

Ensemble data assimilation for earthquake sequences: probabilistic estimation and forecasting of fault stresses

Ylona van Dinther¹, Hans R. Künsch³ and Andreas Fichtner¹

¹*Seismology and Wave Physics, Department of Earth Sciences, ETH Zürich, Sonneggstrasse 5, CH-8092 Zürich, Switzerland. E-mail: y.vandinther@uu.nl*

²*Department of Earth Sciences, Utrecht University, Princetonlaan 4, 3584 CB Utrecht, the Netherlands*

³*Seminar for Statistics, Department of Mathematics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland*

Accepted 2019 February 5. Received 2019 January 28; in original form 2017 September 15

SUMMARY

Our physical understanding of earthquakes, along with our ability to forecast them, is hampered by limited indications on the current and future state of stress on faults. Integrating indirect observations, laboratory experiments and physics-based numerical modelling to quantitatively estimate this evolution is crucial. However, quantitative integrations are tenuous in light of the scarcity and uncertainty of observations and the difficulty of modelling the physics governing earthquakes. We show that observations and prior physical knowledge, along with their errors, can be efficiently integrated through the statistical framework of ensemble data assimilation (EDA), which is adopted from weather forecasting. To evaluate whether fault stress estimation and forecasting is possible, we perform a perfect model test in a subduction zone setup that mimicks a scaled laboratory experiment. Synthetic noised data on velocities and stresses from one point near the surface are assimilated using an Ensemble Kalman Filter. These data update the velocity and stress states throughout 150 ensemble members, whose dynamics is governed by a seismic cycle model. This visco-elasto-plastic forward model forecasts the system's evolution through solving Navier–Stokes equations with a strongly rate-dependent friction coefficient. The ensemble assimilation of data from a single location provides probabilistic estimates of fault stress and dynamic strength evolution, which capture the true solution exceptionally well. This is possible, because the sampled error covariance matrix contains prior information from the physics that relates velocities, stresses and pressure at the surface to those at the fault. In the analysis step, this covariance allows stress and strength distributions to be reconstructed. In the subsequent forecast step the physical equations are solved to propagate the updated states forward in time. This provides probabilistic information on the likelihood of occurrence of the next earthquake in this synthetic laboratory setting. Throughout the ensemble simulations the forecasting ability for large, quasi-periodic events turns out to be significantly better than that of a periodic recurrence model. For example, it only requires an alarm to sound for 17 per cent instead of 68 per cent of the time to forecast 70 per cent of 21 events. We show that combining our prior knowledge of physical laws with observations through a Bayesian framework provides distinct added value with respect to using observations or numerical models independently. This educational test thus shows vast potential for including physics-based information into probabilistic seismic hazard assessment using EDA. To analyze its real world potential assumptions on an exact representation of the physics in a 2-D simplified system remain to be explored.

Key words: Inverse theory; Numerical modelling; Probabilistic forecasting; Earthquake interaction, forecasting, and prediction; Statistical seismology; Dynamics and mechanics of faulting.

1 INTRODUCTION

1.1 Stress estimation and forecasting

Understanding the dynamics of the solid Earth is hampered by limited indications on the state of stress and strength of the Earth's crust. Stress and strength govern deformation over timescales from millions of years for tectonic processes down to hundreds of years for earthquake cycles and seconds for earthquakes. During an earthquake cycle stress builds up in the interseismic period between earthquakes. When accumulated stresses reach the frictional, static strength over a large enough area, an earthquake propagates and releases the elastic stresses in the coseismic period. Knowing the present and forecasting the future state of stress on seismogenic faults is thus a pivotal part of earthquake forecasting. The current state of stress and strength on faults is, however, largely unknown and inaccessible (e.g. Townend 2006 and references therein). This is one of the reasons why earthquakes cannot be predicted in terms of a specific time, location and magnitude (e.g. Geller 1997). Hence earthquake forecasting through providing a probabilistic statement about future earthquake occurrence or ground shaking is the highest achievable goal (e.g. Marzocchi *et al.* 2017). Since stress and strength change due to neighbouring, local and previous slip, creep and fluid flow processes, it is challenging to incorporate such physics-based understanding into a probabilistic assessment of seismic hazard (e.g. Segou *et al.* 2013). To do the best possible, scientists focus on including pieces of physical information through, for example, Coulomb stress transfer (e.g. King *et al.* 1994; Cattania *et al.* 2018), physical approximations driving earthquake simulators (Shaw *et al.* 2018) or physically extrapolating processes influencing different aspects of seismic hazard (e.g. Dalguer *et al.* 2017, and references therein). However, we do have some information on the state of stress on faults from different lines of research.

1.2 Observations

Observations that provide indications on the *in-situ* stress state are gradually improving (e.g. Townend 2006; Heidbach *et al.* 2018 and references therein). Moderate estimates come from nearby or regional borehole measurements (e.g. Zoback *et al.* 2003; Lin *et al.* 2013, *resp.*), earthquake focal mechanisms (e.g. Hardebeck & Michael 2006; Abolfathian *et al.* 2018; Terakawa & Hauksson 2018), active fault orientations and slip directions (e.g. Angelier 1984) and heat flow measurements (e.g. Fulton *et al.* 2013). Alternative data on gradual stress or strength changes could also come from active, coda wave interferometry (Grêt *et al.* 2006), passive image interferometry (Wegler & Sens-Schönfelder 2007), seismic imaging (e.g. Tsuji *et al.* 2014) or b-values estimates of the Gutenberg–Richter earthquake frequency–magnitude relation (e.g. Scholz 2015). Additionally, we have seen a tremendous increase in other observations that provide information on the state of the fault and earthquake physics (e.g. Dragert *et al.* 2001; Moreno *et al.* 2010; Section 4.2). This increase in data is expected to continue or accelerate (e.g. Nagao 2014; Lindsey *et al.* 2017).

All observations face technical and economic difficulties for the following reasons: (i) they are indirect, since they are limited to depths of a few kilometres, while seismogenic faults penetrate to depths of ten(s) of kilometres or rely on inversions that require assumptions, (ii) they are very sparse in space and time,

and (iii) they typically contain large (and potentially non-Gaussian) errors. These difficulties make estimation of the current state of stress on a fault highly challenging. Additionally, they make the next steps of continuous forecasting through space and time even more precarious. However, it is not known what observations are needed to provide reasonable constraints on current and future state of stress. How much data do we need? Which types of data do we need? How often do we need them? How do we account for measurement errors? And what level of accuracy is required?

1.3 Physical models

The evolution of fault stress and strength is extensively modelled using physics-based numerical forward models, which range from tectonic (e.g. Buitter *et al.* 2006; Billen 2008), to earthquake sequence (e.g. Lapusta & Barbot 2012), and earthquake timescales (e.g. Harris *et al.* 2018). Earthquake sequence models rely on observations from laboratory experiments (e.g. Dieterich 1979; Di Toro *et al.* 2011) and physical laws (e.g. Rice 1993; Lapusta *et al.* 2000; Lyakhovsky *et al.* 2011; van Dinther *et al.* 2013a; Erickson *et al.* 2017). They are important to enlarge our physical understanding of earthquake source processes (e.g. Lapusta *et al.* 2019), including understanding seismic versus slow and aseismic fault slip (e.g. Goswami & Barbot, 2018), rupture styles (e.g. Michel *et al.* 2017), rupture propagation to the trench (e.g. Noda & Lapusta 2013), and long-term recurrence patterns (e.g. Herrendoerfer *et al.* 2015). It is more challenging for models directly based on physical laws to reproduce statistical observations of seismicity. Nonetheless, various models capture several key observations, including a magnitude-invariant stress drop (e.g. Lin & Lapusta 2018), a Gutenberg–Richter earthquake magnitude–frequency of occurrence relation (e.g. Dal Zilio *et al.* 2018) and a modified Omori's relation for decay of aftershocks (e.g. Ben-Zion & Lyakhovsky 2006).

They, however, face challenges, since (i) the physics controlling earthquakes is not (well) known, (ii) the dynamics involves non-linearity, and (iii) a wide range of temporal and spatial scales is involved (e.g. Rundle *et al.* 2000, 2003; Ben-Zion 2008; Geller *et al.* 2016). Nonetheless, predictive power can be maintained if care is taken not to compromise essential physics. It is difficult, however, to calibrate these physical models to a current, natural setting and its stress level (e.g. Barbot *et al.* 2012). So how can we use the predictive power of our physics-based forward models for stress estimation and/or earthquake forecasting? Is our current physical understanding of earthquakes good enough for this? How should we improve it? How can we update our models if data becomes available? How do we account for model errors? And on a related remark, how do we account for prior expert knowledge and uncertainties in laboratory experiments?

1.4 Ensemble data assimilation

These questions could be answered quantitatively by ensemble data assimilation, which is well-suited for combining insights from observations, statistics and physical models (e.g. Evensen 2009b). It is extensively developed in meteorology and oceanography, such that all available information from observations and physical models can be combined to estimate and forecast the state of atmospheric or oceanic flow (e.g. Daley 1991; Talagrand 1997; Evensen 2009b; van Leeuwen 2009). Solid earth dynamicists instead aim

to estimate the state of flow in the solid earth and in that face similar—albeit less constrained—challenges (e.g. Fournier *et al.* 2013; de la Torre Guzman *et al.* 2014; Bocher *et al.* 2016; Gregg & Pettijohn 2016).

Data assimilation (or filtering/smoothing in the engineering community) entails solving a ‘time-dependent inverse’ problem for the computation of a probability density function (PDF) that describes the approximate dynamical numerical model solution, conditioned on noisy observations scarcely distributed in time and space (Evensen 2009b). At the same time, the measurement and modelling errors are taken into account to obtain probabilistic estimates of the involved states (i.e. physical variables) and parameters.

This is efficiently achieved through sequential data assimilation methods, which integrate information from one observation until the next. They are more straightforward than variational data assimilation methods (*cf.* Talagrand 1997) and they efficiently subtract time-dependent information in problems involving a loading history (Murakami *et al.* 2012). Within sequential data assimilation methods, we focus on ensemble- or Monte Carlo-based approaches to account for nonlinear dynamics and non-Gaussian probabilities (Section 2).

1.5 Data assimilation for seismicity

EDA, as defined in the sense of time-dependent state estimation, has not been done for estimating or forecasting the fault state for earthquake sequences. However, earthquake scientists have been combining observations with various statistical or physical models and observations in a variety of ways.

The main efforts relating to earthquake forecasting are undertaken by the statistical community based on stochastic point processes that effectively describe earthquake data (for an overview see Werner *et al.* 2011). They use probabilistic rules for generating random events and a range of Monte Carlo integration and Bayesian inference techniques conditioned on paleoseismological or historical earthquake data (e.g. Ogata 1999; Varini 2007). These techniques typically have difficulties to account for (i) measurement and model errors, (ii) non-Gaussian probabilities and (iii) sequential updating as data becomes available. These aspects are tackled by EDA using sequential Monte Carlo methods. The potential of EDA for estimating space–time occurrence of events has been extended to point processes in a pioneering study by Werner *et al.* (2011). Their forward model is a renewal model in which interevent times are randomly drawn, but it cannot exploit prior information available from physics.

From a physical model perspective, several studies have included information from earthquake sequence observations to calibrate their model through an informed trial-and-error procedure (Rundle *et al.* 2006; van Aalsburg *et al.* 2007; Zöller & Ben-Zion 2014; Yoder *et al.* 2015). In a more informed manner, Monte Carlo searches are used to calibrate parameters to reproduce induced seismicity and extrapolate it using physical considerations (e.g. Dempsey & Suckale 2017). These approaches, amongst others, cannot assimilate data as it becomes available. This could be done in an off-line version of EDA that is used to weigh a few tens of scenario’s, which solve for one or two states or parameters on a fault within a quasi-dynamic forward model (Hori *et al.* 2014a,b,c). However, this approach becomes computationally expensive for the large number of scenario’s required, because actual states are not estimated or updated to account for observations. Interestingly, for imaging aseismic slip transients (e.g. Segall & Matthews 1997), an Extended

Kalman Filter has been used to estimate the stressing rate using a rate-and-state friction model (Llenos & McGuire 2011). However, ensemble-based data assimilation methods are preferred for the nonlinear dynamics involved in earthquake sequences.

1.6 Objectives through perfect model test

The purpose of this paper is to explore whether and how EDA could be used for fault stress state estimation and eventually for earthquake forecasting. This rigorous statistical framework for combining observations and physical models has tremendous potential for quantitatively answering the questions posed in Sections 1.2 and 1.3. Actually answering these questions requires extensive future studies, but in light of the challenges posed it is not evident whether EDA is feasible for these goals. Therefore, we start to explore its feasibility in an educational test case involving a perfect model test of a scaled laboratory or analogue experiment, which reproduces earthquake sequences in a subduction zone (Corbi *et al.* 2013). In this synthetic test, a numerical model provides both the data to be assimilated and the corresponding true spatiotemporal solution to be retrieved. Such a test is the only quantitative way to explore and test a new contribution, but avoids dealing with not well-known physics. A comparison between our ensemble results and the truth demonstrates that very reasonable probabilistic estimates of the current and future state of stress, strength and slip at faults can be obtained.

We forward model sequences of earthquakes using the seismo-mechanical version of the seismo-thermo-mechanical modelling framework (e.g. van Dinther *et al.* 2013a,b; 2014). This has been shown to capture the essential physics of the stress evolution within this simplified laboratory model (van Dinther *et al.* 2013b). Stresses evolve in cycles from slow tectonic loading in the interseismic period, to coseismic release through spontaneous frictional instabilities, and postseismic rate-strengthening afterslip and viscoelastic relaxation. These natural processes are captured through solving the first physical principles of conservation of mass and momentum (and energy), while using experimentally derived visco-elasto-plastic constitutive equations (Gerya & Yuen 2007) and a slip-rate-dependent friction (van Dinther *et al.* 2013b). This continuum formulation allows us to directly apply data assimilation techniques, which have been developed to deal with nonlinearities and uncertainties in the atmospheric dynamics described by similar partial differential equations.

This forward modelling approach is described in more detail in Section 2.4. Sections 2.1 to 2.3 explain the framework of EDA and Ensemble Kalman Filters (EnKF), while section 2.5 describes the application of an Ensemble Kalman Filter to earthquake sequences. Section 3 demonstrates that—and explains how—the underlying fault state is estimated at a single point in time (Section 3.1). The reasonable estimation of stress and strength, which govern earthquake propagation, then allow us to fit the pattern of analogue seismic cycles surprisingly well (Section 3.2, as quantified in Appendix A). The resulting potential for earthquake forecasting for this test case is demonstrated in Section 3.3. We then discuss the limitations, implications, difficulties and opportunities for future applications to nature for both fault stress estimation and other solid earth problems (Section 4). These sections contain extensive explanations of the theory, application and interpretation of EDA to provide a basis for understanding for earthquake seismologists. They are, however, also intended as an introduction for other dynamic solid earth scientists, since they face similar challenges and could exploit benefits in a similar fashion. At the same time, data

assimilation specialists are introduced to the study of earthquake sequences.

2 METHODS

This section summarizes EDA methods and focuses on EnKFs. The novice reader is encouraged to read the accompanying Appendix C, when referenced, or to read the unified overview in the online supplements. For more in-depth reading we suggest a book (Evensen 2009b), a review (Carrassi *et al.* 2018) or an implementation summary (Evensen 2003).

2.1 Ensemble data assimilation

Data assimilation methods aim at deriving probabilistic estimates for the temporal evolution of unobserved physical variables (i.e. state and parameters) and their uncertainties. Their goal is to capture the evolving truth by combining information from both observations and a physical model, while taking errors for both information sources into account (see Fig. 1). Data assimilation alternates between two steps. In the first step, the so-called propagation or forward or forecast step, the previous estimate and its uncertainty are propagated forward in time by the physical model to provide the forecast of the current state and its uncertainty. In the second step, the so-called analysis or update or assimilation step, the new observations are used to correct the forecast to obtain a new best estimate. This analysis becomes the input for the next propagation step (Fig. 1a). This procedure ensures that the data is informed by the prior knowledge of the physical equations on how to extrapolate in space and time, while the evolution of the physical state-space model is corrected by the data.

To describe these two steps in more detail, we first introduce some notation in which vectors are denoted in lower case bold and matrices are denoted as upper case bold. State variables are discretized on a 2-D numerical grid with spatial coordinates χ and ψ . They include all physical variable types r needed to calculate the future evolution of the system. For clarity, state vectors of form $x_{\chi, \psi, r}$ and length S are simply denoted as \mathbf{x} and observation vectors of length O are denoted as \mathbf{y} . A time index is omitted as well, since the sense of time within the single assimilation step is clarified through the use of superscript f for forecast or first guess (i.e. coming from previous time $t - 1$ to current time t) and a for analysis (i.e. estimate including all information up to t).

We use $p^f(\mathbf{x})$ to denote the forecast PDF (or density) and $\bar{\mathbf{x}}^f$ to denote its mean (expected value), the best forecast value. Similarly we use $p^a(\mathbf{x})$ for the analysis PDF and $\bar{\mathbf{x}}^a$ for its mean, the best estimate after the new observations have been taken into account. The true state of \mathbf{x} is unattainable and differs from $\bar{\mathbf{x}}^f$ and $\bar{\mathbf{x}}^a$ because of model errors $\boldsymbol{\varepsilon}_x$ due to, for example, propagation and discretization errors, measurement errors $\boldsymbol{\varepsilon}_y$ and unobservable components. We thus write

$$\mathbf{x} = \bar{\mathbf{x}}^f + \boldsymbol{\varepsilon}_x^f = \bar{\mathbf{x}}^a + \boldsymbol{\varepsilon}_x^a. \quad (1)$$

The forecast errors $\boldsymbol{\varepsilon}_x^f$ and assimilation errors $\boldsymbol{\varepsilon}_x^a$ have mean zero and their covariances matrices $\mathbf{C}_{xx}^f = \overline{\boldsymbol{\varepsilon}_x^f (\boldsymbol{\varepsilon}_x^f)^T}$ and $\mathbf{C}_{xx}^a = \overline{\boldsymbol{\varepsilon}_x^a (\boldsymbol{\varepsilon}_x^a)^T}$ describe the uncertainty of the forecast and the analysis, respectively.

In the **propagation step**, the analysis distribution at the previous time ($p^a(\mathbf{x})$) is propagated forward in time according to the dynamics of the physical model to give the forecast distribution at the current time ($p^f(\mathbf{x})$). The implementation of the propagation

step depends on the chosen assimilation method. In case of a linear dynamics, the propagation step can be done analytically by the Kalman Filter (KF Kalman 1960; Appendix C). The best forecast estimate $\bar{\mathbf{x}}^f$ is obtained by propagating $\bar{\mathbf{x}}^a$ from the previous time step. Similarly the forecast covariance \mathbf{C}_{xx}^f can be computed from the analysis covariance one step before. For nonlinear dynamics one can use linearization, leading to the Extended Kalman Filter (EKF; Jazwinski 1970).

When the dynamics of the physical processes is very nonlinear, ensemble methods that use Monte Carlo techniques are preferred. These methods construct ensembles of size N and approximate $p^f(\mathbf{x})$ and $p^a(\mathbf{x})$ by a sum of N Dirac functions (Fig. 1). The forecast ensemble at the current time is obtained by propagating each member of the analysis ensemble at the previous time according to the dynamics of the physical model. The main advantage of ensemble methods is that the physical model propagation of the ensemble correctly forward integrates the error statistics and the spread of the forecasted ensemble around the best estimate. No approximation of the dynamics is necessary.

The **update step** involves the assimilation of observations through using Bayes' theorem (Bayes & Price 1763)

$$p(x|y) \propto p(x)p(y|x). \quad (2)$$

This theorem describes how prior knowledge about a variable x in the form of a prior PDF $p(x)$ is modified by information from an uncertain measurement y . Here $p(y|x)$ describes the distribution of the measurement when the value of x is known, and $p(x|y)$ contains the posterior knowledge about x when the value y has been observed. In data assimilation, the prior is equal to the forecast PDF, which contains our prior knowledge based on the physical evolution of the system and all previous observations. The likelihood $p(y|x)$ follows from assumptions on the measurement error. The posterior is the analysis PDF which is further propagated in time by the physical model until the next data becomes available.

The implementation of the update step typically assumes the observation vector \mathbf{y} to be of the form

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \boldsymbol{\varepsilon}_y, \quad (3)$$

where $\boldsymbol{\varepsilon}_y$ is a vector of measurement errors and the measurement matrix \mathbf{M} locates the measurement location on the numerical grid. \mathbf{M} is of size $S \times S$ and contains zeros everywhere, but for a few ones on the locations where observations are available. If both the model and measurement errors are Gaussian, then the analysis PDF is Gaussian and the analysis mean and covariances can be computed analytically. This leads to the update step of the regular Kalman filter (Section 2.2).

