



## A multi-data stream assimilation framework for the assessment of volcanic unrest

Patricia M. Gregg <sup>\*</sup>, J. Cory Pettijohn

Dept. of Geology, University of Illinois—Urbana-Champaign, 152 Computer Applications Building, 605 E. Springfield Ave., Champaign, IL 61820, United States



### ARTICLE INFO

#### Article history:

Received 2 August 2015

Accepted 9 November 2015

Available online 01 December 2015

#### Keywords:

Data assimilation

Ensemble Kalman Filter

Volcano deformation

InSAR

GPS

Finite element model

### ABSTRACT

Active volcanoes pose a constant risk to populations living in their vicinity. Significant effort has been spent to increase monitoring and data collection campaigns to mitigate potential volcano disasters. To utilize these datasets to their fullest extent, a new generation of model-data fusion techniques is required that combine multiple, disparate observations of volcanic activity with cutting-edge modeling techniques to provide efficient assessment of volcanic unrest. The purpose of this paper is to develop a data assimilation framework for volcano applications. Specifically, the Ensemble Kalman Filter (EnKF) is adapted to assimilate GPS and InSAR data into viscoelastic, time-forward, finite element models of an evolving magma system to provide model forecasts and error estimations. Since the goal of this investigation is to provide a methodological framework, our efforts are focused on theoretical development and synthetic tests to illustrate the effectiveness of the EnKF and its applicability in physical volcanology. The synthetic tests provide two critical results: (1) a proof of concept for using the EnKF for multi dataset assimilation in investigations of volcanic activity; and (2) the comparison of spatially limited, but temporally dense, GPS data with temporally limited InSAR observations for evaluating magma chamber dynamics during periods of volcanic unrest. Results indicate that the temporally dense information provided by GPS observations results in faster convergence and more accurate model predictions. However, most importantly, the synthetic tests illustrate that the EnKF is able to swiftly respond to data updates by changing the model forecast trajectory to match incoming observations. The synthetic results demonstrate a great potential for utilizing the EnKF model-data fusion method to assess volcanic unrest and provide model forecasts. The development of these new techniques provides: (1) a framework for future applications of rapid data assimilation and model development during volcanic crises; (2) a method for hind-casting to investigate previous volcanic eruptions, including potential eruption triggering mechanisms and precursors; and (3) an approach for optimizing survey designs for future data collection campaigns at active volcanic systems.

© 2015 Elsevier B.V. All rights reserved.

### 1. Introduction

In the past decades, the evaluation of volcanic activity has been greatly enhanced by the coupling of remote (e.g., satellite geodetic and global seismic arrays) and local (e.g., GPS, tiltmeters, gas emissions, campaign gravity, and seismometers) observations. Surface deformation observations, in particular, provide critical information about the evolution of a magmatic system and may provide early warning indicators of imminent eruption and the evolution of a magma system during a volcano crisis (Segall, 2013; Biggs et al., 2014; Cannava et al., 2015). Concurrently, thermo-mechanical models of magma reservoirs have significantly advanced our understanding of eruption-triggering mechanisms beyond the temporal and spatial limitations of our observations (Currenti et al., 2007, 2008; Currenti et al., 2011a; Currenti et al., 2011b; Currenti et al., 2012; Gerbault, 2012; Gerbault et al., 2012; Gregg et al., 2013; Currenti and Williams, 2014). This new generation of models

has pushed the field beyond the classic analytical approaches (Mogi, 1958; McTigue, 1987), providing new insights into the mechanics of magma chamber growth and failure. However, combining these recent numerical advancements with increasingly sophisticated observations has proven to be a difficult task. A robust model-data fusion framework is necessary to best utilize the potential of both the volcano observations and geodynamic models.

A classic approach to analyzing volcano observations is to develop inversions and joint inversions of datasets using analytical solutions such as the elastic Mogi model (Massonnet and Feigl, 1998; Battaglia et al., 2003a; Battaglia et al., 2003b; Newman et al., 2012; Parks et al., 2015). Optimization schemes for finite element models can provide similar outcomes for inverting datasets (Carbone et al., 2007; Hickey et al., 2013). While these approaches work well for combining models with one or potentially two data streams, they are static assessments of the system state and do not provide updates or forecasts. For the problem of assessing volcano hazards and combining models with observations in real time, sequential data assimilation techniques are required. The Development of time-sequential data assimilation

\* Corresponding author.

E-mail address: [pgregg@illinois.edu](mailto:pgregg@illinois.edu) (P.M. Gregg).

techniques is therefore essential for the field of volcanology to make significant advancements in volcano monitoring and hazard assessment, toward the eventual goal of eruption forecasting.

