-012-9275-5.
Emerick, A.A., Reynolds, A.C., 2013. Ensemble smoother with multiple data assimilation. Comput. Geosci. 55, 3–15. https://doi.org/10.1016/j.cageo.2012.03.011.
Erslev, E.A., 1991. Trishear fault-propagation folding. Geology 19, 617–620.
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99 (C5), 10143–10162. https://doi.org/10.1029/94JC00572.
Evensen, G., van Leeuwen, P.J., 1996. Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic Model. Mon. Weather Rev. 124, 85–96. https://doi.org/10.1175/1520-0493(1996)124<0085:AOGADF>2.0.CO, 2.
Evensen, G., 2004. Sampling strategies and square root analysis schemes for the EnKF. Ocean Dynam. 54, 539–560.
Evensen, G., 2009a. The ensemble Kalman filter for combined state and parameter estimation. IEEE Control Syst. Mag. 29 (3), 83–104. https://doi.org/10.1109/MCS.2009.932223.
Evensen, G., 2009b. Data Assimilation: the Ensemble Kalman Filter, second ed. Springer, Berlin.
Evensen, 2018. Analysis of iterative ensemble smoothers for solving inverse problems. Comput. Geosci. 22, 885–908.
Faure Walker, J.P., Visini, F., Roberts, G., Galasso, C., McCaffrey, K., Mildon, Z., 2019. Variable fault geometry suggests detailed fault-slip-rate profiles and geometries are needed for fault-based probabilistic seismic hazard assessment (PSHA). Bull. Seismol. Soc. Am. 109 (1), 110–123. https://doi.org/10.1785/0120180137.
Georgsen, F., Røe, P., Syversveen, A.,R., Lia, O., 2012. Fault displacement modelling using 3D vector fields. Comput. Geosci. 16, 247–259. https://doi.org/10.1007/s10596-011-9257-z.
Godefroy, G., Caumon, G., Ford, M., Laurent, G., Jackson, C.A.-L., 2018. A parameteric fault displacement model to introduce kinematic control into modeling faults from sparse data. Interpretation 6 (2). https://doi.org/10.1190/INT-2017-0059.1. B1-B13.
Godefroy, G., Caumon, G., Laurent, G., Bonneau, F., 2019. Structural interpretation of sparse fault data using graph theory and geological rules. Math. Geosci. 51 (8), 1091–1107. https://doi.org/10.1007/s11004-019-09800-0.
Grose, L., Laurent, G., Aillères, L., Armit, R., Jessell, M., Cousin-Dechenaud, T., 2018. Inversion of structural geology data for fold geometry. J. Geophys. Res. Solid Earth 123, 6318–6333. https://doi.org/10.1029/2017JB015177.
Grose, L., Aillères, L., Laurent, G., Armit, R., Jessell, M., 2019. Inversion of geological knowledge for fold geometry. J. Struct. Geol. 119, 1–14. https://doi.org/10.1016/j.jsg.2018.11.010.
Grose, L., Ailleres, L., Laurent, G., Caumon, G., Jessell, M., Armit, R., 2021. Modelling of faults in LoopStructural 1.0. Geosci. Model Dev. (GMD) 14, 6197–6213. https://doi.org/10.5194/gmd-14-6197-2021.
Groshong Jr., R.H., Bond, C., Gibbs, A., Ratliff, R., Wiltschko, D.V., 2012. Preface: structural balancing at the start of the 21st century: 100 years since Chamberlin. J. Struct. Geol. 41, 1–5. https://doi.org/10.1016/j.jsg.2012.03.010.
Hanke, M., 1997. A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. Inverse Probl. 13, 79–95.
Hoffman, K.S., Neave, J.W., 2007. The fused fault block approach to fault network modelling. In: Jolley, S.J., Barr, D., Walsh, J.J., Knipe, R.J. (Eds.), Structurally Complex Reservoirs, vol. 292. Geological Society, London, Special Publications, pp. 75–87.
Houtekamer, P.L., Mitchell, H.L., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129, 123–137. https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO, 2.
Iglesias, M.A., Law, K.J.H., Stuart, A.M., 2013a. Evaluation of Gaussian approximations for data assimilation in reservoir models. Comput. Geosci. 17, 851–885. https://doi.org/10.1007/s10596-013-9359-x.
Iglesias, M.A., Law, K.J.H., Stuart, A.M., 2013b. Ensemble Kalman methods for inverse problems. Inverse Probl. 29, 045001 https://doi.org/10.1088/0266-5611/29/4/045001.