There are two ensemble methods which only differ in how the update step is implemented. The EnKF (Evensen 1994) uses an ensemble version of the Kalman filter update (Section 2.3). It directly corrects the state variables based on the data misfit or innovation of each state variable. The particle filter (e.g. Gordon *et al.* 1993; van Leeuwen *et al.* 2018) does not update state variables directly. It instead reweights each ensemble member based on the data misfit of each ensemble member. This is typically followed by a resampling step to achieve an equally weighted ensemble.

In this paper we apply an EnKF, since it is the current method of choice for large-dimensional nonlinear problems (Evensen 2009b; Fournier *et al.* 2013). Its computational requirements remain affordable for high-dimensional systems, while it is conceptually simpler and easier to implement. Moreover, its performance in practical problems is remarkable. However, a particle filter (or a combination of both, as in e.g. Frei & Künsch 2013) could turn out to be more

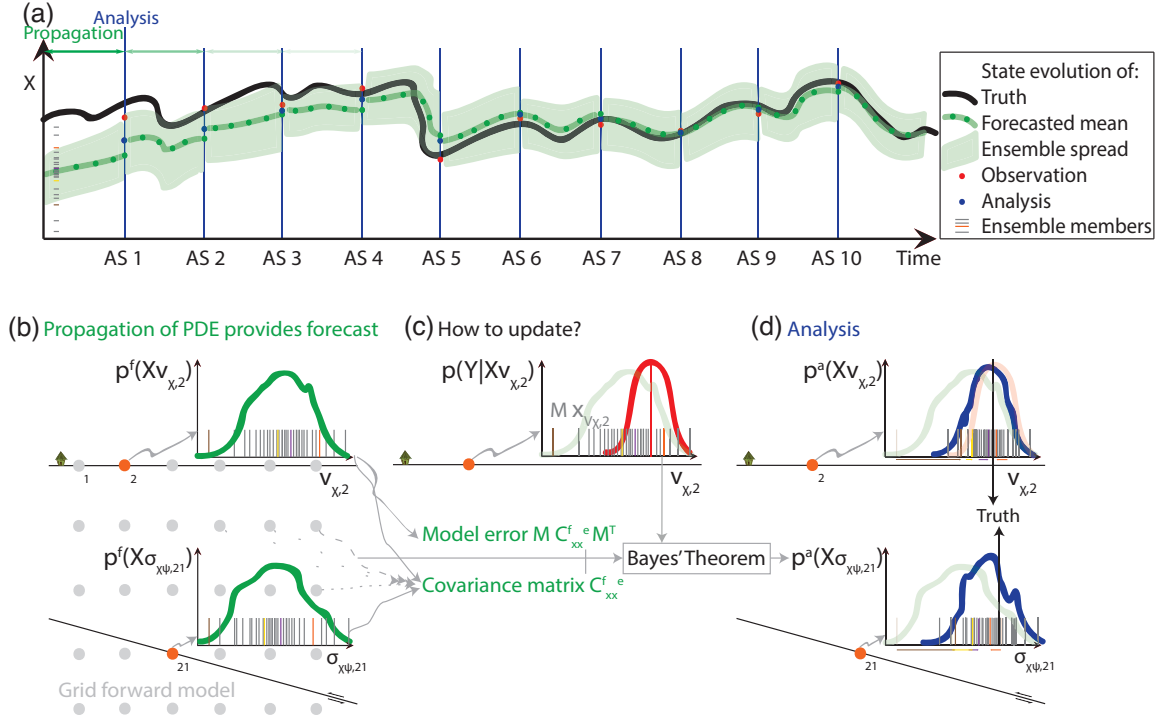


Figure 1. Schematic diagram explaining the concept of Ensemble Kalman Filters. (a) Temporal evolution of a true (black), forecasted (green), analysed (blue) and observed (red) state variable with ensemble data assimilation at each Assimilation Step (AS). The goal is to bring the forecast, propagated based on a physical model, in line with the truth. The concept of the (b) propagation and (d) update steps is illustrated for many ensemble members forming a probability density function of the (b) prior $p^f(x)$, (c) data likelihood $p(y|x)$ and (d) analysis $p^a(x|y)$ for an observed velocity v_x at the surface (at grey grid node 2) and a hidden stress $\sigma_{x,\psi}$ at the fault (at grid node 21). In (d) note that each ensemble member is moved by the coloured arrows to form a posterior, which is the multiplication of the transparent prior and data likelihood. This possibly non-Gaussian posterior ideally approximates the truth (black bar).

appropriate, if underlying assumptions of an EnKF are principally violated. That is if either the measurement error or forecasted prior turns out to be strongly non-Gaussian (e.g. using bi-modal errors as illustrated in Werner *et al.* 2011), or the measurement relation between the observable and the unknown state is too nonlinear.

2.2 Kalman filter

To understand an EnKF, the derivation of the equations, assumptions and challenges for a regular Kalman Filter are explained in Appendix C. In summary, we additionally assume model and measurement errors have a Gaussian distribution with zero mean. Along with a Gaussian initial distribution, this leads to a prior or forecast PDF that is Gaussian with mean $\bar{\mathbf{x}}^f$ and covariance \mathbf{C}_{xx}^f . This model error covariance matrix contains how each component of the state vector covaries with each other component in the state vector. The measurement PDF $p(\mathbf{x}|\mathbf{y})$ is Gaussian with mean $\mathbf{M}\mathbf{x}$ and covariance \mathbf{C}_{yy} . The measurement error covariance matrix \mathbf{C}_{yy} is defined as $\mathbf{e}_y \mathbf{e}_y^T$.

By multiplying the prior and measurement Gaussian densities as in Bayes' theorem and then finding the maximum likelihood estimate $\bar{\mathbf{x}}^a$ (Appendix C and D) one obtains

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}(\mathbf{y} - \mathbf{M}\bar{\mathbf{x}}^f), \quad (4)$$

where matrix \mathbf{K} is usually called the Kalman gain

$$\mathbf{K} = \mathbf{C}_{xx}^f \mathbf{M}^T (\mathbf{M} \mathbf{C}_{xx}^f \mathbf{M}^T + \mathbf{C}_{yy})^{-1}. \quad (5)$$

The Kalman gain is also used to calculate the analysis error covariance matrix \mathbf{C}_{xx}^a (eq. C7).

These equations show how to update the forecasted mean of each state variable based on the data misfit or innovation $\mathbf{y} - \mathbf{M}\bar{\mathbf{x}}^f$ for each measurement. The Kalman gain is a matrix weighting two terms, which together determine the size of the correction by the data misfit. The first term is a transpose of the influence functions $\mathbf{M} \mathbf{C}_{xx}^f$ (size $S \times S$), which quantifies how each state variable relates to the observation. The second term contains the Kalman weights (each of size $O \times O$), or inverted error covariances, which weigh the update for each measurement as a function of the level of confidence in both the model and the measurement. This thus means that update is small if (i) the model forecast at the measurement location lies close to the data, and/or (ii) the Kalman weights are low. The inverted covariance weights are relatively low when either (i) the data error for that measurement is large (i.e. it is better to ignore a deviation of this measurement) and/or (ii) the model spread (or error/variance) for a certain state variable–measurement combination is large (i.e. the model does not know how the state and observation are related and thus how to update the state). However, if the dynamics of the physical processes is very nonlinear, it becomes very difficult to forward propagate the entries into the forecasted model error covariance matrix \mathbf{C}_{xx}^f in particular.

2.3 Ensemble Kalman filter

Forward propagation of nonlinear dynamics is tackled in an EnKF through representing the forecast and analysis distributions by ensembles (Evensen 1994; Burgers *et al.* 1998). In the propagation step, every member (or particle or evolving Markov Chain targeting $p^a(\mathbf{x})$) of the previous analysis ensemble is forward propagated according to the nonlinear partial differential equations (Figs 1a

and \mathbf{b} and Section 2.4) to form the forecast ensemble \mathbf{X}^f (or \mathbf{x}_n^f with $n = 1, \dots, N$). The state variables from every ensemble member are contained in a designated column of the forecast matrix \mathbf{X}^f of size $S \times N$. The forecast ensemble can then be used to approximate the error covariance matrix $\mathbf{C}_{xx}^{f,e}$. This is done for each entry (p, q) by averaging the deviations of all forecasted ensemble members from the mean of the ensemble $\bar{\mathbf{x}}^f$ for each state variable as

$$\begin{aligned} \mathbf{C}_{xx}^{f,e} = (\mathbf{c}_{xx}^{f,e})_{pq} &= \frac{1}{N-1} \sum_{n=1}^N [(x^f)_p - \overline{(x^f)_p}] [(x^f)_q - \overline{(x^f)_q}] \\ &= [\mathbf{X}^f - \bar{\mathbf{X}}^f][\mathbf{X}^f - \bar{\mathbf{X}}^f]^T. \end{aligned} \quad (6)$$

We use a superscript e and write $\mathbf{C}_{xx}^{f,e}$ to emphasize that the error covariance is based on the ensemble. The ensemble update will also correctly propagate the uncertainty forward in time, such that analysed error covariance matrix can also be obtained through sampling the ensemble.

Correct propagation of the error statistics, however, also requires the assimilation of an ensemble of perturbed observations (Burgers *et al.* 1998). For this we generate N artificial measurement errors according to the prescribed Gaussian measurement error distribution $\mathcal{N}(0, \mathbf{C}_{yy})$ and add them in the columns of \mathbf{E}_y and to the actual observation \mathbf{y} , such that

$$\mathbf{Y} = \mathbf{y} + \mathbf{E}_y. \quad (7)$$

For computational efficiency the ensemble measurement error covariance can then be sampled as (Evensen 2003)

$$\mathbf{C}_{yy}^e = \frac{1}{N-1} \sum_{n=1}^N \mathbf{e}_{y,n} \mathbf{e}_{y,n}^T = \overline{\mathbf{E}_y \mathbf{E}_y^T}. \quad (8)$$

These ensemble estimates in terms of the average of the Dirac measures at \mathbf{x}_n^a will converge to the posterior p^a (eqs C3 and C5) as the ensemble size goes to infinity.

Using these ensemble-based approximations, the EnKF calculates the posterior density or analysis ensemble as

$$\mathbf{X}^a = \mathbf{X}^f + \mathbf{K}^e(\mathbf{Y} - \mathbf{M}\mathbf{X}^f), \quad (9)$$

where each column of the perturbed observation matrix \mathbf{Y} (size $O \times N$) and the posterior matrix \mathbf{X}^a (size $S \times N$) again contains one ensemble member. The superscript e in \mathbf{K}^e indicates that the Kalman Gain of size $S \times O$ is calculated using ensemble-based approximations to the error covariances as

$$\mathbf{K}^e = \mathbf{C}_{xx}^{f,e} \mathbf{M}^T (\mathbf{M} \mathbf{C}_{xx}^{f,e} \mathbf{M}^T + \mathbf{C}_{yy}^e)^{-1}. \quad (10)$$

The update eq. (9) means that each forecast member is updated in the same way as the mean is updated in the Kalman filter (eq. C6), but with its own, perturbed observation. In particular, the discussion on how the update equation (eq. 9) depends on the data misfit also applies for the EnKF. The interpretation of these equations in light of our earthquake sequence model is further discussed in the results (Section 3.1).

Finally, the state estimates analysed to reflect the observed data are inserted back into each forward model for the next propagation step. This inclusion is the only adaptation needed within the core of the forward model code to link it to the EDA framework.

In summary, the relaxation of Gaussian assumptions on the prior and model error leads to a solution that lies between a linear Gaussian update and a full Bayesian computation (Figs 1b and d). To

apply these equations, we only need to know (i) the evolution equations from the physical model to forecast \mathbf{X}^f and \mathbf{C}_{xx}^f , and (ii) the measurement values \mathbf{y} and their error ε_y and location for M .

2.4 Continuum modelling of seismicity

The continuum seismo(-thermo)-mechanical modelling framework reproduces cycles of spontaneous analogue earthquakes without predefining fault planes. It is based on a 2-D, coupled thermomechanical code developed for long-term, geodynamic deformation (Gerya & Yuen 2007), which is extended and validated to simulate analogue earthquakes (van Dinther *et al.* 2013b). These references contain a description of the forward model, which is summarized here to understand the data assimilation scheme proposed.

We use a fully staggered Eulerian, conservative finite difference scheme, in combination with a Lagrangian marker-in-cell technique, to implicitly solve Navier–Stokes equations for the conservation of mass and momentum in an incompressible medium

$$\frac{\partial v_\chi}{\partial \chi} + \frac{\partial v_\psi}{\partial \psi} = 0, \quad (11)$$

$$\frac{\partial \sigma'_{\chi\chi}}{\partial \chi} + \frac{\partial \sigma'_{\chi\psi}}{\partial \psi} - \frac{\partial P}{\partial \chi} = \rho \frac{Dv_\psi}{Dt}, \quad (12)$$

$$\frac{\partial \sigma'_{\psi\chi}}{\partial \chi} + \frac{\partial \sigma'_{\psi\psi}}{\partial \psi} - \frac{\partial P}{\partial \psi} = \rho \frac{Dv_\chi}{Dt} - \rho g. \quad (13)$$

This provides us with horizontal and vertical velocities, v_χ and v_ψ , and pressure P (defined as mean normal stress). The last two equations of motion are written in terms of deviatoric stress tensor components (nodal cell normal $\sigma'_{\chi\chi}$ and nodal cell shear $\sigma'_{\chi\psi}$) and include both gravity acceleration g ($=9.81 \text{ m s}^{-2}$) and inertia (density ρ times the Lagrangian time derivative of the respective velocity components $\frac{Dv}{Dt}$).

These equations are solved using constitutive relations that relate deviatoric stresses and strain rates $\dot{\epsilon}_{ij}$ in a nonlinear visco-elasto-plastic manner as

$$\dot{\epsilon}_{ij} = \frac{1}{2\eta_v} \sigma'_{ij} + \frac{1}{2G} \frac{D\sigma'_{ij}}{Dt} + \begin{cases} 0 & \text{for } \sigma'_{II} < \sigma_{\text{yield}} \\ \lambda \frac{\sigma'_{ij}}{2\sigma'_{II}} & \text{for } \sigma'_{II} = \sigma_{\text{yield}}, \end{cases} \quad (14)$$

where i and j are coordinate indices (χ and ψ each), η_v is viscosity, G is shear modulus, D/Dt is the objective corotational time derivative, λ is the plastic multiplier connecting plastic strain rates and stresses and σ'_{II} is the second invariant of the deviatoric stress tensor

$$\sigma'_{II} = \sqrt{\sigma_{\chi\chi}^2 + \sigma_{\chi\psi}^2}. \quad (15)$$

The elastic update of both stress components from the previous time step σ_{ij}^0 is solved using a forward Euler, explicit time integration scheme (Schmalholz *et al.* 2001; Moresi *et al.* 2003; Gerya 2010). Through introducing an effective viscosity corrected for plasticity η_{vp} , eq. (14) can be rewritten as

$$\sigma'_{ij} = 2\eta_{vp} \dot{\epsilon}_{ij} Z + \sigma_{ij}^0 (1 - Z), \quad (16)$$

$$Z = \frac{G \Delta t}{G \Delta t + \eta_{vp}}. \quad (17)$$

Here the viscoelasticity factor Z weighs the contribution of viscous with respect to elastic stresses, as determined by the local shear modulus, viscoplastic viscosity and time step Δt . The contribution

of plasticity comes in through adapting the ductile viscosity to ensure that stresses do not exceed the yield strength, defined in the non-associated Drucker–Prager plastic yielding model (Drucker & Prager 1952). Brittle/plastic faulting occurs at a Lagrangian marker, when the second invariant of the deviatoric stress tensor is equal to (or larger than) a pressure-dependent yield strength σ_{yield}

$$\sigma_{\text{yield}} = C + \mu(V)P, \quad (18)$$

where C is cohesion and $\mu(V)$ is the slip-rate-dependent friction coefficient.

Frictional instabilities, and subsequent healing, are introduced through an invariant formulation of strongly slip-rate-dependent friction formulation (van Dinther *et al.* 2013b following e.g. Burridge & Knopoff 1967 and Cochard & Madariaga 1994) as

$$\mu(V) = \mu_s(1 - \gamma) + \mu_d \frac{\gamma}{1 + \frac{V}{V_c}}. \quad (19)$$

Here γ denotes the amount of rate-induced weakening that is equivalent to $1 - \mu_d/\mu_s$, where μ_s and μ_d are the static and minimum dynamic friction coefficient. The viscoplastic slip rate V is calculated as half a viscoplastic strain rate $\sigma_{\text{yield}}/\eta$ times the grid size $d\chi$, where strain rate is a function of both velocity components through $\dot{\epsilon}_{ij} = \frac{1}{2}(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i})$. This invariant rate-dependence allows for spontaneous fault plane development, as a drop in effective viscosity localizes deformation and ensures local stress–strain rate equilibrium during plastic yielding.

In summary, a synthetic, analogue earthquake occurs spontaneously, when a large enough local stress drop is able to overcome neighbouring stresses and induce plastic yielding there as well. Since local plastic yielding occurs when

$$\sqrt{\sigma_{\chi\chi}^2 + \sigma_{\psi\psi}^2} \geq C + \mu(v_\chi, v_\psi)P, \quad (20)$$

we need to capture both stress and velocity components and pressure along the fault to forecast earthquake occurrence and the corresponding behaviour of the wedge or medium.

2.5 Setup perfect model experiment

We perform a perfect model experiment to test whether EDA could be used to estimate fault stresses and resulting earthquake sequences. The crucial assumption is that the modelled physics perfectly represents the evolution of the system, that is, there is no representativeness error. The numerical model that produces the forecasts is thus run once more to provide synthetic observations, onto which Gaussian noise is added. This synthetic data model represents the true solution, which provides a means to quantify how well hidden state variables are estimated by our EnKF. This approach of a synthetic twin experiment is the best way to test initial feasibility and is a classical requirement tested extensively in geophysical data assimilation applications (e.g. Evensen 2009b).

The script in which the EnKF was implemented (freely available on ETH's Research Data Repository) is checked for consistency of the error statistics (appendix A in Evensen 2003). In our existing mechanical code, only two things are changed; (i) state variables are input, interpolated to markers and output when needed, and (ii) posterior state variables replace prior ones before the plasticity calculation and Navier–Stokes solve (at the start of step 2 in fig. 2 in Gerya & Yuen 2007). The new state temporarily violates the mechanical conservation of mass and momentum, but this does not lead to computational issues and conservation is restored in the next

time step. Communication between the assimilation and forward codes, and coordination for nearly embarrassingly parallel ensemble runs, was facilitated by wrappers written using MPI (freely available as well).

The model setup resembles an analogue or scaled laboratory experiment (Corbi *et al.* 2013; Fig. 2a) to ensure a simple and computationally fast setup in which many events occur within a feasible observational time span. The numerical setup adopts the laboratory settings, parameters and dimensions and captures all observed statistics of this analogue seismic cycle (van Dinther *et al.* 2013b). This rotated setup mimicks a subduction zone in which a rigid oceanic slab subducts landward beneath a viscoelastic upper plate forearc. Within the frictional interface a velocity-weakening frictional patch designates the seismogenic zone in which earthquakes occur. This zone experiences cycles of long interseismic coupling punctuated by rapid coseismic slip and post-seismic recovery (see fig. 7 in van Dinther *et al.* 2013b for the evolution of variables throughout the seismic cycle). This slip simulates an analogue earthquake, which elastically rebounds the landward displacements build up during the interseismic period of locking in between two events (arrows in Fig. 2b). During this interseismic period, the stress state builds up along the interface, and particularly near the downdip limit of the seismogenic zone (colour and contour line, resp., in Fig. 2b). There differential displacement is the largest and stresses thus typically exceed the yield strength first. When that happens over a large enough area, an event nucleates, propagates and consequently releases a distinct part of these stresses during this coseismic event. As in nature, fault stress and strength thus determine the state throughout the entire forearc. Hence, the challenge is to forecast these correctly.