The Kalman filter sequential data assimilation method has been utilized to develop techniques for inverting surface deformation data to investigate transient signals along faults and at active volcanoes (Segall and Matthews, 1997; McGuire and Segall, 2003; Miyazaki et al., 2003, 2011; Fukuda et al., 2004; Murray and Segall, 2005; Ohtani et al., 2010). These investigations have shown the ability to combine time variant data sets with forward models. However, the classic Kalman filter approach utilizes linearized equations that result in instabilities for non-linear application (Schmidt, 1966). As such, they are likely inappropriate for use in volcanic systems which may behave highly nonlinearly, especially during volcano eruption. The Extended Kalman Filter attempts to circumvent the linearization issues by calculating partial derivatives at each time step to linearize any nonlinear portions of a dynamic system. While this approach has been used in analyses of Global Positioning System (GPS) data to provide models of fault slip (e.g., McGuire and Segall, 2003), these formulations are nontransferable from one application to another. As such, Extended Kalman Filter formulation would need to be re-derived for any new volcano applications. Furthermore, the Extended Kalman filter approach is likely intractable for assimilating multiple disparate data streams into complicated finite element models. An additional limitation to the Kalman filter and Extended Kalman filter approaches is that they require storage and forward propagation of a cumbersome covariance matrix making them computationally expensive (Evensen, 1994; Grewal and Andrews, 2008). The Extended Kalman filter is particularly computationally costly because the linearized partial derivatives must be calculated at each time step, precluding the ability to perform calculations off-line to save time and storage space (Julier et al., 2000).

The Ensemble Kalman Filter (EnKF) was developed to address the prohibitive computational costs and linearization issues inherent to the Kalman filter and Extended Kalman filter approaches (Evensen, 1994). The EnKF method uses a Markov chain Monte Carlo (MCMC) approach to estimate the covariance matrix in the Kalman filter and also tracks model state variables and parameters concurrently, all at a fraction of the computational cost (Evensen, 1992, 1994, 2009a, 2009b). While the EnKF method is widely utilized due to its relatively straightforward implementation, interpretation, forecasting capabilities, and usability with nonlinear models (Evensen, 2009a), it has yet to be applied to volcanic investigations. The availability of comprehensive geophysical datasets and geodynamic models of magma chamber evolution, as well as nearly two decades of successful EnKF applications to other diverse, nonlinear physical systems, provides a unique opportunity to apply this robust model-data fusion approach to volcano monitoring.

In this investigation we adapt the EnKF to assimilate ground deformation data into finite element geodynamic models of an evolving magma chamber. As the EnKF approach has never been used for this particular application, our focus is on theoretical formulation and synthetic testing to illustrate the utility of this method for future investigations of volcanic unrest. We begin by reviewing the Kalman filter and Extended Kalman filter, which have been utilized in previous geophysical investigations of ground deformation (e.g., Segall and Matthews, 1997; McGuire and Segall, 2003). Discussing these foundational approaches in model-data fusion provides context for the EnKF advancement. We then present a series of numerical experiments utilizing synthetic GPS and InSAR data for three hypothetical volcanic systems. The EnKF is implemented to assimilate the synthetic data into viscoelastic finite element models to test its performance on single and combined datasets. In particular, temporally rich GPS data are compared to spatially dense and temporally poor InSAR to investigate which data stream provides better information for model forecasts. The results from the synthetic experiments establish a foundation for future applications of the EnKF to natural systems and datasets.

## 2. Model-data fusion

A simple approach to calibrating and testing a model of a dynamic system is to compare model outputs with specific datasets, update parameter values, and then iterate to achieve a better fit. Model-data fusion takes model evaluation a step further by combining observations with a model while utilizing the uncertainties of both. The primary objective of model-data fusion is to use data to improve a model's performance by adjusting unknown model parameters and/or state variables and then correcting a model's trajectory.

In geophysical applications, a classic approach to model-data fusion is solving the linear inverse problem:

$$Hx = d \quad (1)$$

where  $d$  is a vector of data observations,  $x$  is the system or model being used to describe the data and  $H$  is a sensitivity matrix relating the model to the data. From here, one can solve for the system state,  $x$ , by inverting such that:

$$H^T Hx = H^T d \quad (2)$$

and the system state,  $x$ , can be solved for as

$$x = (H^T H)^{-1} H^T d \quad (3)$$

The Gramian matrix,  $Gr$ , is defined as:

$$Gr = H^T H \quad (4)$$

In a linear system where the Gramian matrix is non-singular (i.e., invertible), the solution of  $x$  is referred to as the least-squares solution of the overdetermined linear inversion problem. However, in systems with complex observations and dynamics, more often than not the direct least squares approach is unsolvable. As such, an entire body of research exists to solve the inverse problem.