Iglesias, M.A., 2016. A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. Inverse Probl. 32, 025002 https://doi.org/10.1088/0266-5611/32/2/025002.
Iglesias, M., Yang, Y., 2021. Adaptive regularization for ensemble Kalman inversion. Inverse Probl. 37, 025008 https://doi.org/10.1088/1361-6420/abd29b.
Irving, A.D., Kuznetsov, D., Robert, E., Manzocchi, T., Childs, C.J., 2014. Optimization of uncertain structural parameters with production and observation well data. SPE Reservoir Eval. Eng. 17 (4), 547–558. https://doi.org/10.2118/171558-PA.
Judge, P.A., Allmendinger, R.W., 2011. Assessing uncertainties in balanced cross sections. J. Struct. Geol. 33 (4), 458–467.
Lacerda, J.M., Emerick, A.A., Pires, A.P., 2019. Methods to mitigate loss of variance due to sampling errors in ensemble data assimilation with non-local model parameters. J. Petrol. Sci. Eng. 172, 690–706. https://doi.org/10.1016/j.petrol.2018.08.056.
Laurent, G., Caumon, G., Bouziat, A., Jessell, M., 2013. A parametric method to model 3D displacements around faults with volumentric vector fields. Tectonophysics 590, 83–93.
Ma, X., Bi, L., 2019. A robust adaptive iterative ensemble smoother scheme for practical history matching applications. Comput. Geosci. 23, 415–442. https://doi.org/10.1007/s10596-018-9786-9.
McKay, M.D., Cooper, W.J., Beckman, R.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.
Michie, E.A.H., Mulrooney, M.J., Braathen, A., 2021. Fault interpretation uncertainties using seismic data, and the effects on fault seal analysis: a case study from the Horda Platform, with implications for CO_2 storage. Solid Earth 12, 1259–1286. https://doi.org/10.5194/se-12-1259-2021.
Moretti, I., Lepage, F., Guiton, M., 2006. KINE3D: a new 3D restoration method based on a mixed approach linking geometry and geomechanics. Oil & Gas Science and Technology – Rev. IFP 61 (2), 277–289. https://doi.org/10.2516/ogst:2006021.
Muir, J.B., Tsai, V.C., 2020. Geometric and level set tomography using ensemble Kalman inversion. Geophys. J. Int. 220, 967–980. https://doi.org/10.1093/gji/ggz472.
Oakley, D.O.S., Fisher, D.M., 2015. Inverse trishear modeling of bedding dip data using Markov chain Monte Carlo methods. J. Struct. Geol. 80, 157–172. https://doi.org/10.1016/j.jsg.2015.09.005.
Oakley, D.O.S., Fisher, D.M., Gardner, T.W., Stewart, M.K., 2018. Uplift rates of marine terraces as a constraint on fault-propagation fold kinematics: examples from the Hawkswood and Kate anticlines, North Canterbury, New Zealand. Tectonophysics 724–725, 195–219. https://doi.org/10.1016/j.tecto.2017.12.021.
Pirot, G., Joshi, R., Giraud, J., Lindsay, M.D., Jessell, M.W., 2022. loopUI-0.1: Uncertainty Indicators to Support Needs and Practices in 3D Geological Modelling Uncertainty Quantification. https://doi.org/10.5194/gmd-2021-377. Preprint.
Regalla, C., Fisher, D., Kirby, E., 2010. Timing and magnitude of shortening within the inner fore arc of the Japan Trench. J. Geophys. Res. 115, B03411 https://doi.org/10.1029/2009JB006603.
Røe, P., Georgsen, F., Abrahamsen, P., 2014. An uncertainty model for fault shape and location. Math. Geosci. 46, 957–969. https://doi.org/10.1007/s11004-014-9536-z.
Røe, P., Almendral Vazquez, A., Hanea, R., 2016. Distinguishing signal from noise in history matching – analysis of ensemble collapse on a synthetic data set. In: ECMOR XV-15th European Conference on the Mathematics of Oil Recovery. European Association of Geoscientists & Engineers. https://doi.org/10.3997/2214-4609.201601819.
Røe, P., Goodwin, H., Aker, E., Zdanowicz, H., Kvernelv, V.B., 2021. HAVANA User Manual: Version 8.1.1. Norwegian Computing Center, p. 198.