In order to (i) estimate fault and forearc state, (ii) forecast events and (iii) propagate a complete and balanced forward model, we need to estimate all nodal values for five physical variables (i.e. v_χ , v_ψ , $\sigma'_{\chi\chi}$, $\sigma'_{\psi\psi}$ and P). In addition to the three solution variables, stresses are included to incorporate the history component. They can namely not be deduced from current velocities, but are necessary to calculate the next time step. These $(701 \times 136) \times 5 = 476\,680$ state variables are first scaled with a characteristic value to have more similar variability within error covariance matrices that need to be inverted (Table 1). This ensures that the eigenvalues corresponding to each measurement type do not differ by many orders of magnitude (Evensen 2003). These state variables are added consecutively into the rows of forecast state vector \mathbf{x}^f , which in the ensemble formulation is extended to a forecast matrix \mathbf{X}^f with the entries for the N ensemble members in each column. Larger ensembles typically allow for improved forecasts, so we show results for the largest number tested: 150 ensemble members. However, as little as 20 ensemble members are found to already capture most of the diversity within the models (Appendix B). Subsequently, data obtained from a single point in space are assimilated into these forward models at intervals of 30 time steps, which are fixed for simplicity (or ~ 2 s in this 800 s laboratory run). This point can be interpreted to represent a roughly 5 km deep borehole (the yellow circle in Fig. 2a). To obtain error estimates for each measurement type, we use current state-of-the-art values from the literature (Valley & Evans 2007; El-Mowafy & Bilbas 2016). These errors are then downscaled to our laboratory model setup using the analogue scaling relation developed in Corbi *et al.* (2013; Table 2). These values show that velocities can be obtained fairly accurately using GPS measurements. Stress and pressure estimates within boreholes, however, are very challenging and thus have large and not well known errors (e.g.

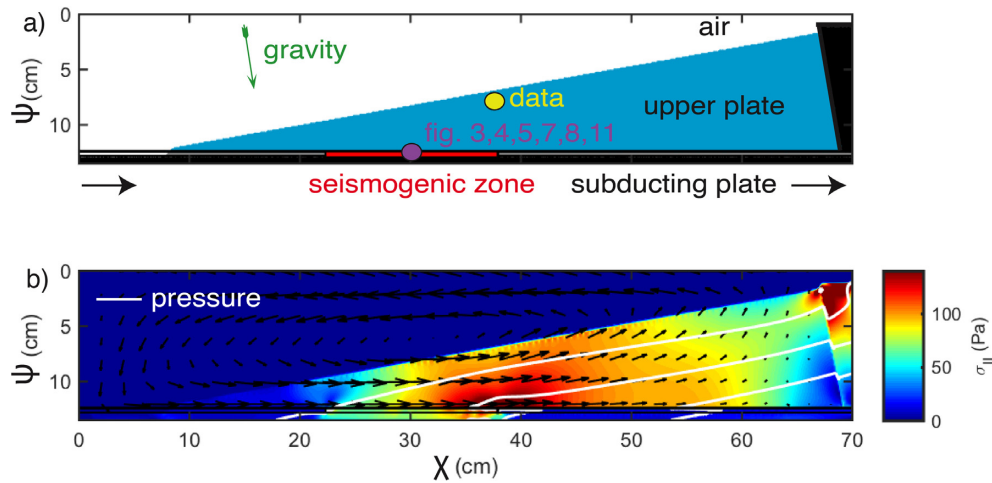


Figure 2. (a) Numerical model setup of an analogue subduction zone adopted from van Dinther *et al.* (2013b). Colours represent different compositional units and dots represent relevant locations. The green arrow is a gravity body force tilted with a 10 degree slab dip, while the black arrow represents a driving boundary condition for subduction velocity. (b) Spatial overview of all state variables (arrows for velocity, colour for second invariant of deviatoric stress and pressure contours) at the first assimilation step in the synthetic data model. Note that velocities are present in the air due to a sticky-air approximation, which typically ensures a free surface in geodynamic models (Cramer *et al.* 2011).

Table 1. Scaling values. Columns 2 and 3 provide characteristic (char.) values used during assimilation for scaling to similar eigenvalues for each data type r . Columns 4–6 give scaled error variances of resp. the model (at assimilation step AS), the data and their combined standard deviation used to evaluate and visualize output.

Data type r	Reasoning for characteristic value	Characteristic value	$(C_{xx}^{f,e})_{rr,AS=1}$	$(C_{yy}^{f,e})_{rr}$	$\sqrt{(C_{xx}^{f,e})_{rr,AS=1} + (C_{yy}^{f,e})_{rr}}$
v_x	Subduction velocity	$39.0 \times 10^{-6} \text{ m s}^{-1}$	1.23	$3.26\text{e-}4$	1.11
v_ψ	Average in wedge	$2.25 \times 10^{-6} \text{ m s}^{-1}$	219	1.31	14.8
σ'_{xx}	Average $\sigma'_{xx\psi}$ build in SeZ	33.2 Pa	$38.1\text{e-}3$	$7.20\text{e-}3$	0.213
$\sigma'_{x\psi}$	Average $\sigma'_{x\psi}$ build in SeZ	33.2 Pa	$7.70\text{e-}3$	$7.20\text{e-}3$	0.122
P	Lithostatic P at downdip SeZ	547 Pa	$1.61\text{e-}4$	$6.70\text{e-}6$	0.0129

Table 2. Data error ε_y values and sources based on current state-of-the-art measurements. Downscaling to laboratory setup values follows scaling relations developed by Corbi *et al.* (2013). Stress terms refer to our numerical grid and the error in each component—and pressure or mean stress—is kept the same for simplicity (i.e. is assumed to be independent of orientation).

Data type	Error (nature)	Error (lab)	Equipment	Location	Reference
v_x	1.5 mm yr^{-1}	$0.67 \times 10^{-6} \text{ m s}^{-1}$	GNSS (GPS)	Western Australia	El-Mowafy & Bilbas (2016)
v_ψ	5.5 mm yr^{-1}	$2.5 \times 10^{-6} \text{ m s}^{-1}$	GNSS (GPS)	Western Australia	El-Mowafy & Bilbas (2016)
σ'_{xx}	5 MPa	2.7 Pa	Wellbore failure and hydraulics	Soultz-sous-Forêts	Valley & Evans (2007)
$\sigma'_{x\psi}$	5 MPa	2.7 Pa	Wellbore failure and hydraulics	Soultz-sous-Forêts	Valley & Evans (2007)
P	5/2 MPa	2.7/2 Pa	Wellbore failure and hydraulics	Soultz-sous-Forêts	Valley & Evans (2007)

Huffman & Saffer 2016). In Appendix B, we analyse the impact of this choice to use these state-of-the-art errors, as they are. We observe that distinctly smaller errors no longer reduce least absolute errors. Finally, the synthetic data are noised using this Gaussian error before providing it to the assimilation algorithm to ensure compatibility with the same challenge of noisy observations in nature. Introducing this additional error into the synthetic data distinctly increases the difficulty of recovering the truth.

We introduce a stochastic component into both the synthetic data simulation and all ensemble member simulations through varying initial distributions of the Lagrangian markers. Over time this leads

to a different timing of events due to the inherently sensitive nature of plasticity (see appendix A1 in van Dinther *et al.* 2013b). Such variable event timing permanently changes stress states and thus makes them diverge slowly. Propagating all divergent stresses for a few characteristic cycles ensures that (i) the true stress state is unknown, and (ii) the prior is as uninformed as possible. This means that, at the start of data assimilation at $t \approx 134 \text{ s}$, the truth is considerably distinct from the mean of each prior (compare e.g. Fig. 3a and c; and the green, prior PDF and black, true line in the right column of Fig. 4d). The true solution is namely located around the 79th, 47th, 81st, 67th and 95th percentile for v_x , v_ψ , $\sigma'_{x\psi}$, σ'_{xx} , and P resp.

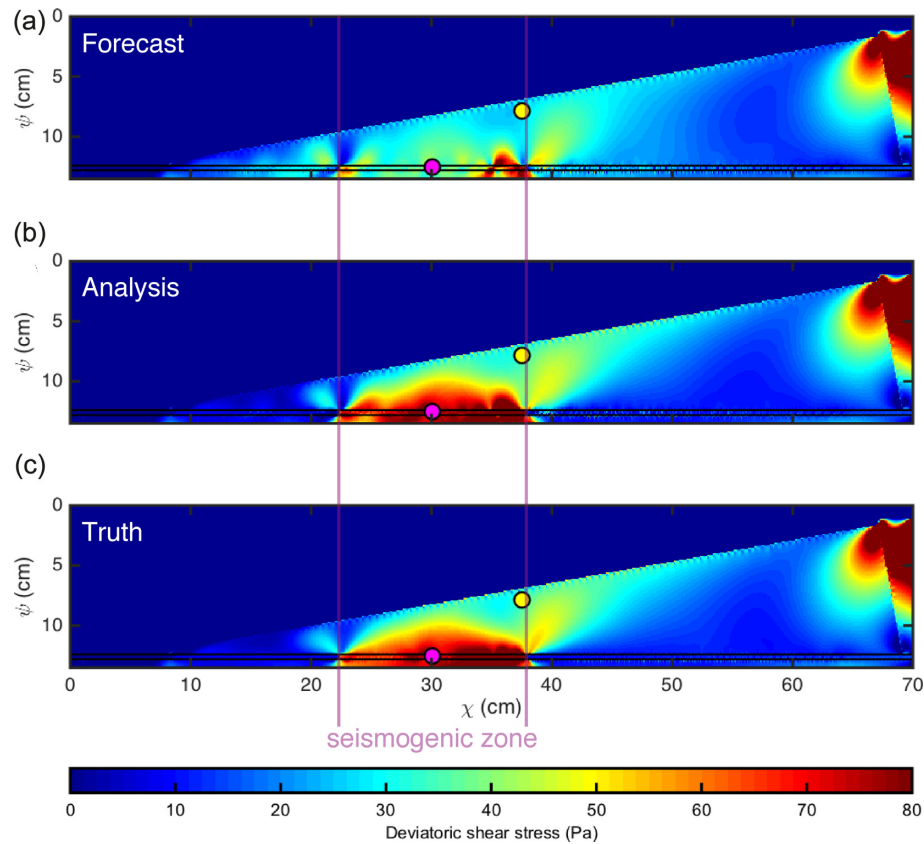


Figure 3. Spatial shear stress recovery visualized for an ensemble member in a postseismic period (a) before and (b) after assimilation of interseismic data from (c) the true target model. This example ensemble member had a distinct velocity misfit at the analogue borehole (yellow circle) at the first assimilation step (red ensemble member in Fig. 4). The magenta circle marks the location where subsequent fault states will be evaluated.

3 RESULTS AND ANALYSIS

Whether EDA can enhance our physical understanding of earthquakes—and ultimately improve their probabilistic forecasting—relies on the possibility to estimate, and forecast, the state of stress and strength at dominant faults.

3.1 Can a fault's state be recovered?

To appreciate what is possible—and why—in a simplified laboratory setup, we start by analysing whether we can capture the unknown stress state from data obtained at a single point in space and time. At the first assimilation step, velocity and stress data from the true synthetic data model are noised and assimilated (see the yellow point in Figs 2a and 3, error values in Table 2 and explanation in Section 2.5). These data are taken from an interseismic period, which intuitively contains little information to constrain the spatial distribution of shear stresses. Velocities indicate slow landward movements, as they do most of the time, and stresses and pressures are increased (Fig. 3c). However, when sampled at the surface, they provide a somewhat distorted picture due to a large measurement error and the large distance to the fault.

To see how much information can be extracted when both modelling and observations are combined through data assimilation, we analyse the shear stress update of one of the ensemble members that is far from the true solution (Fig. 3c). The forecast of that ensemble member shows low stresses throughout most of the

seismogenic zone and forearc (Fig. 3a), which is characteristic for a post-seismic period as a coseismic event just released stresses. After assimilating data obtained at the yellow circle, the analysed shear stress pattern (Fig. 3b) largely resembles that of the targeted true model from which the synthetic data were taken. The EnKF can thus reconstruct the spatial distribution of shear stresses remarkably well.

3.1.1 How to update shear stress in the centre of the seismogenic zone?

To understand why this shear stress pattern can be recovered so well, we consider how shear stress is updated at a node in the centre of the seismogenic zone (magenta point in Figs 2a and 3). To do so, the hidden or unobserved shear stress is plotted on the y-axis, while the observable horizontal velocity is outlined on the x-axis in each subfigure in Fig. 4. Every open circle represents a forecasted ensemble member, which together make up the prior density plotted in green in the top and right frames. The large majority of ensemble members is within the interseismic period and show small positive landward velocities. However, several outliers either indicate post-seismic velocities, whose reversal towards landward velocities is delayed (e.g. the red open circle in Fig. 4), or the nucleation stage of a coseismic event (e.g. green open circle). Fault shear stresses complete a corresponding circular-like pattern with interseismic stresses first increasing slowly with velocity, then reducing rapidly with increasing negative (seaward) coseismic velocities towards a

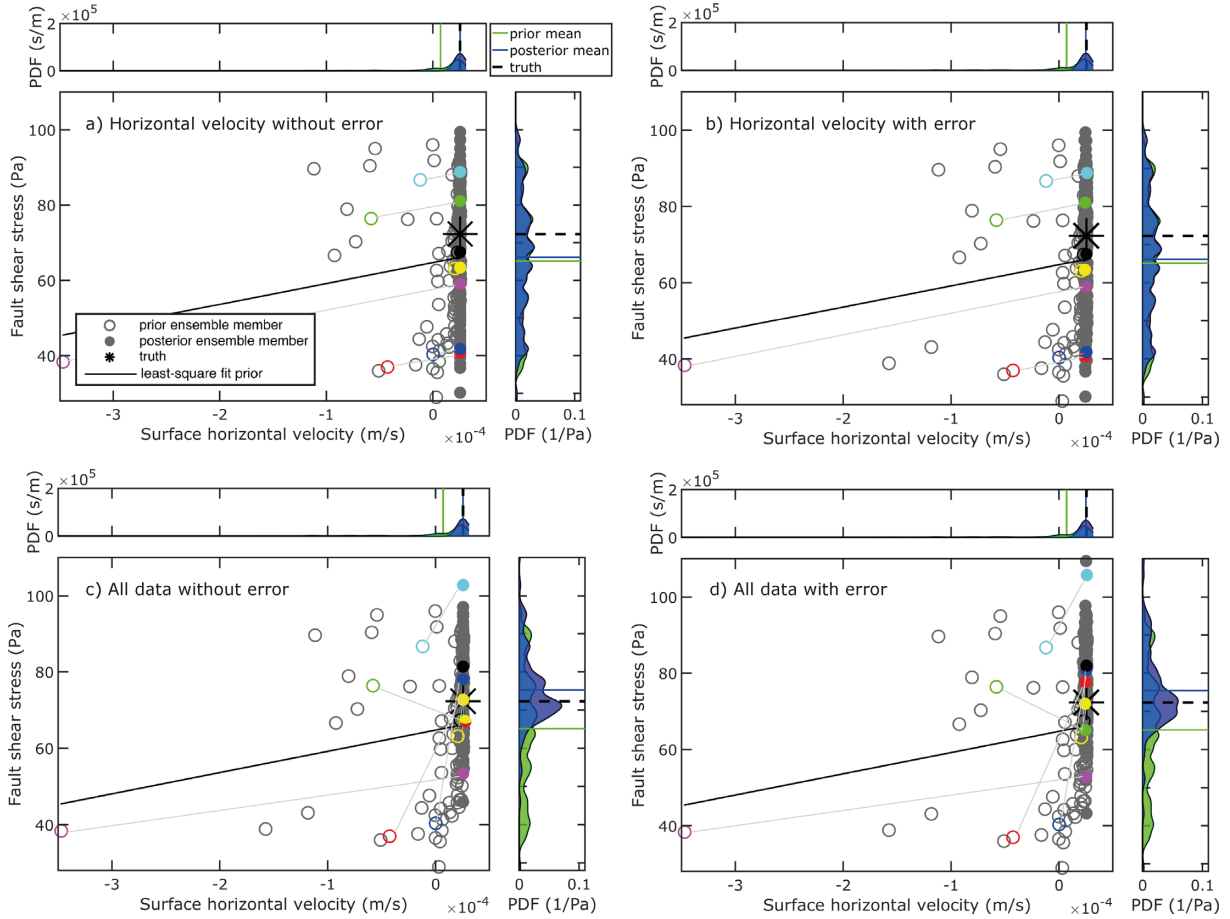


Figure 4. Ensemble results explaining how to update each ensemble member when data is assimilated. Each subfigure shows nodal values of observed horizontal velocity (x) versus unobserved fault shear stress in the centre of the seismogenic zone (y). The main frame contains open (prior) and filled (posterior) circles for each ensemble member (with some examples coloured for identification), a black star for the true target solution and a solid regression line from the linear least-square fit to the prior ensemble members. Subfigures show successively increasing data complexity from (a) one observation of the true horizontal velocity without data error, (b) one noised observation of the horizontal velocity, including corresponding data error, (c) five single location observations of true velocities, stresses and pressure without data error, to (d) five noised observations with data error (reference setting used throughout this paper). The side frames show the corresponding prior (green) and posterior (blue) densities with a dashed line for each mean and the true target line in black.

minimum, and finally a post- and early interseismic rapid recovery. This leads to a wide potential distribution of shear stresses with a maximum near common interseismic values.

An understanding of how the shear stress of each ensemble member should be updated is best provided successively through increasing the complexity of the data assimilated into the model. First, we analyse the assimilation of a single observation of the most readily available horizontal surface velocity without accounting for a data error (Fig. 4a). Since the data is assumed to represent the truth correctly, the observed horizontal velocities of all ensemble members are updated to the observation exactly. The shift along the x -axis from $x_{n,1}^f$ to $x_{n,1}^a$ is thus equal to the data misfit or innovation $y_1 + 0 - Mx_{n,1}^f$. The corresponding update of the shear stress of each ensemble members from $x_{n,2}^f$ to $x_{n,2}^a$ occurs parallel to the slope of the least-square regression (C_{21}/C_{11}) shown for the forecasted prior (follow coloured examples). This can be seen from the fact that the slope of the corresponding update for a single observation (eq. 9) is equal to

$$\frac{x_{n,2}^a - x_{n,2}^f}{x_{n,1}^a - x_{n,1}^f} = \frac{K_2}{K_1} = \frac{(C_{xx}^{f,e})_{21}}{(C_{xx}^{f,e})_{11}}. \quad (21)$$

This is because the data misfit with horizontal velocity is the same for all state variables and thus cancels out in the first rewrite. The scalar Kalman weights are also equivalent for all state variables and thus cancel out in the second rewrite. This update direction emphasizes the least-squares character of the Kalman solution and represents the prior physical knowledge on how the hidden state variable covaries with the observed state variable.

The slope of the update when data errors are included is still parallel to the least-square fit to the prior, because both update terms cancel as well (Fig. 4b). However, the update along the surface horizontal velocity does not reach the noised observation due to (i) assimilation of an ensemble of perturbed observations \mathbf{Y} (eq. 7), and (ii) the data error accounted for in the data error covariance matrix \mathbf{C}_{yy}^e (eq. 9). Assimilation of one interseismic horizontal velocity moved the mean marginally closer to the truth (black star) through increasing the low velocity outliers.