"Statistical data assimilation" is a class of model-data fusion methods that combine models of a dynamic system with observations of the system in an effort to improve our understanding of the system. The foundation of modern data-assimilation methods is Bayes theorem, which updates the probability of a hypothesis (model) by linking it to evidence (data). One class of methods, referred to as "sequential" data assimilation, starts with a best-guess initial condition for a time-forward model that is sequentially updated as observations become available. In other words, the model is integrated forward in time and as measurements are taken they are used to reinitialize the model before the integration continues. For the problem at hand, where we want to understand the evolution of a magma system during a period of volcanic unrest, sequential data assimilation methods are highly suitable.

Considered by many to be the one of the greater advancements in the history of statistical estimation theory, the Kalman filter (Kalman, 1960; Kalman and Bucy, 1961) is an extensively used sequential data assimilation tool for predicting the course of a dynamic system through time (Grewal and Andrews, 2008). The Kalman filter utilizes the uncertainties in both the model and the observations to provide an optimal linear update. There are many different variations of the Kalman filter, some with specific enhancements to deal with nonlinear dynamic systems. Because the Kalman Filter, Extended Kalman Filter, and Ensemble Kalman filter all have similarities, it is easy to get lost in the details and lose sight of why one approach may be preferable to another. As such, in the following sections we review the Kalman filter and subsequent updates to the Kalman filter approach including the Extended Kalman Filter (EKF) and the Ensemble Kalman Filter (EnKF). This background provides necessary context for why the EnKF approach is chosen for the problem of assimilating multiple datasets into finite element models to evaluate volcanic activity.

## 2.1. The Kalman Filter

The Kalman Filter is an optimization scheme for estimating the instantaneous state of a linear, dynamic system in a time-forward fashion by updating the model of the system whenever observations become available (Kalman, 1960; Kalman and Bucy, 1961). The Kalman filter is extensively utilized in the fields of signal processing, physical oceanography, and climate modeling, among many others. In geophysics, the Kalman filter has been previously implemented for effectively analyzing Global Positioning System (GPS) observations to investigate ground surface displacement at slipping faults and at active volcanoes (Segall and Matthews, 1997; Larson et al., 2001). The full Kalman filter has been derived and discussed in a number of books and publications. For an excellent undergraduate level textbook on the Kalman Filter and its variations see: Grewal and Andrews (2008). However, to compare the Kalman filter to its subsequent enhancements, we provide the basic formulation below in a succinct overview of the Kalman filter approach.

The Kalman filter scheme begins at time,  $t_k$ , when a measurement,  $d$ , is made which provides information that can be applied to update the state of a model,  $x$ . As posed in Eq. (1), linearity is assumed to relate the state of the model with the collected data. Following the notation of Evensen (1992; 2009), a state estimate vector for the model,  $\psi_k^t$ , and a vector of measurements,  $d_k$ , are given at time,  $t_k$ . The underlying true system state,  $\psi^t$ , includes the vector of unbiased model system noise,  $q_k$ . If the true state,  $\psi^t(x)$  is discretized, it can be represented by the state vector,  $\psi^t$ , where it is assumed that the true state of the system evolves in time according to a dynamic model:

$$\psi_k^t = \Gamma\psi_{k-1}^t + q_{k-1} \quad (5)$$

$$\psi_0^t = \Psi_0 + v \quad (6)$$

where  $\Gamma$  is the linear model operator (matrix) between each time step, and  $\Psi_0$  is the initial condition with error,  $v$  (Table 1). The model error at each time step,  $q$ , takes into account the errors associated with the model setup, numerical approximations, and boundary conditions. It then follows that the forecasted model state,  $\psi^f$ , at time  $t_k$  is related to the update of the analyzed model,  $\psi^a$ , such that:

$$\psi_k^f = \Gamma\psi_{k-1}^a \quad (7)$$

$$\psi_0^a = \Psi_0 \quad (8)$$

In other words, given a best estimate,  $\psi^a$ , for the model state,  $\psi$ , at time  $t_{k-1}$ , a forecast,  $\psi^f$ , is calculated at time  $t_k$  using Eq. (7).

The error covariance matrix for the forecast,  $C_k^f$ , at time  $t_k$  is then defined as:

$$C_k^f = \overline{(\psi_k^f - \psi_k^t)^2} \quad (9)$$

the error covariance for the model state is defined as:

$$C_k^a = \overline{(\psi_{k-1}^a - \psi_{k-1}^t)^2} \quad (10)$$

and the model error covariance is defined as:

$$Q_k = \overline{q_{k-1}^2} \quad (11)$$

The difference between the true and forecast model state,  $\psi^t - \psi^f$ , is found by subtracting Eq. (7) from Eq. (5):

$$\psi_k^t - \psi_k^f = \Gamma(\psi_{k-1}^t - \psi_{k-1}^a) + q_{k-1} \quad (12)$$