Schaaf, A., Bond, C.E., 2019. Quantification of uncertainty in 3-D seismic interpretation: implications for deterministic and stochastic geomodeling and machine learning. Solid Earth 10, 1049–1061. https://doi.org/10.5194/se-10-1049-2019.
Schweizer, D., Blum, P., Butscher, C., 2017. Uncertainty assessment in 3-D geological models of increasing complexity. Solid Earth 8, 515–530. https://doi.org/10.5194/se-8-515-2017.
Seiler, A., Aanonsen, S.I., Evensen, G., Rivenæs, J.C., 2010a. Structural surface ucertainty modeling and updating using the ensemble Kalman filter. SPE J. 15 (4), 1062–1076. https://doi.org/10.2118/125352-PA.
Seiler, A., Aanonsen, S.I., Evensen, G., Lia, O., 2010b. An elastic grid approach for fault uncertainty modelling and updating using the Ensemble Kalman filter. In: Paper 130422-MS, Presented at the SPE EUROPEC/EAGE Annual Conference and Exhibition. https://doi.org/10.2118/130422-MS. Barcelona, Spain, 14-17 June.
Skauvold, J., Eidsvik, J., 2018. Data assimilation for a geological process model using the ensemble Kalman filter. Basin Res. 30, 730–745. https://doi.org/10.1111/bre.12273.
Stein, M., 1987. Large sample properties of simulations using Latin hypercube sampling. Technometrics 29 (2), 143–151.
Stockmeyer, J.M., Shaw, J.H., Billingsley, L.T., Plesch, A., Wales, M., Lavin, L.C., Knox, R., Finger, L., 2018. Geomechanical restoration as a tool for fractured reservoir characterization: application to the Permian Basin, west Texas. AAPG (Am. Assoc. Pet. Geol.) Bull. 102 (1), 103–128. https://doi.org/10.1306/03231716076.
Suppe, J., 1983. Geometry and kinematics of fault-bend folding. Am. J. Sci. 283, 684–721.
Tso, C.-H.M., Iglesias, M., Wilkinson, P., Kuras, O., Chambers, J., Binley, A., 2021. Efficient multiscale imaging of subsurface resistivity with uncertainty quantification

using ensemble Kalman inversion. Geophys. J. Int. 225, 887–905. https://doi.org/10.1093/gji/ggab013.
Walsh, J.J., Watterson, J., 1987. Distributions of cumulative displacement and seismic slip on a single normal fault surface. J. Struct. Geol. 9 (8), 1039–1046.
Wellmann, F., Caumon, G., 2018. 3-D Structural geological models: concepts, methods, and uncertainties. Adv. Geophys. 59, 1–121. https://doi.org/10.1016/bs.agph.2018.09.001.
Woodward, N.B., 2012. Evaluation, analysis and prediction of geologic structures. J. Struct. Geol. 41, 76–85. https://doi.org/10.1016/j.jsg.2012.02.012.
Wu, L., Thorsen, R., Ottesen, S., Meneguolo, R., Hartvedt, K., Ringrose, P., Nazarian, B., 2021. Significance of fault seal in assessing CO_2 storage capacity and containment risks – an example from the Horda Platform, northern North Sea. Petrol. Geosci. 27, petgeo2020–p2102. https://doi.org/10.1144/petgeo2020-102.
Yang, L., Hyde, D., Grujic, O., Scheidt, C., Caers, J., 2019. Assessing and visualizing uncertainty of 3D geological surfaces using level sets with stochastic motion. Comput. Geosci. 122, 54–67. https://doi.org/10.1016/j.cageo.2018.10.006.
Yielding, G., 2015. Trapping of buoyant fluids in fault-bound structures. In: Richards, F. L., Richardson, N.J., Rippington, S.J., Wilson, R.W., Bond, C.E. (Eds.), Industrial Structural Geology: Principles, Techniques and Integration, vol. 421. Geological Society, Special Publications, pp. 29–39.
Yielding, G., Freeman, B., 2016. 3-D seismic-structural workflows-Examples using the Hat Creek fault system. In: Krantz, B., Ormand, C., Freeman, B. (Eds.), 3-D Structural Interpretation: Earth, Mind, and Machine: AAPG Memoir 111, pp. 155–171.
Zhang, Y., Oliver, D.S., 2010. Improving the ensemble estimate of the Kalman gain by bootstrap sampling. Math. Geosci. 42, 327–345. https://doi.org/10.1007/s11004-010-9267-8.