Additionally, also assimilating data on vertical velocity, stresses and pressure significantly improves the mean and maximum likelihood estimate. At the same time the uncertainty is reduced distinctly, thus constraining fault shear stress remarkably well (Fig. 4c). Without a data error (and with uncorrelated data errors) this is achieved

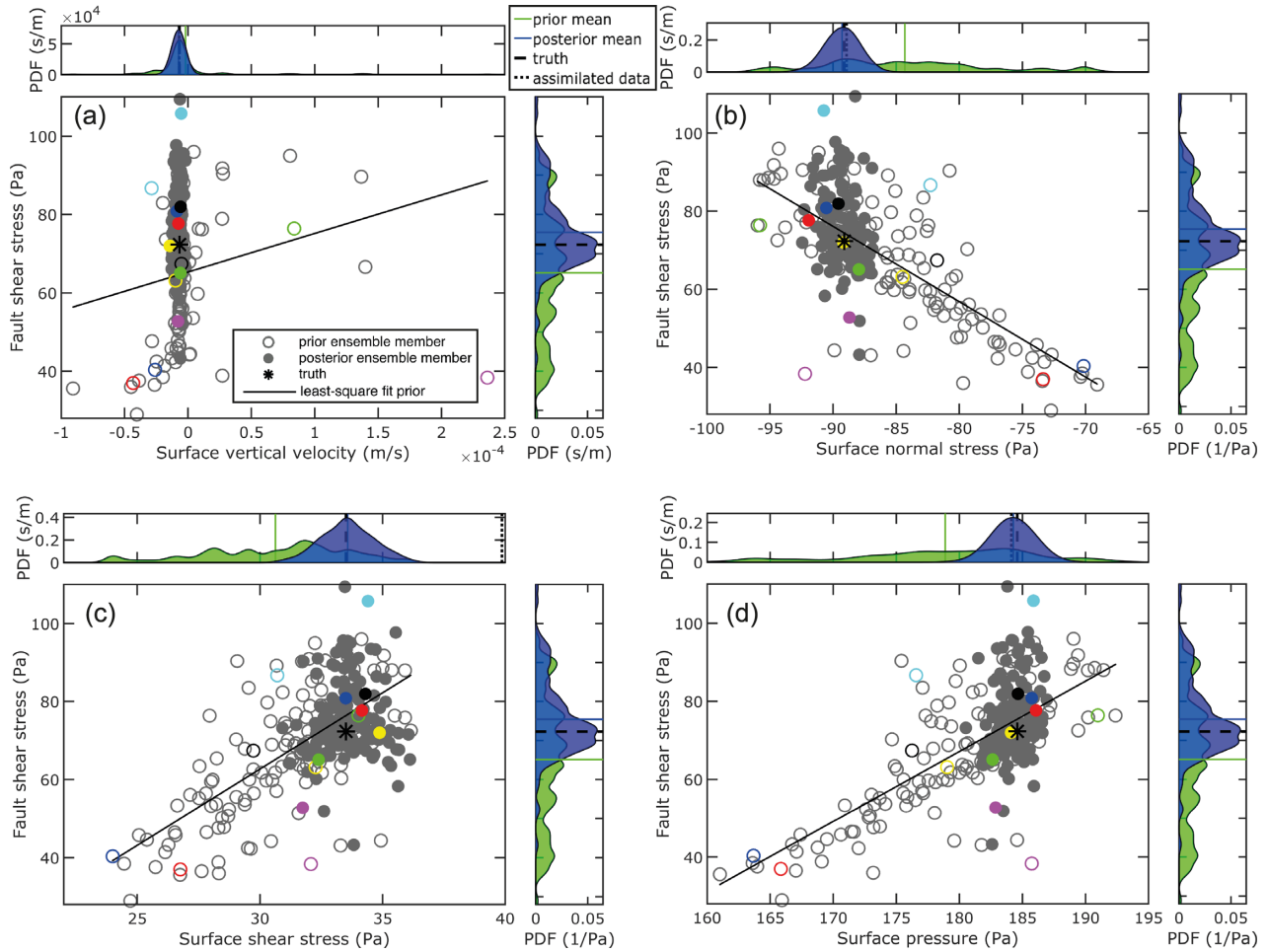


Figure 5. Ensemble panels of the reference models update of fault shear stress shown as a function of each additional surface observation of (a) vertical velocity, (b) normal stress, (c) shear stress and (d) pressure (see the caption of Fig. 4(d) for more explanation).

through updating each ensemble member similarly as described for horizontal velocity, but then sequentially assimilating one observation after the other. So four least-square fits through each prior combination (Fig. 5) are consecutively projected onto the visualized plane to arrive at the posterior locations in Fig. 4(c).

Accounting for a measurement error leads to changes in the update of the observed (as for Figs 4a and b) and unobserved components. The unobserved update also accounts for the relative weights of various data errors, which are present on the respective off-diagonal elements of the inverted data error covariance (which is diagonal but for random noise). These combined changes introduce uncertainties that increase the spread of the posterior. However, despite including large errors on stresses and pressure, the posterior still peaks clearly around the true fault shear stress.

To understand the contribution of the four other observation types, Fig. 5 shows the updates of fault shear stress in their respective observation planes. Stresses and pressure are observed to covary constructively with the estimated fault shear stress. Ensemble members that experience more compression at the surface (i.e. more positive shear stress and pressure and less negative normal stress), also experience more compression at the fault. This prior physical knowledge builds the model covariances that indicate how to update the solution in response to each observed data type. How much each ensemble member is updated is mainly determined by

its data misfit with respect to each observation. If the horizontal velocity misfit is large, then the update will be largely parallel to the regression through horizontal velocity (e.g. for magenta member in Figs 4d and 5a). In case of a large stress or pressure misfit, their respective least-square fit is followed instead (e.g. for red or blue member in Figs 5b–d).

These step-wise explanations of how to update the shear stress hold for all nodal points within the model, although covariances between the data point and other nodal points vary. This can be seen when analysing an additional point near the downdip limit of the seismogenic zone, where stresses reach the strength first and most events thus nucleate (Appendix E). Updating the fault's shear stress there shows that only assimilating horizontal velocity data performs better, such that the addition of stress data is less crucial (compare Figs E1a and b).

In summary, these updates for a single state variable highlight the necessity of a physical link between this unobservable variable and the observable state variables, as stored within the sampled error covariance matrix of the model $\mathbf{C}_{xx}^{f,e}$.

3.1.2 How to update shear stress throughout space?

The constructive covariance for one point in Section 3.1.1 needs to be extended to all nodes to evaluate whether all shear stresses

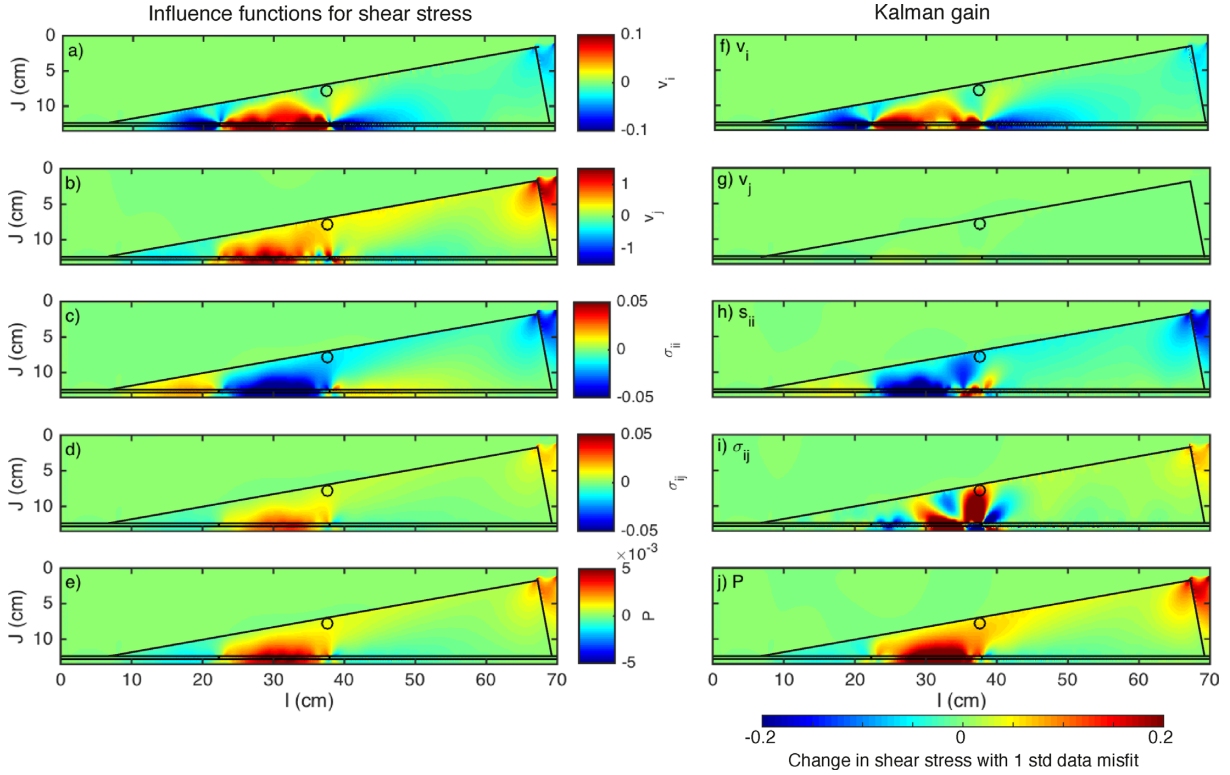


Figure 6. Spatial information available from physical model on how to update shear stresses at every nodal point. The left-hand column shows transposed influence functions ($\mathbf{C}_{xx}^{f,e} \mathbf{M}^T$) illustrating how shear stress covaries with observations of (a) v_x , (b) v_y , (c) σ_{xx} , (d) σ_{xy} and (e) P at the open dot. No scaling is applied with respect to the assimilation (Table 1). The right-hand column shows the part of the Kalman gain \mathbf{K} that illustrates how to update shear stresses throughout the model as a function of a one standard deviation data misfit on innovation or data misfit for observations of (f) v_x , (g) v_y , (h) σ_{xx} , (i) σ_{xy} and (j) P at the open dot. Note that the minor change in patterns between left-hand and right-hand columns is due to the influence of the cross terms involving the off-diagonal components of the model (and data) error covariances. To evaluate the relative contributions of these data types to the shear stress at each node, the update is normalized through scaling (multiplying) \mathbf{K} with the standard deviation of the data misfit $\sqrt{(c_{xx}^{f,e})_{rr,t=1} + (c_{yy}^{f,e})_{rr}}$ (Table 1). This only reduced the visual magnitude of pressure slightly.

can be updated to conform observations. By visualizing the transposed influence functions $\mathbf{M}\mathbf{C}_{xx}^{f,e}$ for each observation, it can be seen that reasonable covariances with shear stresses are also present for other unobserved points throughout the model (Figs 6a–e). These influence functions show that shear stresses near the seismogenic zone should increase, when measured velocities are faster than forecasted velocities (Figs 6a–b). So shear stresses should increase for slower co-, post- and earlier interseismic models, as their stresses have already dropped too much during the previous event from which velocities are still recovering. Similarly, when shear stresses and pressures in proximity of the seismogenic zone are too low with respect to their positively correlated values at the surface, they need to be increased (Figs 6d–e). The same increase to more compression holds for the normal stress components, which are negative in compression (Fig. 6c). Overall, the spatial pattern of how shear stress covaries with each data type is rather smooth. This likely allows for a reasonable shear stress estimate for all points.

To complete the information on how to update shear stress (or any other state variable type), these transposed influence functions need to be weighted by the inverted model plus data error covariance matrices (each 5×5) to form the Kalman gain \mathbf{K} (eq. 9). \mathbf{K} subsequently needs to be multiplied by the misfit for each data entry to get the update. This update is normalized through scaling \mathbf{K}

with the standard deviation of the innovation $\sqrt{(c_{xx}^{f,e})_{rr,t=1} + (c_{yy}^{f,e})_{rr}}$ to be able to compare colours between data types r . Figs 6(f)–(j) thus show how the shear stress on each node should be updated due to a one standard deviation data misfit of each plotted data type r . The changes in colour intensity between the left-hand and right-hand columns result from the weighting with model and data error, which is fixed for all state variable and ensemble member updates (for this assimilation step). This shows that a one standard deviation misfit in vertical velocity hardly induces a change in shear stress (Fig. 6g) due to its large model error $\mathbf{M}\mathbf{C}_{xx}^{f,e} \mathbf{M}^T$ (Table 1). Since there is hardly any prior physical knowledge on how to update shear stress according to vertical velocity (Fig. 5a), this data misfit is largely ignored. The largest contribution from the data error comes from distinctly decreasing the relative contribution of observed stresses and pressure to a level where their influence—accidentally—becomes similar to that of the horizontal velocity. All together, though, the contribution of a data misfit with observed shear stresses remains largest. Finally, the results of each subplot in the right-hand column needs to be multiplied with the scaled innovation to arrive at the full update for shear stress.

In summary, these results demonstrate that shear stresses can be estimated reasonably well even from observations from a single point in time and space.

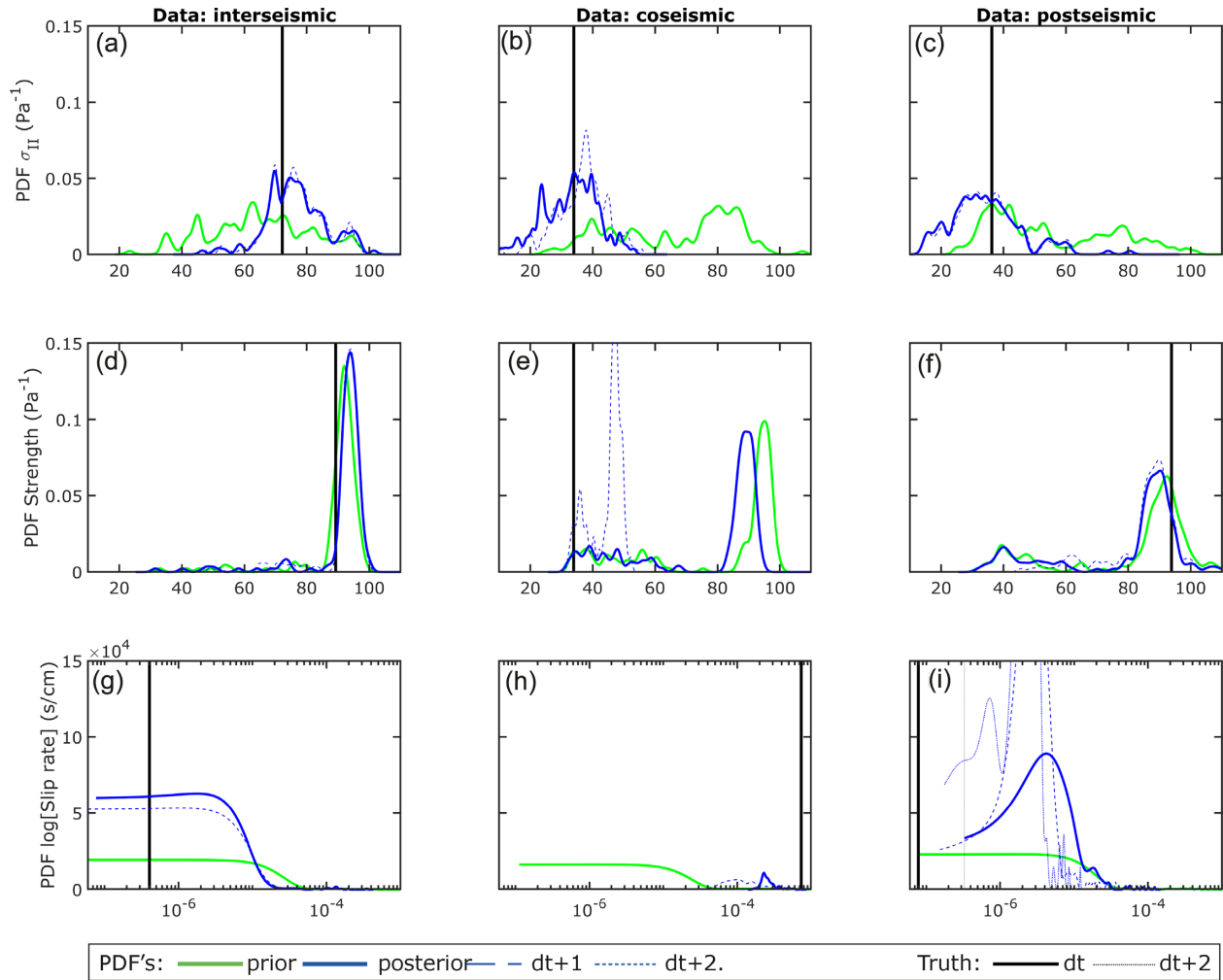


Figure 7. Examples of probability density functions (PDF) of fault stress (a–c), strength (d–f) and slip rate (g–i) at three snapshots in time (columns) chosen at each first occurrence of a specific phase of the seismic cycle within the synthetic data: interseismic (left at first Assimilation Step, AS), coseismic (centre at third AS) and post-seismic (right at 46th AS). PDF's (all with area 1) that are green indicate the prior based on a forecast of the physical model (and previous data assimilated for centre and right), while the blue ones indicate the posterior after data are assimilated. The blue dashed PDF indicates the distribution after a forward modelling time step to illustrate the impact and relevance of the results. The black lines indicate the corresponding truth. Analysed at the same node in the centre of the seismogenic zone.

3.1.3 Can stress and strength be recovered to revert or initiate events?

The successful estimates of shear stresses do not yet provide enough information to assess whether a synthetic event can occur or not. The other four state variable types play a role as well. Stresses normal to our grid contribute to the second invariant of the deviatoric stress tensor (eq. 15, for simplicity referred to as stress), which is used for evaluation of the Drucker–Prager yielding criterion (eq. 20). Pressure and horizontal and vertical velocity contribute to the slip-rate-dependent yield strength (eq. 18, for simplicity referred to as strength). We evaluate whether stress and strength can be estimated in every phase of the seismic cycle through analysing the first occasion at which data representative for each phase is assimilated (Fig. 7).

The first column features the assimilation of interseismic data and corresponding probabilistic estimates of stress, strength and slip rate (Figs 7a, d and g, resp.). The blue posterior density or

PDF of stress captures the black true line well and distinctly better than the green prior PDF. However, the estimation of the current static strength did not improve. The slip rate estimate did improve, since several ensemble members featuring co- or post-seismic characteristics were updated to interseismic velocities towards the land instead. These results thus confirm suggestions from Figs 3 and 4 that the occurrence of events can indeed be inhibited or reverted.

In turn when coseismic data are assimilated, it is necessary—and not straightforward—that events can also be initiated in a significant number of ensemble members. Figs 7(b), (e) and (h) deals with the assimilation of data featuring a coseismic event, since the high slip rate of the true model leads to a low strength and a corresponding drop in stress. Assimilation of this data ultimately leads to very low stresses and strengths throughout the ensemble. This indicates that events are initiated in all ensemble members simultaneously, such that the stress state can be efficiently adapted towards the true state.

This distinct correction makes coseismic data more efficient and informative for data assimilation than inter- or post-seismic data.

Finally, information on the state of the fault could also come from data obtained during the post-seismic phase (Figs 7c, f and i). That the targeted synthetic data model is within the (early) post-seismic phase is evident from its very low stress and regular strength level. Interestingly, the low stresses are already forecasted by the mode of the prior. This forecasting value can be credited to the successful assimilation of data at 45 previous points in time and the following physical forward propagation there upon. However, it is only after assimilating these postseismic data that stresses at the fault are indeed estimated to be very low in the large majority of the ensemble. The strength estimate is reasonable, although it is not improved with respect to the prior estimate. Meanwhile, the slip rate estimates are improved, particularly in the forward model time steps following the assimilation (compare dotted blue PDF with dotted black line).

In summary, stress and strength state can thus be updated directly in each ensemble member to better capture the true stress state of the synthetic data model. This is facilitated by the prior knowledge of the evolution of physical variables within each phase of the seismic cycle and the explicit state updating included within the EnKF scheme. In a Particle Filter the absence of explicit state updating would need to be compensated by a very large numbers of ensemble members and effective resampling.

3.2 How well is analogue seismicity captured?

3.2.1 How well are states captured through time?

This section extends the spatial state estimation from a single assimilation round to a complete analogue experiment with 30 cycles of analogue earthquakes (Corbi *et al.* 2013). First, we analyse the temporal evolution of all state variable types, and strength excess, for the same marker in the centre of the seismogenic zone (Fig. 8). This figure depicts the ensemble through different statistical measures in red, the true evolution in black and shows grey bars when data are assimilated. Prior to the start of data assimilation at $t \approx 134$ s, the range of the ensembles for velocities and stresses mostly encompass the minimum and maximum values of the true evolution. This is because the parameters are known in a perfect model test. However, there is no notion about the timing of analogue events.

After data assimilation starts, both velocity estimates demonstrate a remarkable sense of timing of events (Figs 8a and b). Horizontal velocity, which is equivalent to slip rate of the fault, shows that even the occurrence of the first event is somewhat suggested by the seaward reversal of the mean (Fig. 8a, as shown in Figs 7d–f). Subsequently, more than 97.5 per cent of the ensemble members forecast that no velocity change will occur for a while. Just prior to the second event about 10 per cent of the ensemble members indicate an event is happening. However, it is only captured by all ensemble

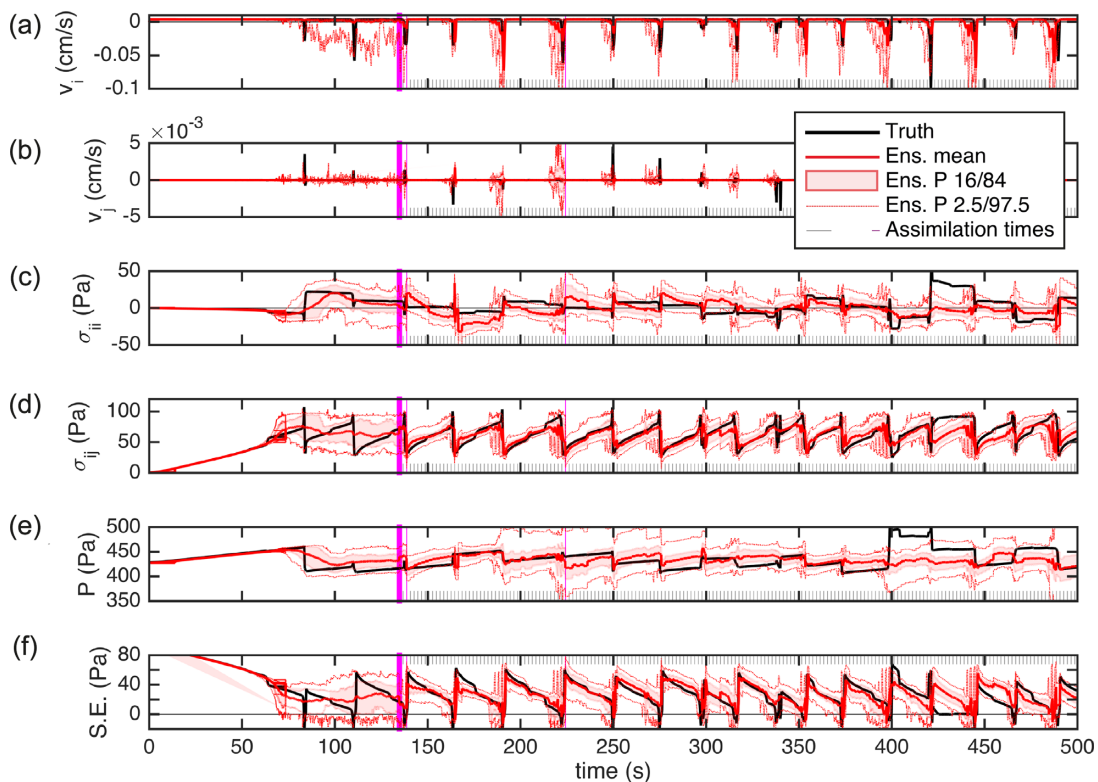


Figure 8. Temporal evolution of (a–e) all state types and (f) strength excess (S.E.) at a random fault marker in the centre of the seismogenic zone (Fig. 2, magenta circle). Strength excess is pressure- and slip-rate-dependent yield strength (eq. 18) minus second invariant of the deviatoric stress tensor. It gives an indication of the proximity of the next event, being high when the next event is far away. The black lines indicate the unknown targeted truth, while the red lines represent the probabilistic estimate of the resulting ensemble (see the legend). The thick line is the mean, while different Percentiles P are chosen to approximate plus and minus, one and two standard deviations of this ensemble distribution. Data assimilation starts at the magenta line, which thus indicates the described first assimilation step (Section 3.1). Subsequent constant assimilation times are denoted in grey at the bottom (or top) of each panel. Snap shots shown in Fig. 7 are shown in magenta. For visualization purposes a zoom of 0–500 s instead of 0–800 s is shown.