**Table 1**  
Notations.

Symbol	Description
$a$	Magma chamber horizontal radius
$A$	Model ensemble matrix
$A^a$	EnKF analysis matrix
$b$	Magma chamber vertical radius
$C$	Covariance matrix
$C^f$	Covariance for the model forecast
$C^a$	Covariance for the analyzed model state
$CBP$	Chamber boundary pressure
$d$	Data, observations
$D$	Depth to magma chamber center
$D$	Data matrix
$e$	Measurement noise vector
$E$	Young's modulus
$G$	Shear modulus
$Gr$	Gramian Matrix
$H$	Sensitivity matrix relating the model to data
$K$	Kalman Gain
$M$	Measurement matrix
$P$	Pressure
$P_L$	Lithostatic pressure
$q$	Model noise
$Q$	Model error covariance matrix
$R$	Measurement covariance matrix
$t$	Time
$v$	Initial condition error
$V$	Volume
$x$	The model describing the data
$X$	Ensemble covariance matrix
$Y$	Ensemble perturbation matrix
$\Gamma$	Matrix operator between model time steps
$\eta$	Viscosity
$\mu$	Population mean
$v$	Poisson's ratio
$\sigma$	Stress
$\sigma_s$	Population standard deviation
$\psi^a$	Model analysis
$\psi^f$	Model forecast
$\psi^t$	The true system state
$\Psi_0$	Initial condition at $t = 0$

From this point the forecast covariance equation is derived as follows:

$$C_k^f = \Gamma \overline{(\psi_{k-1}^t - \psi_{k-1}^a)^2} \Gamma^T + \overline{q_{k-1}^2} + 2\Gamma \overline{(\psi_{k-1}^t - \psi_{k-1}^a)q_{k-1}} = \Gamma C_{k-1}^a \Gamma^T + Q_{k-1} \quad (13)$$

The measurement vector,  $d$ , is related to the true state vector,  $\psi^t$ , using the measurement matrix,  $M$  and a vector containing the unbiased measurement noise,  $e$ , at time  $t_k$  as:

$$d_k = M\psi_k^t + e_k \quad (14)$$

Any measurements that can be related to the state variables by the linear operator  $M$  can be used in this assimilation scheme. In this investigation, GPS and InSAR data are used, but any dataset (e.g., gravity, topography, and electromagnetic data) that provides an observation related to the system dynamics can be assimilated using this approach. The measurement error covariance matrix is then defined as:

$$R_k = \overline{ee^T} \quad (15)$$

The optimal estimate of the model state given the available data is provided by the inverse estimate  $\psi^a$  as follows:

$$\psi_k^a = \psi_k^f + K_k(d_k - M_k\psi_k^f) \quad (16)$$

where the Kalman gain at time  $t_k$ , is:

$$K_k = C_k^f M_k^T \left[ M_k C_k^f M_k^T + R_k \right]^{-1} \quad (17)$$

Thus, the optimal estimate of the standard Kalman filter is given by the analysis of:

$$\psi_k^a = \psi_k^f + C_k^f M_k^T \left[ M_k C_k^f M_k^T + R_k \right]^{-1} (d_k - M_k \psi_k^f) \quad (18)$$

One of the primary computational limitations of the Kalman filter is that the error covariance matrix for the model state is stored and propagated in time. For models with a high-dimensional state vector, the Kalman filter approach becomes computationally unfeasible due to the size of the covariance matrix and significant run times. However, strategies have been developed to mitigate the computational limitations of the Kalman filter such as performing some of the calculations offline to decrease the computational load. For example, while the Kalman filter is a real-time assimilation scheme in that it estimates the current state of a system based on observations in real time, the Kalman gain does not need to be calculated at each time step. In most cases, there is very little variation (<5%) in the Kalman gain between observation times (Grewal and Andrews, 2008). As such, the same Kalman gain may be used for multiple time steps. Ultimately, the largest road-block to using a classic Kalman filter approach for the volcano problem is its assumption of linearity. To model the dynamic case of a volcano experiencing unrest, a variation of the Kalman filter that provides strategies for nonlinearity is necessary.

## 2.2. The Extended Kalman Filter

As highlighted in the previous section, the effectiveness of the Kalman filter data assimilation approach is limited by the assumption of linearity of the model. As illustrated above, the Kalman filter derives an update for the model state in terms of means and covariances, which depend on the model's linearity and assume that the underlying probability distributions are Gaussian. As many systems are nonlinear, previous researchers have investigated nonlinear updates to the Kalman filter approach. One of the most widely utilized nonlinear approaches is the Extended Kalman filter (EKF). The framework for the EKF, first proposed by S. F. Schmidt, is also referred to as the "Kalman–Schmidt Filter" (Schmidt, 1966). The EKF recognizes the nonlinearity of the model state and uses partial derivatives to approximate model covariance. The difference between the Kalman filter and EKF approach is that the EKF calculates the partial derivatives of the estimated state vector so that the full nonlinear functions are used to estimate the update. In other words, the EKF attempts to linearize nonlinear portions of the system model so that a linear Kalman filter can be applied (Julier et al., 2000).

The EKF is a commonly used data assimilation method in navigation and GPS analyses. In geophysics, the EKF has been applied successfully to estimate the spatial and temporal distribution of slip rates from GPS observations of ground motion during earthquakes (McGuire and Segall, 2003; Miyazaki et al., 2003, 2011; Murray and Segall, 2005; Miyazaki et al., 2006). Results from these investigations have shown that the "Extended Network Inversion Filter", a technique based on the EKF, provides an effective means for filtering and analyzing continuous GPS data.

While there have been many successful applications of the EKF, there are several limitations to the approach. The EKF may provide a better estimation of a nonlinear system than the classic Kalman filter, but its effective application has primarily been on problems that are nearly-linear (Gordon et al., 1993; Julier et al., 2000; Julier and Uhlmann, 2004; Grewal and Andrews, 2008). This is because the EKF is quasi-linear; the nonlinearity of the model is approximated linearly using the derived partial derivatives and the linear Kalman filter is still

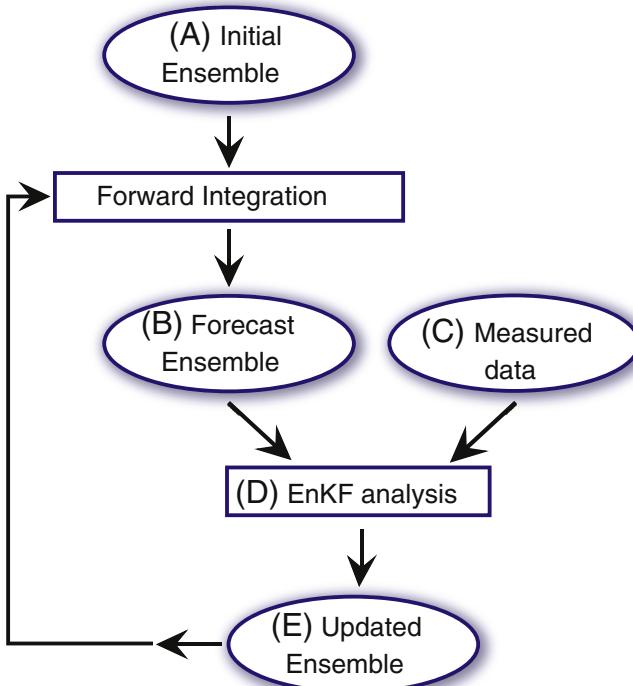
used. Furthermore, because partial derivatives must be derived explicitly for the system being modeled, the resulting EKF formulation cannot be transferred to EKF implementations on other models. As such, the effort expended for developing an EKF application is limited to the problem for which it was specifically derived. Another critical limitation to the EKF is the computational cost, which is much greater than the already computationally expensive Kalman filter (Gordon et al., 1993; Julier et al., 2000; Julier and Uhlmann, 2004). The partial derivatives must be calculated in real-time to provide the linearization estimates for the model trajectory. As such, offline computation is not possible because the implementation parameters are a function of the real-time state estimates (Grewal and Andrews, 2008). For a coupled multi-physics finite element approach, the EKF is in most cases not tractable due to the complexities of calculating the partial derivatives at each model state and their computational load.

As summarized by Julier and Uhlmann (2004), "35 years of experience in the estimation community has shown that [EKF] is difficult to implement, difficult to tune, and only reliable for systems that are almost linear on the time scale of the updates."

## 2.3. The Ensemble Kalman Filter

As summarized above, the Kalman filter method assumes that all of the probability density functions are Gaussian. The utilization of linearized equations for Kalman filter analyses results in instabilities for nonlinear applications (Evensen, 1992). Furthermore, the Kalman filter methods are computationally expensive because they require storage and forward propagation of the error covariance matrix. The Extended Kalman Filter (EKF) was developed to quasi-linearize nonlinear models so that the linear Kalman filter can be used more effectively. However, the EKF has proved a poor performer in highly nonlinear cases (Julier and Uhlmann, 2004). Moreover, Evensen (1992) discovered a closure problem in the error covariance evolution equation when using the EKF approach for nonlinear implementations. The issues with utilizing EKF for nonlinear problems motivated several efforts to find a workaround to improve the behavior of the EKF (Evensen, 1992, Evensen, 1993; Julier et al., 2000) and resulted in the development of other assimilation schemes including the Ensemble Kalman Filter (Evensen, 1994).