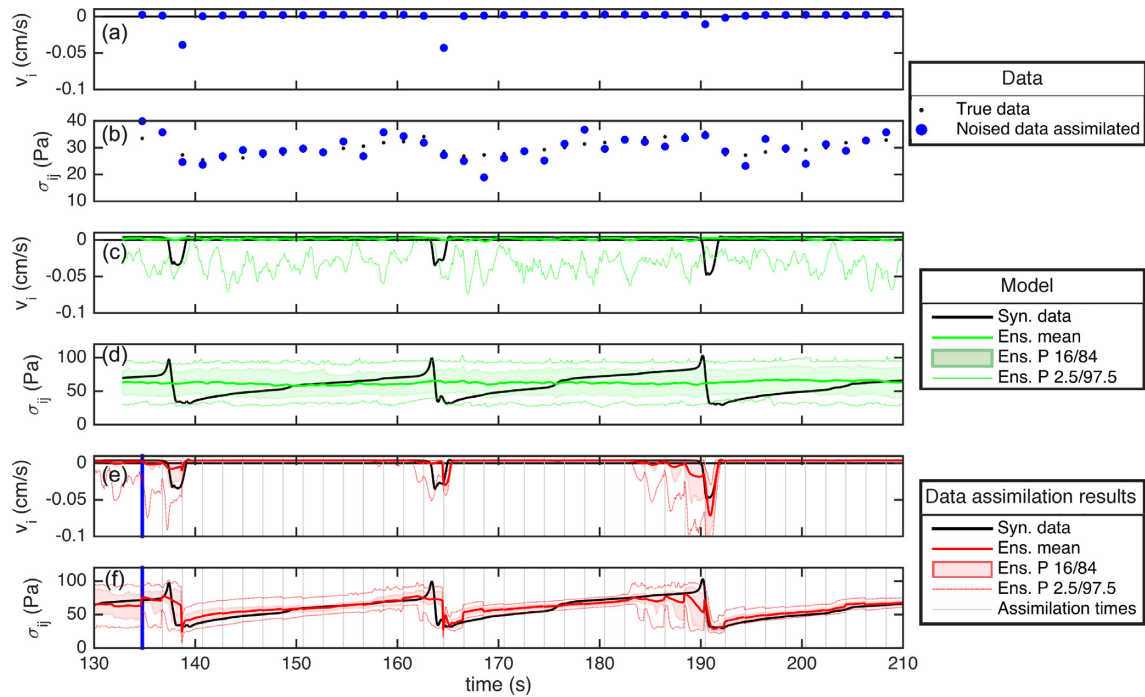


Figure 9. Information present solely within (a,b) the data and (c,d) the physical model, which together lead to (e,f) the data assimilation result. A zoom of the initial time period is shown for horizontal velocity and shear stress to better visualize the quality of the fit.

members once coseismic data are assimilated. Subsequently, the third interseismic period is forecasted to be interseismic. Seconds before the third event, a minor portion of ensemble members experiences events. One to two seconds before the event the large majority of ensemble members experiences events, indicating that the onset of an event is highly likely right before it happens. Subsequent interseismic period as forecasted as such, while events are typically preceded by a smaller or larger portion of ensemble members. However, towards the end, larger portions of members and larger times before the actual event, an event is already indicated as more likely.

These indications for the likelihood of events are also observed in the pattern and uncertainties of fault stresses, in particular for the most relevant shear stresses (Figs 8c–f). As soon as data assimilation starts, both the timing and magnitude of static stress drop are captured particularly well (Fig. 8c). However, the dynamic stress increase due to the approaching rupture front cannot be captured by the mean of an ensemble. The subsequent rate of interseismic stress build up is again captured well in most interseismic periods, occasionally even with large certainty. Normal stresses are captured somewhat, albeit for two extended periods of exceptional extension (Fig. 8d). Pressure estimates in the first half approximate the truth somewhat better than before data assimilation, although with large uncertainty (Fig. 8e). In the second half of the experiment pressure estimates deviate from the truth during two exceptional periods of extension. Since these pressure variations are not relevant and hardly impact static strength (Fig. 8f), we suggest to exclude pressure estimates in future applications.

The most crucial variable to capture is the strength excess (strength minus stress), since this directly determines the occurrence of plastic yielding and runaway events. This thus provides most information for those interested in forecasting events. Fig. 8(f) shows the evolution of strength excess is captured remarkably well,

and with uncertainty levels that are mostly within one standard deviation from the mean. Its evolution shows that after events, stresses have dropped and strength excess is thus large, thereby making the occurrence of an event very unlikely. As the strength excess decreases towards zero events become more likely. A distinct number of events seem to occur around the time when the strength excess reaches zero for a distinct portion of the ensemble members. However, fitting strength excess is more difficult near the downdip limit of the seismogenic zone, where events nucleate (Fig. E2).

3.2.2 How much information within model and data?

The degree to which fault stresses and dynamic strength are captured is remarkably high. To appreciate how surprising this is, and how much value is added through data assimilation, we analyse the information already present within the physical model and synthetic data separately (Fig. 9).

From periodic observations near the surface, it is difficult to know how many events occurred (Figs 9a–b). Horizontal velocity gives a reasonable idea, but events can be missed due to the constant sampling interval. The additional information from the assimilated shear stress (blue dot), which is noised to mimic reality, introduces uncertainty when identifying how many, and when, events occur. Several jumps in stress instead result from a randomly drawn error, which is roughly on the order of the stress drop measurable at the surface.

The ensemble of physical models provides a very good estimate of the average level of velocity and stress (Figs 9c–d), since the parameters are known in this perfect model test. However, a sense of event timing is not present at all. Velocities seem to rebound seaward regularly which could be interpreted as the occurrence of up to about 20 events.

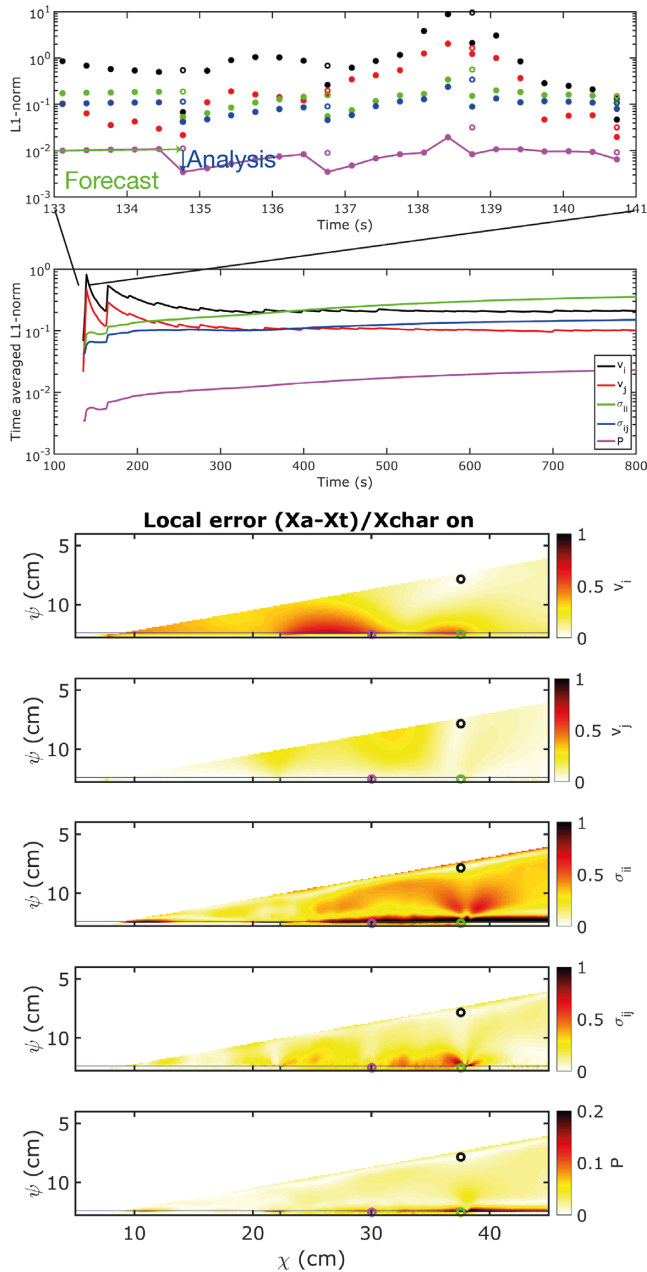


Figure 10. Error or goodness of fit through (a and b) time and (c–g) space. (a) Zoom of L1 norm evolution for each state variable type (colour, legend in b), where filled dots give the final value at each fifth forward time step and the open dot shows the forecast before assimilation. Spatial averaging was applied over the region of interest shown in (c–g) in which boundary conditions, such as the sticky air, were assigned zero. (b) Additional time averaging at each assimilation time step shows the L1-norm converges within the duration of the analogue experiment, such that errors can be compared between different experiments. (c–g) Spatial distribution of the local error ($\frac{\bar{X} - x}{\bar{X}_{char,r}}$) averaged over every assimilation step after assimilation. The open circles indicate the locations of observation (black), central seismogenic zone shown throughout paper (magenta) and point near downdip limit shown in Appendix E (green). The vertical blue lines indicate seismogenic zone limits, where horizontal ones show fault limits.

The combined data assimilation clearly indicate the occurrence of three events (Figs 9e–f). Moreover, they also estimate a reasonable static stress drop subsequent interseismic build up rate. Data assimilation thus added pronounced value with respect to using physical models or observations independently.

3.2.3 What is the error?

Since the true solution is known in this perfect model test, the goodness of fit between the best estimate of the ensemble, its analysed mean and the true solution is quantified in detail (please see Appendix A). In summary, we use a normalized L1 norm (eqs A1 and A2) to avoid large sensitivity to outliers. Through time this error typically increases during the forecasting period and decreases again when data are assimilated (Fig. 10a). In space we observe that the error typically increases away from the observation point and with depth (Figs 10c–g). This means that the deviations from the truth shown throughout the main text are relatively large (e.g. Fig. 8) and maximum near the nucleation region (Fig. E2).

The L1 norm averaged over space and time is used to compare different experiments and select optimal parameters for the data assimilation scheme (for details please see Appendix B and Fig. B1). In summary, this assimilation parameter sensitivity study reveals that EDA is feasible within a reasonable range of parameters, although care should be taken to set them carefully in order to optimize performance. Interestingly, this sensitivity study shows that state-of-the-art measurements of stress are suitable for improving EDA, despite their large error (Appendix B; Fig. B1c). In fact, reducing this error by more than a half seems to remove the variability needed to make the assimilation work well.

3.3 Can we forecast analogue seismicity?

Purely quantifying the fit of the ensemble to the truth (Sections 3.2.3, A and B; Figs 10 and B1) is rather arbitrary for this system with very nonlinear dynamics. Neither is an excellent fit with a delay in time as meaningful. Another way of quantifying the performance of EDA for the purpose of seismicity is to analyse its forecasting ability. We analyse forecasting ability with respect to a periodic recurrence model, whose mean recurrence interval is updated as time progresses. This is done to provide an honest benchmark for this simplified setup, which exhibits a quasi-periodic recurrence of quasi-characteristic events (coefficient of variation of the recurrence interval is about 0.5; Corbi *et al.* 2013; van Dinther *et al.* 2013b). Note that this time-series is neither time- nor slip-predictable, since corresponding regression coefficients are about 0.1.

Forecast-ability from a decision-making standpoint is often addressed using an error diagram (Helmstetter & Sornette 2003), as shown Fig. 11. An optimal forecast would forecast 100 per cent of the events correctly through sounding an alarm for very limited time, thereby passing near the diagrams origin at (0,0). If one includes no information and forecasts randomly, results fall along the black thin line connecting (0,1) and (1,0). To identify events we trigger an alarm when one of six different velocity thresholds is passed. For EDA results alarms are sound when 10 per cent of the ensemble members pass these thresholds.

For the highest velocity threshold, which is interpreted to represent the largest events (i.e. reddest line), data assimilation results

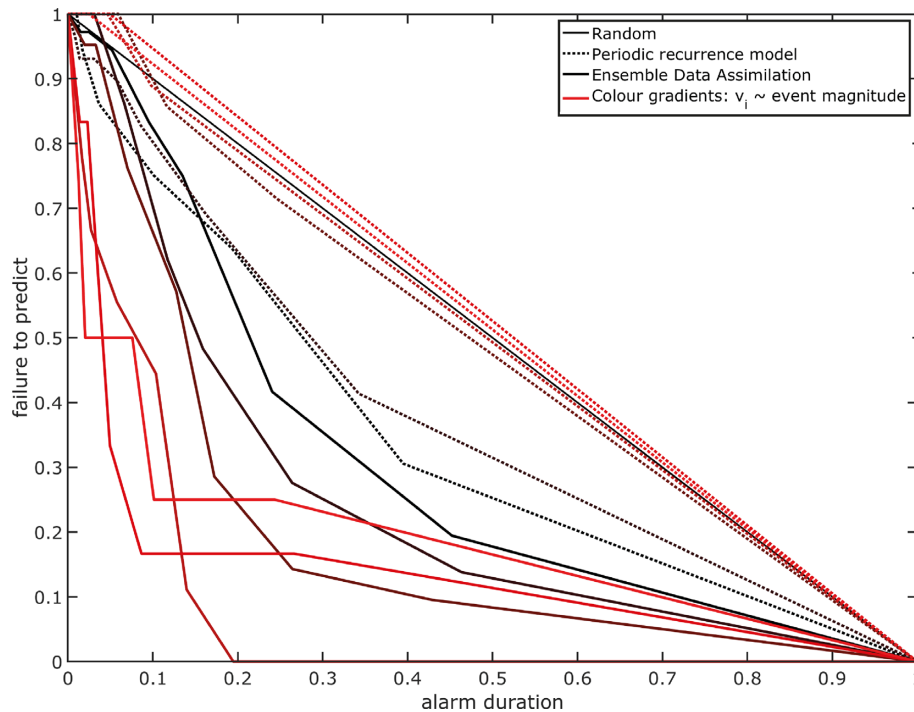


Figure 11. Earthquake forecasting potential quantified through an error diagram used for decision making. A perfect prediction would run from (0,1) to (1,0) via the origin (0,0), while a random estimate would run straight (thin black line). Colours indicate different velocity thresholds characterizing minimum magnitudes, where black is small ($v_{\chi} \leq 0 \text{ cm s}^{-1}$) and red is large ($v_{\chi} > -0.07 \text{ cm s}^{-1}$) and colours are interpolated linearly using RGB values. The dotted lines represent results from a periodic recurrence model, whereas solid lines represent ensemble data assimilation results obtained when sounding an alarm when 10 per cent of the ensemble members passes each velocity threshold. Results are obtained at a random marker in the centre of the seismogenic zone, where resp. 36, 29, 21, 9, 6 and 4 events occur for the increasing velocity thresholds used.

approach the optimal path through the origin surprisingly well. When sounding an alarm 2 per cent of the time, already half of them (2/4) is forecasted correctly. For slightly lower thresholds 9 out of 9 events are forecasted within 18 per cent of the time. However, a periodic recurrence model needs, respectively, 53 and 100 per cent of the time to forecast these black swans. That performance is thus similar to a random forecast. Data assimilation also significantly outperforms the periodic recurrence model for 21 moderately large events (dark red line). To predict 70 per cent of them correctly a periodic recurrence model needs to sound an alarm for 68 per cent of the time, while the data assimilation results only need to sound an alarm for 17 per cent of the time. It is only when we also try to forecast the smaller events ($v_{\chi} > 0 \text{ cm s}^{-1}$ for black line) that the performance of the periodic recurrence model starts to approach that of the data assimilation. The periodic recurrence model predicts 70 per cent of these 36 events correctly in 41 per cent of the time, while the EnKF outperforms that as it needs only 35 per cent of the time.

This single point quantification substantiates the general impression from Fig. 11 that data assimilation results go a long way towards the optimal point of (0,0), whereas the periodic model remains close to the random line for the large events. To appreciate this impact it is good to realize that good earthquake forecasts for natural systems typically only perform slightly better than random (e.g. Helmstetter & Sornette 2003). However, we acknowledge that this evaluation remains partially arbitrary and sensitive to threshold choices. Nonetheless, it does show that this simplified system for which the physical model is known is forecastable and even really well for its large events.

4 DISCUSSION

Our perfect model test shows that remarkable probabilistic estimates of the evolving state of stress and dynamic strength can be obtained by sequentially assimilating data. Synthetic, noised data from a single borehole effectively correct the evolution and forecast of a physical forward model represented by 150 ensemble members. Previous, pioneering data assimilation studies for earthquake sequences already indicated that there might be potential of data assimilation from a statistic point-source (Werner *et al.* 2011), scenario-based (e.g. Hori *et al.* 2014c) and data scoring, physical model perspective (van Aalsburg *et al.* 2007). Here we prove the concept that EDA methods can indeed be used—as they are—for forecasting of synthetic, analogue earthquakes. This demonstrates a clear potential for usage in probabilistic seismic hazard assessment (Section 4.2) and potentially for other solid earth directions (Section 4.3). This remarkable performance, however, only holds for this simplified, synthetic perfect model test, which has distinct limitations for the data assimilation test, setup and forward model (Section 4.1).

4.1 Limitations

The strongest assumption of any perfect model test is that the physics is known perfectly. However, the physics governing earthquakes is not known well and involves highly nonlinear processes (e.g. Rundle *et al.* 2000). We approach this by assuming basic descriptions

based on (i) first physical principles in terms of Navier–Stokes equations, which are the relevant processes in all of geophysics, and (ii) plasticity with strongly slip-rate-dependent friction, which are well observed in laboratory experiments. Furthermore, we emphasize that ignoring a representativeness error is a necessary first step in any test of concept of any data assimilation application and adds indispensable value for exploratory studies. When pursued further, data assimilation can also guide the improvement of our physical understanding, as happened in applications for paleoclimatology. Another strong assumption in a perfect model test for state estimation is that the parameters are known. *In-situ* parameters are, however, not known, although we have small-scale approximations based on abundant laboratory experiments (e.g. Di Toro *et al.* 2011). This information along with an expert interpretation of natural settings (e.g. Barbot *et al.* 2012) and their uncertainties can be included in the prior of this Bayesian framework. EDA thus allows us to include the best information available from all three sources (observations, physical models and laboratory experiments) to provide a better, probabilistic answer. However, the unavoidable presence of errors in the forecast model, initial ensemble statistics and parameters are expected to decrease the strong performance of the EnKF used in this study. This decrease could be significant, as shown in certain cases for forecasting of extreme atmospheric events (Zhang *et al.* 2006).