The Ensemble Kalman Filter (EnKF) method uses a Markov Chain Monte Carlo (MCMC) approach to estimate the covariance matrix in the Kalman filter and also tracks model state variables and parameters concurrently at a fraction of the computational cost of the Kalman filter and EKF (Evensen, 1992, 1994, 2009a, 2009b). The EnKF approach begins by taking a starting model and randomly perturbing the parameters based on *a priori* uncertainties to build an initial model ensemble (Fig. 1). The initial ensemble suite may contain hundreds of model runs or even thousands. However, tests by Evensen (1994) indicate that 100 ensembles provide a sufficient probability distribution for computing the covariance matrixes. In the volcano monitoring case, the initial ensemble comprises of a magma reservoir with parameters describing the system's geometry, chamber pressurization or volume change, and host rock rheology that are perturbed by colored or white noise to produce a suite or "ensemble" of starting models. The initial ensemble of models is forward integrated to produce a forecast ensemble (Step B in Fig. 1). In the case of an inflating magmatic source, the forecast ensemble includes the time forward calculations of deformation from a finite element model. When a time is reached when new observations are available, the measured data and data uncertainties are combined with the forecast ensemble using the EnKF analysis. Any observations that describe the system state can be assimilated in this scheme. The EnKF analysis updates the covariance matrix, which consequently updates the parameters and state variables used for the model ensemble (Step E in Fig. 1). The updated ensemble is then forward integrated until the next time step when observations are available. At each time step a new model forecast is created and the covariance matrix is



**Fig. 1.** The Ensemble Kalman Filter (EnKF) workflow, after Seiler et al. (2001). (A) The initial model, which includes basic model uncertainties. In the case of a volcanic system, the initial ensemble is the pressurized magma reservoir model calculated with a variety of initial parameter values. (B) The Forecast ensemble is derived from forward integration of the initial ensemble. The model is updated for time steps when measured data (C) is available. The measured data and data uncertainties are used in conjunction with the forecast ensemble and model uncertainties in the MCMC EnKF loop (D) to arrive at an updated ensemble (E).

updated. All previous information is discarded. However, the system's previous state is included by how it influences the evolution of the covariance matrix and model forecast.

The EnKF overcomes two of the major hurdles to the Kalman filter and Extended Kalman Filter (EKF) approaches. First, the covariance matrix is estimated from the probability distributions provided by the MCMC approach. As such, there is no need to linearize the model. Second, as mentioned above, the error covariance matrix must be stored and propagated forward in the Kalman filter, making it highly computationally expensive. For example, if the dynamic model has  $n$  unknowns, the error covariance matrix,  $C$ , has  $n^2$  unknowns, and the evolution of  $C$  in time requires  $2n$  integrations. As such, both the EKF and the Kalman filter are only practical for low-dimensional dynamic models. With EnKF, all of the required error statistics are directly calculated from the ensemble, resulting in significant computational savings.

EnKF data assimilation has been shown to be effective in the case of highly nonlinear problems and forecasting applications for a variety of fields such as hydrology, oil reservoir modeling, climatology, ecosystem modeling, and physical oceanography (vanLeeuwen and Evensen, 1996; Seiler et al., 2001; Allen et al., 2003; Bertino et al., 2003; Brusdal et al., 2003; Natvik and Evensen, 2003; Williams et al., 2005; Lisaeter et al., 2007; Skjervheim et al., 2007; Evensen, 2009a, 2009b; Wilson et al., 2010; Wilson and Özkan-Haller, 2012; Wilson et al., 2014). One of the reasons EnKF has gained significant traction is its straightforward conceptual formulation and ease of implementation. For example, the EnKF does not require the derivation of partial derivatives or adjoint equations, and does not require integration backwards in time. Practically speaking, once the EnKF framework is developed, the system model can be swapped out as modeling advancements are made without requiring recoding of the EnKF formulation. In a field such as volcanology where modeling approaches are frequently updated, this is

particularly advantageous. Furthermore, efforts by groups around the world to utilize EnKF for data assimilation have resulted in many resources and support for new EnKF applications. For all of the reasons mentioned above, the EnKF is highly suited for problem of forecasting the dynamics of an active volcano.

As of this writing, the EnKF has not yet been applied to or tested for its use for data assimilation during periods of volcanic unrest. As such, in the next sections we have constructed a synthetic experiment to test the applicability of the EnKF approach for assimilating surface deformation data into a model of an expanding magma chamber. We first provide details about our system model, which is adapted from previous finite element modeling efforts (Gregg et al., 2012; Gregg et al., 2013; Gregg et al., 2015). The synthetic data are then calculated to approximate GPS and InSAR observations at a volcano experiencing unrest. The EnKF method is then provided, including the adaptation for the volcano problem. Finally, we explore three synthetic unrest cases and test the EnKF performance to illustrate its viability for future investigations.

### 3. Methods

#### 3.1. Magma chamber finite element model

Our numerical approach builds upon previous finite element model (FEM) developments (Grosfils, 2007; Del Negro et al., 2009; Gregg et al., 2012; Gregg et al., 2013; Grosfils et al., 2015; Gregg et al., 2015). We utilize COMSOL Multiphysics 5.1 in combination with MATLAB 2014b to implement a viscoelastic FEM for an inflating magma reservoir (Fig. 2A). COMSOL is utilized for mesh development and solving for stress and strain in response to applied loads in a linear viscoelastic solid using a Maxwell model:

$$\frac{d\varepsilon}{dt} \propto \frac{\sigma}{\eta} + \frac{1}{G} \frac{d\sigma}{dt} \quad (19)$$

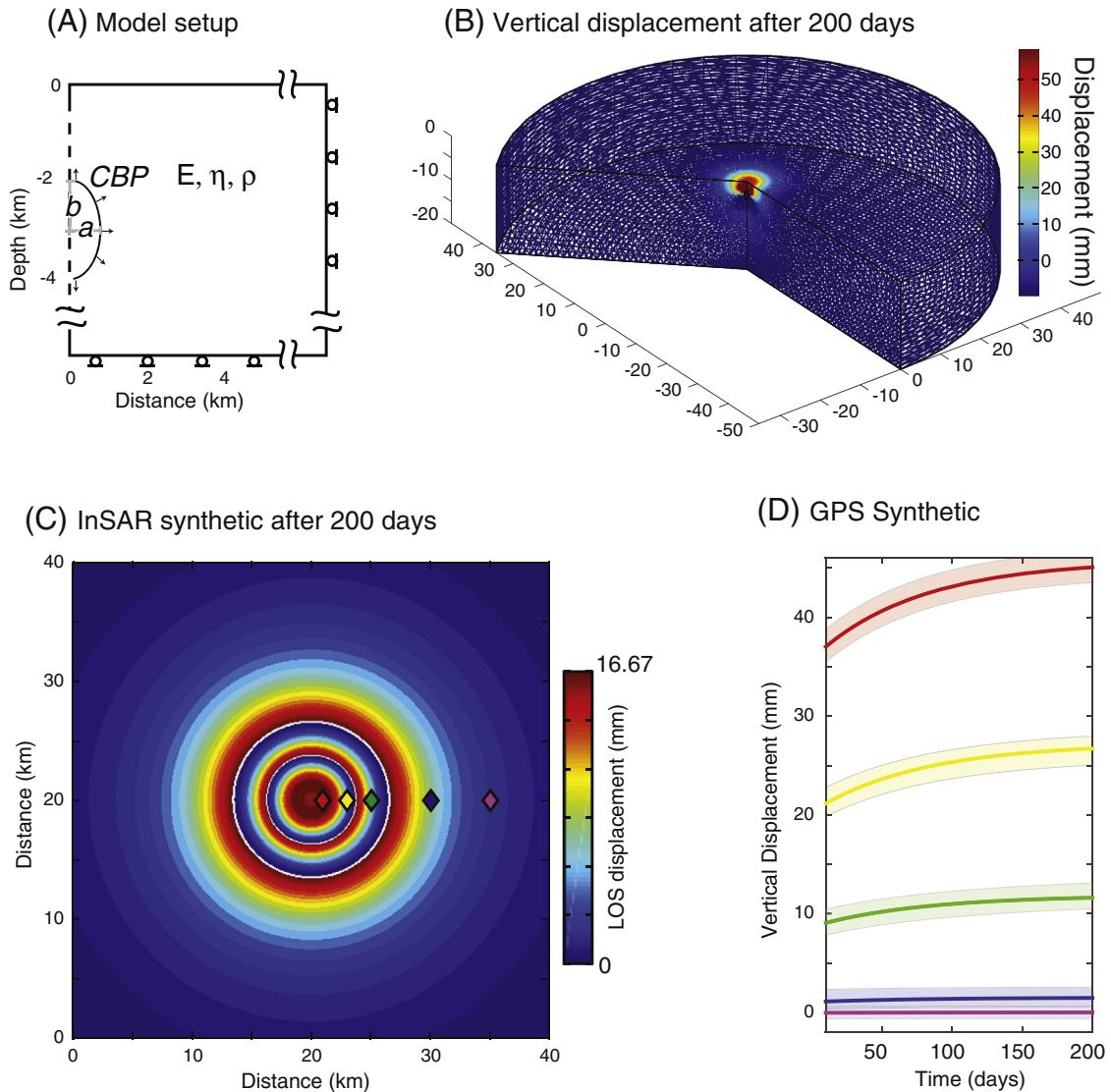
where  $\varepsilon$  is strain,  $\sigma$  is stress,  $\eta$  is viscosity, and the shear modulus,  $G$ , is a function of the Young's modulus,  $E$ :  $G = E/2(1 + v)$ , where  $v$  is Poisson's ratio. In this synthetic investigation we assume a constant Young's modulus of 60 GPa, and viscosity,  $\eta = 5 \times 10^{16}$  Pa s. Future data assimilation investigations of active volcanic systems may choose to leave Young's modulus and viscosity as unknown variables and allow the data assimilation approach to estimate the values, through the assimilation of seismic data sets for example.