Another important simplification of this study is the simulation of an analogue setup, which can be scaled up to a forearc in a subduction zone. The viscoelastic properties of gelatin lead to a relatively long coseismic period with respect to a short interseismic period (Corbi *et al.* 2013; van Dinther *et al.* 2013b). This means that coseismic outliers occur more frequently than in nature (Fig. 4), which could improve the update through increased covariances between observed and unobserved variables. However, even without these additional outliers, a physical relation exists between surface horizontal velocity and fault shear stress ($R^2 \approx 0.25$). This results from a gradual change of interseismic locking, which is also observed in large-scale seismo-thermo-mechanical models (fig. 4a in van Dinther *et al.* 2013a) and GPS observations (e.g. Mavrommatis *et al.* 2014).

The forward modelling approach is appropriate for the analogue setup simulated (van Dinther *et al.* 2013b). For applications to natural seismicity, improved versions of this 2-D, low temporal resolution approach should be used. There the limited temporal resolution is tackled with adaptive time stepping and rate-and-state-dependent friction (Herrendoerfer *et al.* 2018) and the 2-D setup is extended to 3-D (Pranger *et al.* 2017). These future setups include the most accepted physical description of earthquake sequences, although these physical descriptions might need to be improved upon as well. EDA could help with that through identifying and evaluating physical descriptions, while in the mean time including a stochastic error to introduce model variability. In terms of a reasonable physical description, it is important that key observations from earthquake sequences are reproduced. Applications of this approach show agreement with observations in terms of on- and off-megathrust seismicity characteristics (albeit rupture speeds van Dinther *et al.* 2013a, 2014), Gutenberg–Richter statistics (Dal Zilio *et al.* 2018), surface displacements (Dal Zilio *et al.* 2019; Preiswerk *et al.* 2015) and long-term seismicity patterns (Herrendoerfer *et al.* 2015; Dal Zilio *et al.* 2019).

4.2 Future challenges and opportunities

These limitations show that these promising results are really only a first step into the direction of fault stress estimation and earthquake forecasting. An extension to natural faults and earthquakes remains an open question and a challenging one for the future. The estimation of state and parameters in a 3-D complex geometry requires extensive computational power and observations. In addition, the large number of states and parameters could become limited by the curse of dimensionality (e.g. Bengtsson *et al.* 2008). The hope would be that, with considerable efforts, the controlling variables can be isolated and estimated in a meaningful manner, whereas the null space primarily contains non-significant variables.

Fortunately, the rapidly evolving development of EDA in atmospheric and ocean sciences provides opportunities for how to approach very high-dimensional and dynamic, nonlinear problems. These applications typically deal with billions of variables, that is, more than $O(10^9)$ states (van Leeuwen *et al.* 2018) and $O(10^2)$ parameters, without considering spatial and temporal variability (Ruiz *et al.* 2013). Example improvements, include (i) covariance inflation (to take errors of the forward model into account) and covariance regularization and localization (to counteract spurious correlations due to limited ensemble size) (e.g. Chapter 15 in Evensen 2009b), (ii) nudging of various parts of the equations (Lakshmivarahan & Lewis 2013) and (iii) optimization of model equations via proposal densities (e.g. van Leeuwen 2009, 2016). Furthermore, parameter estimation could be efficient via (i) appending them to the state vector (e.g. Jazwinski 1970; Evensen 2009a), (ii) using an Ensemble Kalman Smoother to exploit all information available in time (e.g. Evensen & Van Leeuwen 2000) or (iii) through using variational data assimilation (*cf.* in Talagrand 1997) to in one batch account for all temporal observations (e.g. Kano *et al.* 2015 for friction parameters). Finally, if either dynamics is too strongly nonlinear, the brief disruption of dynamical balances is too severe, or assumptions on Gaussianity of the data error/likelihood and model error/prior are not valid enough, particle filters could be used for both state (e.g. van Leeuwen 2009) and parameter estimation (e.g. Doucet & Tadić 2003). To avoid their degeneracy for high-dimensional problems, various combinations between these two filter families have recently been suggested (van Leeuwen 2009), such as the Merging Particle Filter (Nakano *et al.* 2007), Equivalent-Weights Particle filter (Van Leeuwen 2010) and the Ensemble Kalman Particle Filter (Frei & Künsch 2013).

Additionally, significantly more spatial data than used in this study is available. Here we used data from a single location to facilitate our understanding and demonstrate the information already available from one point. This perfect model test indicates that relatively little data are required to capture this solid system, which varies much more smoothly and slowly through space and time than chaotic meteorological and oceanographic fields that require a lot of data. Besides the availability of more of the same or related data sources (Sections 1.2 and 2), additional constraints could come from instrumental and historical earthquake catalogues (that provide time, location, size and direction of displacement), various paleoseismological records (e.g. Grant & Gould 2004), laboratory experiments (e.g. Di Toro *et al.* 2011), high-rate GPS and interferometric synthetic aperture radar (InSAR) (e.g. Ge *et al.* 2000; Simons *et al.* 2002), creep-, strain- and tilt-meters (e.g. Wyatt *et al.* 1994), gravity measurements (e.g. Mikhailov *et al.* 2004) and seafloor displacements and pressure/temperature measurements from ocean floor observatories (e.g. Kaneda *et al.* 2014). However, combining

them in a meaningful manner—usually through separated assimilation phases and by applying localization—is a challenge in itself. In doing so it is important to know that our study shows that the large data error is not necessarily limiting their use, as long as it is accounted for properly. Moreover, our study indicated the value of stress (or stress changes) being measured through time. While stress monitoring is common in mining and other underground engineering applications (Amadei & Stephansson 1997), applications related to understanding and monitoring seismicity do typically not monitor stresses through time. This is more expensive and challenging, but our results show that stresses near the surface—even with their large error—can provide very useful indications about the evolving fault stress state.

These opportunities for improvements should be explored in a brick-by-brick approach, such that the application of EDA for rigorous physics-based earthquake forecasting can be tuned and evaluated rigorously. The choice of appropriate case studies will be important and should focus on (i) simplified, controlled laboratory or analogue setups (e.g. to evaluate the role of representativeness errors), and (ii) controlled, natural scale setups, such as for induced seismicity and (iii) natural settings with large amounts of spatial and temporal data with respect to the characteristic event recurrence time. In each of these directions perfect model tests are needed to evaluate with what methodologies and observations the truth can be best captured. Such synthetic studies also provide an economically feasible test case to evaluate how much and what types of data are needed and with what accuracy where and when are needed.

4.3 Implications for solid Earth sciences

This simplified, proof of concept indicates the potential for a range of dynamic applications within the solid Earth sciences, since observations and physical models are improving in these directions as well. Examples where our physical knowledge can extend the data in space and time, include crustal and lithosphere dynamics, geochemistry, petrology, geological reconstructions and engineering geology. In the latter direction, data assimilation has recently been successfully applied to tunnel excavation (Nguyen & Nestorović 2016) and parameter estimation for debris flow run-out (Brezzi *et al.* 2016). These two studies—in combination with our test of concept for estimating faulting—indicate the potential for both forecasting of instabilities in mines, mountain glaciers and monitoring of nuclear waste repositories. New applications can make a quick estimate of whether data assimilation could bring merit by running a suite of different models to check whether a meaningful correlation exists between observed state variables versus controlling unobserved ones (as in Fig. 4). We hope that the understanding provided within this paper guides solid earth scientists and sparks their interests into data assimilation methods for estimation of fault state evolution and beyond.

5 CONCLUSIONS

We demonstrated the remarkable probabilistic estimation and forecasting of the state of stress on an earthquake-prone fault in a perfect model test. Synthetic, noised velocity and stress observations from a single location were assimilated into 150 ensemble members of a seismic cycle model using an EnKF. Within the specified limitations of this perfect model test (i.e. physics is known perfectly) in an analogue subduction setup, we showed that distinct value was

added by EDA with respect to using information from either the physical model or data alone. To familiarize solid earth scientists with the potential of data assimilation, well-developed in weather forecasting, we provided extensive explanations on EDA and EnKFs.

Shear stresses at a fault, and throughout the medium, were estimated remarkably well based on observations at one point in space and one interseismic time. The prior physical model knowledge provides how the observation at the surface covaries with state variables throughout the model. The corresponding least-square fit is stored in the sampled model error covariance matrix. This, along with its spread and the measurement error, indicates how—and how much—to update each unobserved state variable as a function of each data misfit. Typically, when stress data indicated more compression, compression was enlarged throughout most of the seismogenic zone. This update was complicated by a large measurement error on stresses, which was on the order of the near-surface stress drop. To account for this, a lower weight was attributed to these measurements and uncertainty in the posterior estimate was increased. When horizontal velocity data indicated larger landward velocities, this indicated that the true model is slightly more advanced in the interseismic period and fault velocities and stresses should be increased accordingly. Updates according to vertical velocity data were negligible due to a large model spread and consequently large model error.

Analysing the update of fault stress and strength shows that data from each seismic cycle phase was able to add meaningful information to directly update the state of stress. Interseismic data reverted already reduced stress states to a state more compatible to latter stages of the seismic cycle. Coseismic data initiated analogue earthquakes throughout the ensemble, such that stresses were reduced in most ensemble members. Post-seismic data typically lowered fault stresses to better synchronize stress levels.

Relevant updates during each phase of the seismic cycle ensured that the evolution of the true fault shear stress (largest stress component) and horizontal velocity (largest strength component) were captured remarkably well. Interseismic periods were typically forecasted to be interseismic by the whole ensemble, while prior to events different portions of the ensemble started to indicate an event is about to come. At the same time, both interseismic build up rate and static drop of shear stress—an indication for event magnitude—are traced adequately, occasionally even with small uncertainty. Less important variables, such as vertical fault velocity and normal stress and pressure, are only estimated to a moderate degree. Ultimately, the controlling strength excess was still estimated surprisingly well.

Quantification of the goodness of (interseismic) fit using the L1 norm typically shows the expected increase of misfit during the forecasting period and a decrease due to data assimilation. An assimilation parameter sensitivity study showed that the current large stress errors did not inhibit the use of stress observations. Opposite to intuition, at least half of their current size was required to introduce enough variability into the filtering results for them to work well.

Since a fault's stress and strength determine whether it slips, being able to estimate and forecast them, allows us to continuously provide and update the probability of the occurrence of a synthetic analogue earthquake. The EnKF was shown to significantly outperform a periodic recurrence model for this quasi-periodic time series. For small events (i.e. low velocity threshold) an alarm needed to be sounded for 35 per cent instead of 41 per cent of the time to predict 70 per cent of them correctly. However, for the forecasting of 70 per cent of the 21 larger events, it significantly outperformed

the periodic benchmark by sounding an alarm only 17 per cent instead of 68 per cent of the time. Or, even only 7 per cent instead of 72 per cent of the time for the largest, four events.

This educational test suggests that combining observations and physics-based modelling through the statistical framework of EDA has clear potential for earthquake forecasting. This potential, however, needs to be explored step-by-step, since several challenges for the application to natural settings remain.

ACKNOWLEDGEMENTS

We are grateful to Didier Sornette and Jeff Anderson for discussions on data assimilation and/or earthquake forecasting, to Dave May for discussing the lay-out of the code coupling our assimilation and forward model codes, and to Keith Evans for insights on stress estimation from boreholes. We thank Takane Hori, Egill Hauks-son and an anonymous reviewer, since their comments distinctly improved the readability of this manuscript. YD expresses her gratitude to Taras Gerya for allowing her to build on top of his I2ELVIS code. Finally, we are thankful for the hospitality of the Seismology group at Utrecht University, where parts of this manuscript have been written. Simulations have been run at the Swiss National Supercomputing Center under project number s741.

REFERENCES

- Abolfathian, N., Martínez-Garzón, P. & Ben-Zion, Y., 2018. Spatiotemporal variations of stress and strain parameters in the San Jacinto fault zone, *Pure appl. Geophys.*, 1–24, doi.org/10.1007/s00024-018-2055-y.
- Amadei, B. & Stephansson, O., 1997. Monitoring of stress change, in *Rock Stress and Its Measurement*, Nolan, D. & Speed, T., pp. 361–385, Springer.
- Angelier, J., 1984. Tectonic analysis of fault slip data sets, *J. geophys. Res.*, **89**(B7), 5835–5848.
- Barbot, S., Lapusta, N. & Avouac, J.P., 2012. Under the hood of the earthquake machine: toward predictive modeling of the seismic cycle, *Science*, **336**(6082), 707–710.
- Bayes, T. & Price, R., 1763. An essay towards solving a problem in the doctrine of chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, *Phil. Trans. R. Soc.*, **53**(0), 370–418.
- Ben-Zion, Y., 2008. Collective behavior of earthquakes and faults: continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes, *Rev. Geophys.*, **46**(4), doi:10.1029/2008RG000260.
- Ben-Zion, Y. & Lyakhovsky, V., 2006. Analysis of aftershocks in a lithospheric model with seismogenic zone governed by damage rheology, *Geophys. J. Int.*, **165**(1), 197–210.
- Bengtsson, T., Bickel, P. & Li, B., 2008. Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems, in *Probability and Statistics: Essays in Honor of David A. Freedman*, pp. 316–334, Institute of Mathematical Statistics.
- Billen, M.I., 2008. Modeling the dynamics of subducting slabs, *Annu. Rev. Earth Planet. Sci.*, **36**, 325–356.
- Bocher, M., Coltice, N., Fournier, A. & Tackley, P.J., 2016. A sequential data assimilation approach for the joint reconstruction of mantle convection and surface tectonics, *Geophys. J. Int.*, **204**(1), 200–214.
- Brezzi, L., Bossi, G., Gabrieli, F., Marcato, G., Pastor, M. & Cola, S., 2016. A new data assimilation procedure to develop a debris flow run-out model, *Landslides*, **13**, 1083–1096.
- Buiter, S., Babeyko, A., Ellis, S., Gerya, T., Kaus, B., Kellner, A., Schreurs, G. & Yamada, Y., 2006. The numerical sandbox: comparison of model results for a shortening and an extension experiment, *Analog. Numer. Model. Crustal-Scale Process.*, **253**, 29–64.
- Burgers, G., van Leeuwen, P.J. & Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, **126**(6), 1719–1724.
- Burridge, R. & Knopoff, L., 1967. Model and theoretical seismicity, *Bull. seism. Soc. Am.*, **57**(3), 341–371.
- Carrassi, A., M.B., L., B. & G.E., 2018. Data assimilation in the geosciences: an overview of methods, issues, and perspectives, *WIREs Clim. Change*, **9**, doi:10.1002/wcc.535.
- Cattania, C. et al., 2018. The forecasting skill of physics-based seismicity models during the 2010–2012 Canterbury, New Zealand, earthquake sequence, *Seismol. Res. Lett.*, **89**(4), 1238–1250.
- Cochard, A. & Madariaga, R., 1994. Dynamic faulting under rate-dependent friction, *Pure appl. Geophys.*, **142**(3/4), 419–445.
- Corbi, F., Funicello, F., Moroni, M., van Dinther, Y., Mai, P.M., Dalguer, L.A. & Faccenna, C., 2013. The seismic cycle at subduction thrusts: 1. Insights from laboratory models, *J. geophys. Res.*, **118**, 1483–1501.
- Cramer, F. et al., 2011. A benchmark comparison of numerical topography - what are suitable sticky air parameters?, *Geophys. J. Int.*, **200**, 1–12.
- Daley, R., 1991. Atmospheric Data Analysis. *Cambridge Atmospheric and Space Science Series*, Cambridge University Press.
- Dalguer, L.A., Fukushima, Y., Irikura, K. & Wu, C., 2017. Best practices in physics-based fault rupture models for seismic hazard assessment of nuclear installations, *Pure appl. Geophys.*, **174**(9), 3325–3329.
- Dal Zilio, L., van Dinther, Y., Gerya, T. & Avouac, J. 2019. Bimodal seismicity in the Himalaya controlled by fault friction and geometry, *Nat. Commun.*, **10**, 48.
- Dal Zilio, L., van Dinther, Y., Gerya, T. & Pranger, C., 2018. Seismic behaviour of mountain belts controlled by plate convergence rate, *Earth planet. Sci. Lett.*, **482**, 81–92.
- de la Torre Guzman, J., Babaei, M., Shi, J.-Q., Korre, A. & Durucan, S., 2014. Coupled flow-geomechanical performance assessment of CO₂ storage sites using the Ensemble Kalman Filter, *Energy Procedia*, **63**, 3475–3482.
- Dempsey, D. & Suckale, J., 2017. Physics-based forecasting of induced seismicity at Groningen gas field, the Netherlands, *Geophys. Res. Lett.*, **44**(15), 7773–7782.
- Dieterich, J., 1979. Modeling of rock friction 1. Experimental results and constitutive equation, *J. geophys. Res.*, **84**, 2161–2168.
- Di Toro, G. et al., 2011. Fault lubrication during earthquakes, *Nature*, **471**(7339), 494–499.
- Doucet, A. & Tadić, V.B., 2003. Parameter estimation in general state-space models using particle methods, *Ann. Instit. Stat. Math.*, **55**(2), 409–422.
- Dragert, H., Wang, K. & James, T.S., 2001. A silent slip event on the deeper cascadia subduction interface, *Science*, **292**(5521), 1525–1528.
- Drucker, D. & Prager, W., 1952. Soil mechanics and plastic analysis for limit design, *Q. Appl. Math.*, **10**, 157–165.
- El-Mowafy, A. & Bilbas, E., 2016. Quality control in using GNSS CORS network for monitoring plate tectonics: a Western Australia case study, *J. Survey. Eng.*, **142**, doi:10.1061/(ASCE)SU.1943-5428.0000157.
- Erickson, B.A., Dunham, E.M. & A., K., 2017. A finite difference method for off-fault plasticity throughout the earthquake cycle, *J. Mech. Phys. Solids*, **109C**, 50–77.
- 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.
- Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dyn.*, **53**(4), 343–367.
- Evensen, G., 2009a. The ensemble Kalman filter for combined state and parameter estimation, *IEEE Control Syst. Mag.*, **29**(3), 83–104.
- Evensen, G., 2009b. *Data assimilation - The Ensemble Kalman Filter*, Springer.
- Evensen, G. & Van Leeuwen, P.J., 2000. An ensemble Kalman smoother for nonlinear dynamics, *Mon. Weather Rev.*, **128**(6), 1852–1867.
- Fournier, A., Nérger, L. & Aubert, J., 2013. An ensemble Kalman filter for the time-dependent analysis of the geomagnetic field, *Geochem. Geophys. Geosyst.*, **14**(10), 4035–4043.
- Frei, M. & Künsch, H.R., 2013. Bridging the ensemble Kalman and particle filters, *Biometrika*, **100**(4), 781–800.
- Fulton, P.M. et al., 2013. Low coseismic friction on the Tohoku-Oki fault determined from temperature measurements, *Science*, **342**(6163), 1214–1217.

- Ge, L. *et al.*, 2000. GPS seismometers with up to 20 Hz sampling rate, *Earth Planets Space*, **52**(10), 881–884.
- Geller, R.J., 1997. Earthquakes cannot be predicted, *Science*, **275**, 1616–1617.
- Geller, R.J., Mulargia, F. & Stark, P., 2016. Why we need a new paradigm of earthquake occurrence, in *Subduction Dynamics From Mantle Flow to Mega Disasters*, eds. Morra, G., Yuen, D.A., King, S.D., Lee, S.-M., Lee, S. & Stein, S., Wiley.
- Gerya, T., 2010. Dynamical instability produces transform faults at mid-ocean ridges, *Science*, **329**, 1047–1050.
- Gerya, T. & Yuen, D., 2007. Robust characteristics method for modelling multiphase visco-elasto-plastic thermo-mechanical problems, *Phys. Earth planet. Inter.*, **163**(1–4), 83–105.
- Gordon, N., Salmond, D. & Smith, A., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *Proc. IEEE*, **140**, 107–113.
- Goswami, A. & Barbot, S., 2018. Slow-slip events in semi-brittle serpentinite fault zones, *Scientific Reports*, **8**, 6181, doi:10.1038/s41598-018-24637-z.
- Gottwald, G.A. & Majda, A.J., 2013. A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks, *Nonlinear Process. Geophys.*, **20**(5), 705–712.
- Grant, L.B. & Gould, M.M., 2004. Assimilation of paleoseismic data for earthquake simulation, *Pure appl. Geophys.*, **161**(11–12), 2295–2306.
- Gregg, P.M. & Pettijohn, J.C., 2016. A multi-data stream assimilation framework for the assessment of volcanic unrest, *J. Volc. Geotherm. Res.*, **309**(C), 63–77.
- Grêt, A., Snieder, R. & Özbay, U., 2006. Monitoring in situ stress changes in a mining environment with coda wave interferometry, *Geophys. J. Int.*, **167**(2), 504–508.
- Hardebeck, J.L. & Michael, A.J., 2006. Damped regional-scale stress inversions: methodology and examples for southern California and the Coalinga aftershock sequence, *J. geophys. Res.*, **111**(B11310), 1–10.
- Harris, R.A. *et al.*, 2018. A suite of exercises for verifying dynamic earthquake rupture codes, *Seismol. Res. Lett.*, **89**(3), 1146–1162.
- Heidbach, O. *et al.*, 2018. The World Stress Map database release 2016: Crustal stress pattern across scales, *Tectonophysics*, **744**, 484–498.
- Helmstetter, A. & Sornette, D., 2003. Predictability in the epidemic-type aftershock sequence model of interacting triggered seismicity, *J. geophys. Res.*, **108**(B10), 2482–18.
- Herrendoerfer, R., Gerya, T. & van Dinther, Y., 2018. An invariant rate- and state-dependent friction formulation for viscoelastoplastic earthquake cycle simulations, *J. geophys. Res.*, **123**(6), 5018–5051.
- Herrendoerfer, R., van Dinther, Y., Gerya, T. & Dalguer, L., 2015. Earthquake supercycle in subduction zones controlled by the width of the seismogenic zone, *Nat. Geosci.*, **8**, 471–474.
- Hori, T., Hyodo, M., Miyazaki, S. & Kaneda, Y., 2014a. Numerical forecasting of the time interval between successive M8 earthquakes along the Nankai Trough, southwest Japan, using ocean bottom cable network data, *Mar. Geophys. Res.*, **35**(3), 285–294.
- Hori, T., Hyodo, M., Nakata, R., Miyazaki, S. & Kaneda, Y., 2014b. A forecasting procedure for plate boundary earthquakes based on sequential data assimilation, *Oceanography*, **27**(2), 94–102.
- Hori, T., Miyazaki, S., Hyodo, M., Nakata, R. & Kaneda, Y., 2014c. Earthquake forecasting system based on sequential data assimilation of slip on the plate boundary, *Theor. Appl. Mech. Japan*, **62**, 179–189.
- Houtekamer, P.L. & Mitchell, H.L., 1998. Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, **126**(3), 796–811.
- Huffman, K. & Saffer, D., 2016. In situ stress magnitudes at the toe of the Nankai Trough accretionary prism, offshore Shikoku island, Japan, *J. geophys. Res.*, **121**(2), 1202–1217.
- Jazwinski, A., 1970. *Stochastic Processes and Filtering Theory*, vol. **64**, Academic Press.
- Kalman, R., 1960. A new approach to linear filtering and prediction problems, *J. Basic Eng.*, **82** (1), 35–45.
- Kaneda, Y. *et al.*, 2014. Advanced technology and application of the ocean floor observatory around the Nankai trough, Southwestern Japan, in *OCEANS 2014*, IEEE.
- Kano, M., Miyazaki, S., Ishikawa, Y., Hiyoshi, Y., Ito, K. & Hirahara, K., 2015. Real data assimilation for optimization of frictional parameters and prediction of afterslip in the 2003 Tokachi-oki earthquake inferred from slip velocity by an adjoint method, *Geophys. J. Int.*, **203**(1), 646–663.
- King, G.C.P., Stein, R.S. & Lin, J., 1994. Static stress changes and the triggering of earthquakes, *Bull. seism. Soc. Am.*, **84**(3), 935–953.
- Lakshmivarahan, S. & Lewis, J.M., 2013. Nudging methods: a critical overview, in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, eds. Park, S.K. & Xu, L., Data Assimilation for Atmospheric.
- Lapusta, N. & Barbot, S., 2012. Models of earthquakes and aseismic slip based on laboratory-derived rate- and state friction laws, in *The Mechanics of Faulting: From Laboratory to Real Earthquakes*, pp. 153–207, eds. Bizzarri, A. & Bhat, H.S., Research Signpost.
- Lapusta, N. *et al.*, 2019. Modeling Earthquake Source Processes: from Tectonics to Dynamic Rupture, *Report to the National Science Foundation*.
- Lapusta, N., Rice, J., Ben-Zion, Y. & Zheng, G., 2000. Elastodynamic analysis for slow tectonic loading with spontaneous rupture episodes on faults with rate- and state-dependent friction, *J. geophys. Res.*, **105**(B10), 23 765–23 789.
- Lindsey, N.J., Martin, E.R., Dreger, D.S., Freifeld, B., Cole, S., James, S.R., Biondi, B.L. & Ajo-Franklin, J.B., 2017. Fiber-optic network observations of earthquake wavefields, *Geophys. Res. Lett.*, **44**(23), 11 792–11 799.
- Lin, W. *et al.*, 2013. Stress state in the largest displacement area of the 2011 Tohoku-Oki earthquake, *Science*, **339**(6120), 687–690.
- Lin, Y.Y. & Lapusta, N., 2018. Microseismicity simulated on asperity-like fault patches: on scaling of seismic moment with duration and seismological estimates of stress drops, *Geophys. Res. Lett.*, **45**(16), 8145–8155.
- Llenos, A.L. & McGuire, J.J., 2011. Detecting aseismic strain transients from seismicity data, *J. geophys. Res.*, **116**(B6), 197–17.
- Lyakhovsky, V., Hamiel, Y. & Ben-Zion, Y., 2011. A non-local visco-elastic damage model and dynamic fracturing, *J. Mech. Phys. Solids*, **59**, 1752–1776.
- Marzocchi, W., Taroni, M. & Falcone, G., 2017. Earthquake forecasting during the complex Amatrice-Norcia seismic sequence, *Sci. Adv.*, **3**, 7, doi:10.1126/sciadv.1701239.
- Mavrommatis, A.P., Segall, P. & Johnson, K., 2014. A decadal-scale deformation transient prior to the 2011 M_w 9.0 Tohoku-oki earthquake, *Geophys. Res. Lett.*, **44**, 4486–4494.
- Michel, S., Avouac, J.-P., Lapusta, N. & Jiang, J., 2017. Pulse-like partial ruptures and high-frequency radiation at creeping-locked transition during megathrust earthquakes, *Geophys. Res. Lett.*, **44**(16), 8345–8351.
- Mikhailov, V., Tikhotsky, S., Diamant, M., Panet, I. & Ballu, V., 2004. Can tectonic processes be recovered from new gravity satellite data?, *Earth planet. Sci. Lett.*, **228**(3–4), 281–297.
- Moreno, M., Rosenau, M. & Oncken, O., 2010. 2010 Maule earthquake slip correlates with pre-seismic locking of Andean subduction zone, *Nature*, **467**, 198–202.
- Moresi, L., Dufour, F. & Mühlhaus, H., 2003. A Lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials, *J. Comput. Phys.*, **184**(2), 476–497.
- Murakami, A., Shuku, T., Nishimura, S.-i., Fujisawa, K. & Nakamura, K., 2012. Data assimilation using the particle filter for identifying the elastoplastic material properties of geomaterials, *Int. J. Numer. Anal. Methods Geomech.*, **37**(11), 1642–1669.
- Nagao, 2014. What is required for data assimilation that is applicable to big data in the solid earth science?, in *17th International Conference on Information Fusion*, IEEE, Salamanca, Spain.
- Nakano, S., Ueno, G. & Higuchi, T., 2007. Merging particle filter for sequential data assimilation, *Nonlinear Process. Geophys.*, **14**(4), 395–408.
- Ng, G.-H.C., McLaughlin, D., Entekhabi, D. & Ahanin, A., 2011. The role of model dynamics in ensemble Kalman filter performance for chaotic systems, *Tellus*, **63**(5), 958–977.

- Nguyen, L.T. & Nestorović, T., 2016. Nonlinear Kalman filters for model calibration of soil parameters for geomechanical modeling in mechanized tunneling, *J. Comput. Civil Eng.*, **30**(2), 04015025–11.
- Noda, H. & Lapusta, N., 2013. Stable creeping fault segments can become destructive as a result of dynamic weakening, *Nature*, 1–6.
- Ogata, Y., 1999. Estimating the hazard of rupture using uncertain occurrence times of paleoearthquakes, *J. geophys. Res.*, **104**(B8), 17995–18014.
- Pranger, C., van Dinther, Y., Herrendoerfer, R., Le Pourhiet, L., May, D. & Gerya, T., 2017. Earthquakes as plastic failure on spontaneously evolving faults, in *Geophysical Research Abstracts of EGU General Assembly*, Vol. **20**, Vienna, Austria.
- Preiswerk, L., van Dinther, Y. & Gerya, T.V., 2015. A secondary zone of uplift caused by megathrust earthquakes, *Pure appl. Geophys.*, T42–B08.
- Rice, J., 1993. Spatio-temporal complexity of slip on a fault, *J. geophys. Res.*, **98**, 9885–9907.
- Ruiz, J.J., Pulido, M. & Miyoshi, T., 2013. Estimating model parameters with ensemble-based data assimilation: a review, *J. Meteorol. Soc. Japan. Ser. II*, **91**(2), 79–99.
- Rundle, J., Turcotte, D. & Shcherbakov, R., 2003. Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems, *Rev. Geophys.*, **41**, doi:10.1029/2003RG000135
- Rundle, J.B., Turcotte, D.L. & Klein, W., 2000. *Geocomplexity and the Physics of Earthquakes*, Vol. **120**, American Geophysical Union.
- Rundle, P.B., Rundle, J.B., Tiampo, K.F., Donnellan, A. & Turcotte, D.L., 2006. Virtual California: fault model, frictional parameters, applications, *Pure appl. Geophys.*, **163**(9), 1819–1846.
- Schmalholz, S., Podladchikov, Y. & Schmid, D., 2001. A spectral/finite difference method for simulating large deformations of heterogeneous, viscoelastic materials, *Geophys. J. Int.*, **145**(1), 199–208.
- Scholz, C.H., 2015. On the stress dependence of the earthquake b value, *Geophys. Res. Lett.*, **42**(5), 1399–1402.
- Segall, P. & Matthews, M., 1997. Time dependent inversion of geodetic data, *J. geophys. Res.*, **102**(B10), 22 391–22 409.
- Segou, M., Parsons, T. & Ellsworth, W., 2013. Comparative evaluation of physics-based and statistical forecasts in Northern California, *J. geophys. Res.*, **118**(12), 6219–6240.
- Shaw, B.E., Milner, K.R., Field, E.H., Richards-Dinger, K., Gilchrist, J.J., Dieterich, J.H. & Jordan, T.H., 2018. A physics-based earthquake simulator replicates seismic hazard statistics across California, *Sci. Adv.*, **4**(8), doi:10.1126/sciadv.aau0688.
- Simons, M., Fialko, Y. & Rivera, L., 2002. Coseismic deformation from the 1999 mw 7.1 Hector mine, California, earthquake as inferred from insar and GPS observations, *Bull. seism. Soc. Am.*, **92**(4), 1390–1402.
- Talagrand, O., 1997. Assimilation of observations, an introduction, *J. Meteorol. Soc. Japan Ser. 2*, **75**, 81–99.
- Terakawa, T. & Hauksson, E., 2018. Absolute stress fields in the source region of the 1992 Landers earthquake, *J. geophys. Res.*, **123**(10), 8874–8890.
- Townend, J., 2006. What do faults feel? Observational constraints on the stresses acting on seismogenic faults, in *Earthquakes: Radiated Energy and the Physics of Faulting*, pp. 313–327, American Geophysical Union.
- Tsuji, T., Kamei, R. & Pratt, R.G., 2014. Pore pressure distribution of a mega-splay fault system in the nankai trough subduction zone: Insight into up-dip extent of the seismogenic zone, *Earth planet. Sci. Lett.*, **396**, 165–178.
- Valley, B. & Evans, K.F., 2007. Stress state at Soultz-Sous-Forêts to 5 km depth from wellbore failure and hydraulic observations, in *Thirty-Second Workshop on Geothermal Reservoir Engineering*, pp. 1–10, Stanford University, Stanford, California.
- van Aalsburg, J., Grant, L.B., Yakovlev, G., Rundle, P.B., Rundle, J.B., Turcotte, D.L. & Donnellan, A., 2007. A feasibility study of data assimilation in numerical simulations of earthquake fault systems, *Phys. Earth Planet. Inter.*, **163**(1–4), 149–162.
- van Dinther, Y., Gerya, T., Dalguer, L., Mai, P., Morra, G. & Giardini, D., 2013a. The seismic cycle at subduction thrusts: insights from seismo-thermo-mechanical models, *J. geophys. Res.*, **118**, 6183–6202.
- van Dinther, Y., Gerya, T.V., Dalguer, L.A., Corbi, F., Funicello, F. & Mai, P.M., 2013b. The seismic cycle at subduction thrusts: 2. Dynamic implications of geodynamic simulations validated with laboratory models, *J. geophys. Res.*, **118**, 1502–1525.
- van Dinther, Y., Mai, P.M., Dalguer, L.A. & Gerya, T.V., 2014. Modeling the seismic cycle in subduction zones: the role and spatiotemporal occurrence of off-megathrust events, *Geophys. Res. Lett.*, **41**(4), 1194–1201.
- van Leeuwen, P.J., 2009. Particle filtering in geophysical systems, *Mon. Weather Rev.*, **137**(12), 4089–4114.
- Van Leeuwen, P.J., 2010. Nonlinear data assimilation in geosciences: an extremely efficient particle filter, *Q. J. Royal Meteorol. Soc.*, **136**(653), 1991–1999.
- van Leeuwen, P.J., 2016. *Particle Filters for Nonlinear Data Assimilation in High-Dimensional Systems*, Tech. rep., Department of Meteorology, University of Reading, United Kingdom, Taught at Data-Assimilation Research Centre (DARC). Available at: <http://www.met.reading.ac.uk/xv901096/research/ParticleFilter-LectureNotes.pdf>.
- van Leeuwen, P.J., Künsch, H.R., Nerger, L., Potthast, R. & Reich, S., 2018. Particle filters for applications in geosciences, *Earth*, arXiv:1807.10434.
- Varini, E., 2007. A Monte Carlo method for filtering a marked doubly stochastic Poisson process, *Stat. Methods Appl.*, **17**(2), 183–193.
- Wegler, U. & Sens-Schönfelder, C., 2007. Fault zone monitoring with passive image interferometry, *Geophys. J. Int.*, **168**(3), 1029–1033.
- Werner, M., Ide, K. & Sornette, D., 2011. Earthquake forecasting based on data assimilation: sequential monte carlo methods for renewal point processes, *Nonlinear Process. Geophys.*, **18**(1), 49–70.
- Whitaker, J.S. & Hamill, T.M., 2002. Ensemble data assimilation without perturbed observations, *Mon. Weather Rev.*, **130**(7), 1913–1924.
- Wyatt, F.K., Agnew, D.C. & Gladwin, M., 1994. Continuous measurements of crustal deformation for the 1992 Landers earthquake sequence, *Bull. seism. Soc. Am.*, **84**(3), 768–779.
- Yoder, M.R., Schultz, K.W., Heien, E.M., Rundle, J.B., Turcotte, D.L., Parker, J.W. & Donnellan, A., 2015. The Virtual Quake earthquake simulator: a simulation-based forecast of the El Mayor-Cucapah region and evidence of predictability in simulated earthquake sequences, *Geophys. J. Int.*, **203**(3), 1587–1604.
- Zhang, F., Meng, Z. & Aksoy, A., 2006. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: Perfect model experiments, *Mon. Weather Rev.*, **134**(2), 722–736.
- Zoback, M. et al., 2003. Determination of stress orientation and magnitude in deep wells, *Int. J. Rock Mech. Min. Sci.*, **40**(7), 1049–1076.
- Zöller, G. & Ben-Zion, Y., 2014. Large earthquake hazard of the San Jacinto Fault Zone, CA, from long record of simulated seismicity assimilating the available instrumental and paleoseismic data, *Pure appl. Geophys.*, **171**, 2955–2965.

SUPPORTING INFORMATION

Supplementary data are available at [GJI](#) online.

Figure S1. Schematic diagram explaining the concept of EnKFs. (a) Temporal evolution of a true (black), forecasted (green), analysed (blue) and observed (red) state variable with sequential data assimilation at each Assimilation Step (AS). The goal is to bring the forecast, propagated based on a physical model, in line with the truth. The concept of the (b) propagation and (d) update steps is illustrated for many ensemble members forming a PDF of the (b) prior $p^f(x)$, (c) data likelihood $p(y|x)$ and (d) analysis $p^a(x|y)$ for an observed velocity v_x at the surface (at grey grid node 2) and a hidden stress $\sigma_{x,\psi}$ at the fault (at grid node 21). In (d) note that each ensemble member is moved by the coloured arrows to form a posterior, which is the multiplication of the transparent prior

and data likelihood. This possibly non-Gaussian posterior ideally approximates the truth (black bar).

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

APPENDIX A: WHAT IS THE ERROR?

Since the true solution is known in this perfect model test, the goodness of fit between the best estimate of the ensemble, its analysed mean $\bar{\mathbf{x}}^a$ and the true solution \mathbf{x} can be quantified. The resulting local error ϕ_r , at each state variable p at a single assimilation time step can thus be written as

$$\phi_r = \frac{\bar{\mathbf{x}}_r^a - \mathbf{x}_r}{x_{\text{char},r}}. \quad (\text{A1})$$

The error is normalized with the characteristic value $x_{\text{char},r}$ of each data type (Table 1) to ensure that the most relevant values near the characteristic one are weighted most, while allowing for comparability from one nodal location to the next (although not between state variable types). This error estimate mainly serves to provide a quantitative means to compare different experiments and select optimal parameters for the data assimilation scheme (Appendix B). The related choice of a misfit function for the state vector through time is a tricky one, which again depends on a choice on what to penalize. We chose to use the robust L1 norm (or least absolute error) to avoid a large sensitivity to outliers. We divide it by the number of occurrence in space or time to have a better understanding of its magnitude. These measures thus purely quantify the goodness of fit and focus on an interseismic fit, whereas they do not penalize event timing or forecast-ability.

We calculate the L1 norm for the relevant parts of the state vector s (excluding boundary conditions and their influence regions, as shown in Figs 10 c–g) at each time as

$$\|\phi_s\|_1 = \sum_{s=1}^S |\phi_s|/S. \quad (\text{A2})$$

The evolution of this error for each state variable type shows that it typically increases during the forecasting period (Fig. 10a), where these forecasts rely purely on the physical forward modelling. This divergence of stress and pressure estimates due to variable event timing provides the variability needed to make data assimilation work. The temporal estimates for interseismic velocities behave more erratically, since their error is influenced by the timing of assimilation with respect to the stage in the seismic cycle. During the analysis step, in which data is assimilated, errors are reduced.

Averaged over all assimilation time steps, we find that the time-averaged L1 norm for velocities decreases and stabilizes quickly, as the interseismic signal starts to dominate (Fig. 10b). The time-averaged L1 norm of stresses nearly converges with time, although the amount of divergence cannot be fully counteracted by assimilating noised data. Interestingly, using this averaging scheme for evaluating data errors in stresses shows that current stress measurements with their large error are suitable for improving EDA (Appendix B; Fig. B1c). In fact, reducing this error by more than a half seems to remove the variability needed to make the assimilation work well.

The spatial distribution of the local error after 800 s of simulation shows that the error typically increases away from the observational

location and mainly with depth (Figs 10c–g). It is thus typically largest near the fault, throughout whose seismogenic zone it also increases with depth. The largest error is thus typically found near the downdip limit. In that nucleation region, the occurrence of minor events not affecting the surface, introduces ambiguities, which cannot be captured as well as other locations in the model can. This distribution thus indicates a degree of sensitivity for the temporal results shown throughout the paper. The central seismogenic zone marker thus showed the evolution for intermediate errors (from a fault perspective), whereas the figures in Appendix E illustrate a good enough fit even for the most difficult, albeit most important region. That downdip fit is essential, since the nucleation of the large events rupturing the whole seismogenic zone determine whether stresses can be captured more updip and whether data assimilation works for event timing.

APPENDIX B: HOW SENSITIVE ARE THESE RESULTS TO DATA ASSIMILATION PARAMETERS?

The near convergence of the error estimates with time observed above ensures that we can compare error estimates from one experiment to another to select the optimal parameters for the data assimilation procedure (Fig. B1). Over the wide test range, the L1 norm can vary up to a few times or factor 5 (or 10 in case of zero data error), which makes the difference between a scheme that works well or does not quite work. In summary, increasing the number of ensemble members N , decreases the L1 norm as $1/\sqrt{N}$ for the normal stress, as expected from Monte Carlo sampling (Fig. B1a). However, this convergence rate is only half of that for the other state variable types.

The duration of the forecasted interval without assimilation is most sensitive to the error measure used (Figs B1b, e and f). The error decreases distinctly with increasing number of forward time steps per assimilation interval. However, this partially results from a lower chance of measuring the error at a coseismic time, which typically leads to larger errors. This increased weight for an interseismic fit, and underweighting for event occurrence and timing, inhibits one from capturing the stress evolution (Fig. B1f). Through weighting in event forecasting, both visually and using Fig. 11, we selected an intermediate interval before stress errors start to fluctuate of 30 forward time steps (or about 15 data points per recurrence interval). This efficiently prevents too much divergence of stresses, but does not oversample to introduce unconstrained, partially random updates.

Additionally, we analysed the role of the data error to understand what would be possible if the error on data measurements reduced in the future. Interestingly, data errors that are less than half of what is currently feasible only increase the stress and pressure error (Fig. B1c). Second, opposite to intuition, results for zero data error are truly bad. This extreme divergence for very precise, sparse observations is also observed in other seismological (Hori *et al.* 2014c) and meteorological studies (e.g. Houtekamer & Mitchell 1998; Ng *et al.* 2011). The mechanisms leading to this filter divergence are discussed, also for the case of small or even zero measurement errors, if observations are sparse (e.g. Gottwald & Majda 2013). It is related to errors in the estimation of the background covariance and to dynamic properties of the forward model. To get some physical intuition, one can also see it as there not being enough variability in the system, as random noise can not be added to observations

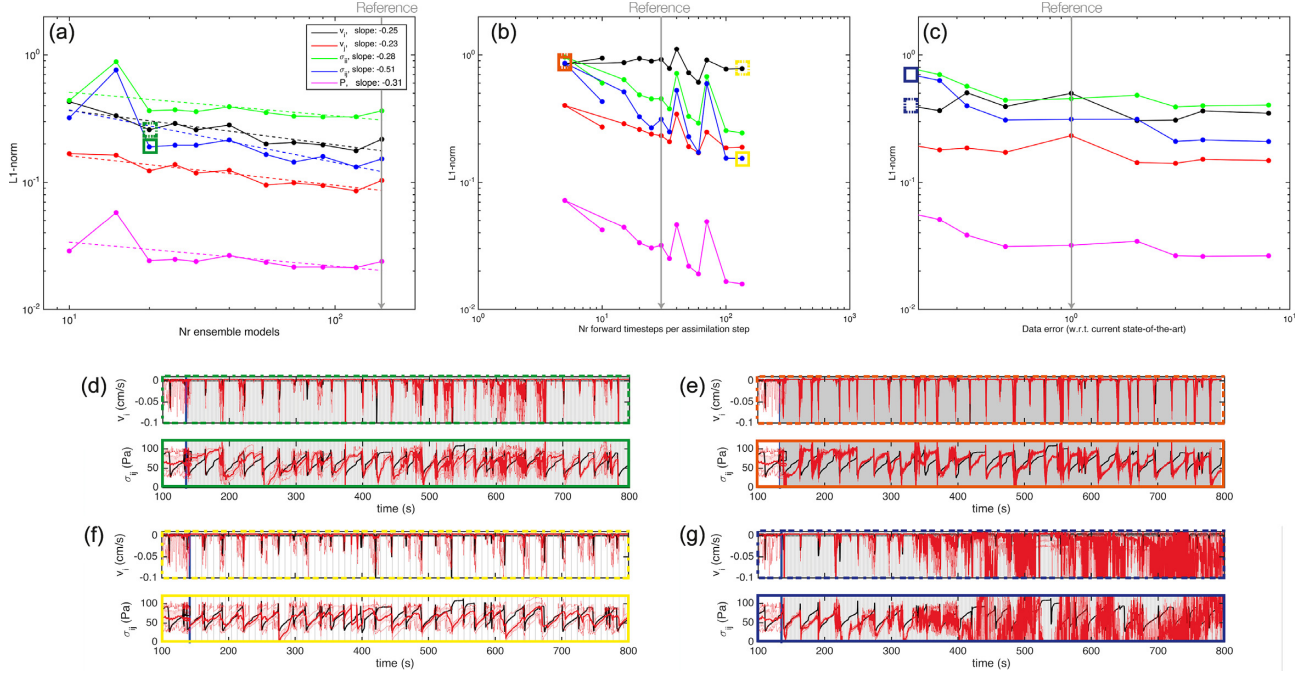


Figure B1. Sensitivity of data assimilation parameters through L1 norm in space and time (i.e., eq. (A2) averaged over time) shown for each state variable type (colour). Role of (a) number of ensemble members, including a power-law fit, (b) size of constant interval at which to assimilate data (error calculated for X_f at each five forward time steps for comparability) and (c) data error normalized with respect to current state-of-the-art error. Grey line shows parameters for reference model from paper, while this sensitivity study was instead run with 20 ensemble members, whose temporal horizontal velocity and shear stress results are shown in (d). Additional examples in (e–g) are intended to illustrate the meaning of our error formulation and relate to the coloured squares in (a–c). For legend and explanation plots see Figs 8 and 9. Examples are shown for the random marker in the centre of the seismogenic zone only. Comparability of results is ensured through using a fixed seed for the random number generator.

(Burgers *et al.* 1998). This small spread can be seen by the very narrow range of the ensemble estimates in the first part in Fig. B1(g). In our case this impact is likely severe, since stress errors are relatively large with respect to the background errors or system noise (Whitaker & Hamill 2002). Overall, the large stress and pressure errors are thus not limiting our data assimilation performance. Instead, when accounting for the large errors properly, they introduce a required amount of variability. One thus does not need to wait with using them, until the boreholes measures improve.

In summary, data assimilation parameter selection is relevant to optimize performance, although feasible results can be obtained over a broad enough range of parameters. These results over the full durations of experiments show that most optimal results would be obtained by assimilating data with half of their current state-of-the-art error into 120 ensemble members every 35 forward time steps.

APPENDIX C: KALMAN FILTER

To derive the regular Kalman Filter equations (Kalman 1960), we start from a discretized state and linear observation system in eqs (1) (first half) and (3), where the errors are unknown. We treat these errors as random variables and formulate the statistical hypothesis that they are unbiased (i.e. $\bar{\epsilon}_x = 0$ and $\bar{\epsilon}_y = 0$) and unrelated (i.e. $\bar{\epsilon}_x \bar{\epsilon}_y = 0$). Additionally, we assume that we know (i) the model error covariance matrix \mathbf{C}_{xx}^f , which contains how each component of the state vector covaries with each other component ($\mathbf{C}_{xx}^f = \bar{\epsilon}_x \bar{\epsilon}_x^T$), and (ii) the measurement error covariance matrix \mathbf{C}_{yy} ($= \bar{\epsilon}_y \bar{\epsilon}_y^T$). Finally, for the regular KF equations, we assume a normal Gaussian distribution for both the model and data errors. Therefore,

the unknown true state \mathbf{x} has a prior density p^f of ϵ_x around the known forecast $\bar{\mathbf{x}}^f$

$$p^f(\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\bar{\mathbf{x}}^f - \mathbf{x})^T (\mathbf{C}_{xx}^f)^{-1} (\bar{\mathbf{x}}^f - \mathbf{x})\right). \quad (\text{C1})$$

The term ‘prior density’ emphasizes that this density describes the uncertainty about the state before the new measurements \mathbf{y} become available. The density of the new measurements \mathbf{y} depends on the unknown true state at each measurement location $\mathbf{M}\mathbf{x}$ as

$$p(\mathbf{y}|\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{M}\mathbf{x})^T (\mathbf{C}_{yy})^{-1} (\mathbf{y} - \mathbf{M}\mathbf{x})\right). \quad (\text{C2})$$

Inserting these two Gaussian distributions into Bayes’ theorem (eq. 2) by multiplying them leads to a Gaussian analysis or posterior density p^a of the true state (Fig. 1d)

$$p^a(\mathbf{x}) = p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p^f(\mathbf{x}) = \exp\left(-\frac{1}{2}\mathcal{J}[\mathbf{x}]\right), \quad (\text{C3})$$

where

$$\mathcal{J}[\mathbf{x}] = (\bar{\mathbf{x}}^f - \mathbf{x})^T (\mathbf{C}_{xx}^f)^{-1} (\bar{\mathbf{x}}^f - \mathbf{x}) + (\mathbf{y} - \mathbf{M}\mathbf{x})^T (\mathbf{C}_{yy})^{-1} (\mathbf{y} - \mathbf{M}\mathbf{x}). \quad (\text{C4})$$

This misfit functional \mathcal{J} weights both the distance to the forecast $\bar{\mathbf{x}}^f$ and the distance to the observations \mathbf{y} . To approximate the true state \mathbf{x} , we find the maximum likelihood estimate $\bar{\mathbf{x}}^a$, located at the mean of the Gaussian posterior density, by fitting it into an identity formulation

$$\mathcal{J}[\mathbf{x}] = (\mathbf{x} - \bar{\mathbf{x}}^a)^T (\mathbf{C}_{xx}^a)^{-1} (\mathbf{x} - \bar{\mathbf{x}}^a) + \text{non-}\mathbf{x} \text{ terms}. \quad (\text{C5})$$

Through calculating the gradient of eq. (C4), setting it equal to zero and reordering of terms (Appendix D) one can prove that the

maximum likelihood model is

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{C}_{xx}^f \mathbf{M}^T (\mathbf{M} \mathbf{C}_{xx}^f \mathbf{M}^T + \mathbf{C}_{yy})^{-1} (\mathbf{y} - \mathbf{M} \bar{\mathbf{x}}^f) \quad (\text{C6})$$

with the analysed error covariance matrix

$$\mathbf{C}_{xx}^a = \mathbf{C}_{xx}^f - \mathbf{K} \mathbf{M} \mathbf{C}_{xx}^f. \quad (\text{C7})$$

APPENDIX D: PROOF FOR DERIVATION OF KALMAN FILTER EQUATIONS

This appendix shows how to derive the Kalman Filter equations (eqs C6–5) from the variational function in eq. (C4).

First we multiply out the terms in eq. (C4), and group them into terms quadratic in \mathbf{x} and linear in \mathbf{x} , respectively, to (through realizing covariance matrices are symmetric) obtain

$$\begin{aligned} \mathcal{J}[\mathbf{x}] = & \mathbf{x}^T ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M}) \\ & \mathbf{x} - 2 \mathbf{x}^T ((\mathbf{C}_{xx}^f)^{-1} \bar{\mathbf{x}}^f + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{y}) \\ & + \text{terms that do not contain } \mathbf{x}. \end{aligned}$$

The gradient of $\mathcal{J}[\mathbf{x}]$, with respect to the true state \mathbf{x} , is

$$\begin{aligned} \frac{d\mathcal{J}}{d\mathbf{x}} = & 2 ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M}) \mathbf{x} \\ & - 2 ((\mathbf{C}_{xx}^f)^{-1} \bar{\mathbf{x}}^f + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{y}). \end{aligned} \quad (\text{D1})$$

Through setting this gradient equal to zero, \mathcal{J} is minimized at the maximum likelihood estimate

$$\begin{aligned} \bar{\mathbf{x}}^a = & ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M})^{-1} \\ & \times ((\mathbf{C}_{xx}^f)^{-1} \bar{\mathbf{x}}^f + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{y}). \end{aligned} \quad (\text{D2})$$

To rewrite this into the expression for the best estimate of the mean state $\bar{\mathbf{x}}^a$ in eq. (C6) first add and subtract $\mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M} \bar{\mathbf{x}}^f$ in the second factor of eq. (D2)

$$\begin{aligned} \bar{\mathbf{x}}^a = & ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M})^{-1} \\ & \times ((\mathbf{C}_{xx}^f)^{-1} \bar{\mathbf{x}}^f + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M} \bar{\mathbf{x}}^f + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{y} \\ & - \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M} \bar{\mathbf{x}}^f) \end{aligned}$$

and regroup terms to isolate the first factor of eq. (D2) in front of $\bar{\mathbf{x}}^f$ to obtain

$$\begin{aligned} \bar{\mathbf{x}}^a = & ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M})^{-1} \\ & \times ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M}) \bar{\mathbf{x}}^f + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} (\mathbf{y} - \mathbf{M} \bar{\mathbf{x}}^f), \end{aligned}$$

which is equal to

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M})^{-1} \mathbf{M}^T (\mathbf{C}_{yy})^{-1} (\mathbf{y} - \mathbf{M} \bar{\mathbf{x}}^f).$$

To arrive at the Kalman equation for the mean (eq. C6) the factor in front of the innovation has to be shown to be equal to the Kalman gain in eq. (5), that is,

$$\begin{aligned} & ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M})^{-1} \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \\ & = \mathbf{C}_{xx}^f \mathbf{M}^T (\mathbf{M} \mathbf{C}_{xx}^f \mathbf{M}^T + \mathbf{C}_{yy})^{-1}. \end{aligned} \quad (\text{D3})$$

By multiplying both sides from the right by $\mathbf{M} \mathbf{C}_{xx}^f \mathbf{M}^T + \mathbf{C}_{yy}$ and realizing that the left-hand side can be rewritten through

$$\mathbf{M}^T (\mathbf{C}_{yy})^{-1} (\mathbf{M} \mathbf{C}_{xx}^f \mathbf{M}^T + \mathbf{C}_{yy}) = (\mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M} + (\mathbf{C}_{xx}^f)^{-1}) \mathbf{C}_{xx}^f \mathbf{M}^T,$$

both sides come down to $\mathbf{C}_{xx}^f \mathbf{M}^T$. This proves that equation (D3) holds. While deriving the maximum likelihood estimate in equation (D2), we also obtained an estimate for the error covariance matrix \mathbf{C}_{xx}^a within the Gaussian posterior of eq. (C5)

$$\begin{aligned} \mathcal{J}[\mathbf{x}] = & \mathcal{J}[\bar{\mathbf{x}}^a] + (\mathbf{x} - \bar{\mathbf{x}}^a)^T ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M}) \\ & \times (\mathbf{x} - \bar{\mathbf{x}}^a). \end{aligned} \quad (\text{D4})$$

The covariance matrix here thus has to be shown to be the same as \mathbf{C}_{xx}^a in the Kalman filter equation (eq. C7), that is,

$$\mathbf{C}_{xx}^f - \mathbf{K} \mathbf{M} \mathbf{C}_{xx}^f = ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M})^{-1}.$$

This is proven to hold by inserting the Kalman gain \mathbf{K} in the form of the left-hand side of eq. (D3) into the second part of the left-hand side and rewriting it as

$$\begin{aligned} \mathbf{K} \mathbf{M} \mathbf{C}_{xx}^f = & ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M})^{-1} \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M} \mathbf{C}_{xx}^f \\ = & ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M})^{-1} \\ & \times (\mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M} + (\mathbf{C}_{xx}^f)^{-1} - (\mathbf{C}_{xx}^f)^{-1}) \mathbf{C}_{xx}^f \\ = & ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M})^{-1} \\ & \times [(\mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M} + (\mathbf{C}_{xx}^f)^{-1}) \mathbf{C}_{xx}^f - \mathbf{I}] \\ = & \mathbf{C}_{xx}^f - ((\mathbf{C}_{xx}^f)^{-1} + \mathbf{M}^T (\mathbf{C}_{yy})^{-1} \mathbf{M})^{-1}. \end{aligned}$$

APPENDIX E: TARGET NEAR DOWNDIP LIMIT

Assimilation results for a single targeted point at the fault were shown for a random point located in the centre of the seismogenic zone (Figs 4, 5, 8 and 9). Varying the location of this point along the seismogenic zone influences the visual goodness of fit and interpretation of the results somewhat. Although Figs 3 and 6 show shear stress updates and covariances for all nodal points, for completeness we also show results for a random point located near the downdip limit of the seismogenic zone. Near to that point most events nucleate, which makes this an important albeit sensitive location. Minor events, which do not affect near-surface data or stresses far from its location, occur relatively frequently, thereby making it difficult to capture all events exactly. The true solution illustrates this as the strength excess is often near zero within the interseismic period (Fig. E2f) indicating a near critically stressed fault region. Fig. E2 shows that stresses and strengths are still captured fairly well. Surprisingly enough, they even seem to capture some small events occasionally (Fig. E2). In terms of contributions to the update, Fig. E1 shows that (at least for this first assimilation step) horizontal velocity data is followed more, as a larger correlation with that exists at this downdip position (i.e. lower model error).

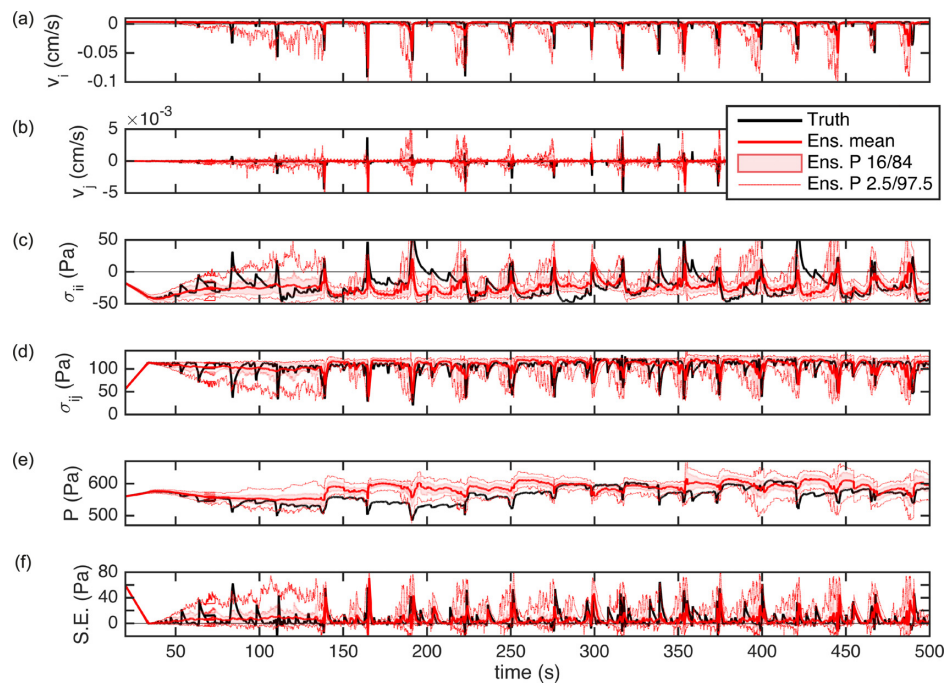


Figure E1. Data assimilation updates of shear stress at a single point at the fault near the downdip limit of the seismogenic zone. Data complexity is step by step increased from (a) one observation of the true horizontal velocity without data error, (a) one noisy observation of the horizontal velocity, including corresponding data error, (c) five observations of true velocities, stresses and pressure without data error, to (d) five noisy observations with data error (reference setting used throughout this paper). For further explanation see Fig. 4.

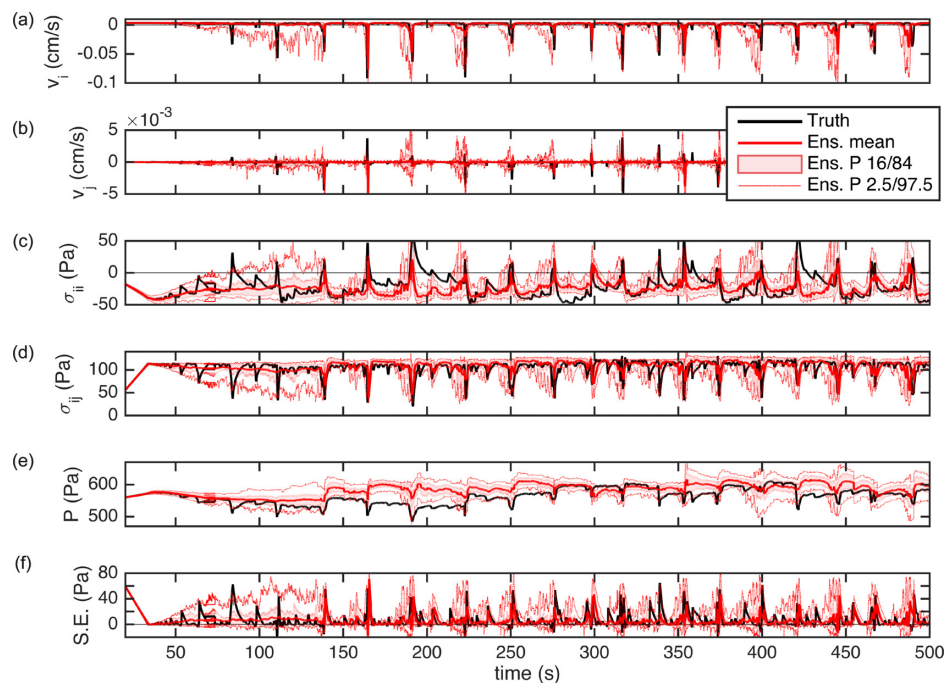


Figure E2. Temporal evolution of (a–e) all state types and (f) strength excess (S.E.) at a random fault marker near the downdip limit of the seismogenic zone. The black lines indicate the truth, while the red lines represent the probabilistic estimate of the resulting ensemble through percentiles P. For further explanations see Fig. 8.