The magma chamber is modeled as a pressurized void (an analog for an ideal fluid) where the geometry and position of the magma chamber are variable. The pressure applied along the chamber boundary is scaled as a function of the lithostatic or confining pressure as follows:

$$CBP = \Delta P - P_L \quad (20)$$

where the chamber boundary pressure, CBP, is equal to the change in pressure,  $\Delta P$ , minus the lithostatic confining pressure,  $P_L$ .

The reader is referred to previous studies for detailed benchmarking of the pressurized void FEM approach to volcano deformation using COMSOL for both elastic and viscoelastic rheologies (Grosfils, 2007; Del Negro et al., 2009; Gregg et al., 2012; Hickey and Gottsmann, 2014). We recognize that this is an ideal condition that is unlikely in a natural volcanic system. However, one of the advantages of utilizing an EnKF approach is that it considers that the model is imperfect and as such, provides a means for testing for gaps in our understanding of how to model the system. Furthermore, since the model is utilized outside the data assimilation analysis scheme, users of this methodology can easily swap in a preferable model as modeling advancements are made.



**Fig. 2.** Modeling approach. (A) Model setup. The left boundary condition is axial symmetry and zero displacement, roller conditions are implemented along the right and bottom boundaries. The top of the model is free to deform. The magma reservoir is a pressurized void with variable geometry. Displacement occurs due to an applied Chamber Boundary Pressure, CBP, which produces an instantaneous chamber volume change,  $\Delta V$ . The resultant  $\Delta V$  is dependent on the depth and initial size of the magma chamber. (B) The axis-symmetric 4D vertical Synthetic (Model 2 in Table 1) after 200 days. The synthetic 4D model is used to produce: (C) synthetic InSAR observations of total displacement in the line of sight (LOS) of the satellite; and (D) synthetic GPS observations of horizontal and vertical displacement at 5 GPS stations spaced from the center of the model as indicated by the diamonds on (C).

### 3.2. Synthetic Test Cases

For the synthetic test, the goal is to investigate whether the EnKF analysis scheme can estimate the parameter values assumed in the synthetic model. Synthetic data are generated using the surface deformation outputs from a pseudo-4D model of an inflating magma reservoir. Three synthetic cases were run to investigate the performance of the EnKF approach on a highly non-unique system: (1) a spherical magma chamber; (2) a vertical, prolate ellipsoid; and (3) a horizontal, oblate ellipsoid (Tables 3–5). In each case the FEM was run for 200 days to produce synthetic InSAR and GPS time series data. Fig. 2B illustrates the vertical displacement observed for the prolate ellipsoid after 200 days. InSAR data are collected every 10 days as single component, line of sight (LOS) displacement field with a resolution of 100 m and 10% standard deviation (Fig. 2C). Three component (north-south, east-west, up-down) GPS data are collected at 5 synthetic GPS stations located at distances of 1, 3, 5, 10, and 15 km from the center of the magma chamber. GPS data are collected every 24 h and have a standard deviation of 1% (Fig. 2D).

The GPS synthetics represent temporally dense data with a three-component deformation field, but with very low spatial resolution. InSAR on the other hand, provides very dense spatial information, but with only the single component total deformation field provided every 10 days.

The goal of this synthetic approach is twofold. First, we want to show how EnKF performs in these highly non-unique test cases. Second, we want to test whether temporally dense GPS is more or less “valuable” than spatially dense InSAR observations in the EnKF model updates.

### 3.3. Data assimilation using EnKF

The EnKF formulation closely follows the methods outlined by Evensen (2003) and is described below. In application, the EnKF analysis scheme begins by using a Monte Carlo approach to produce a suite of  $N$  model ensemble members,  $\psi_i$ , by perturbing  $n$  model states with white noise. Specifically, we utilize a model of a deforming magma chamber in a viscoelastic host rock and perturb

the model parameters and variables to provide the initial ensemble. The resulting model ensemble matrix,  $A$ , is an  $n \times N$  matrix, defined as:

$$A = (\psi_1, \psi_2, \dots, \psi_N) \quad (21)$$

The  $n \times n$  ensemble covariance matrix,  $X$ , is defined as:

$$X = A'(A')^T/(N-1) \quad (22)$$

The observation or data matrix is built from perturbations of discrete measurements. In the case of this investigation, the observations are provided from synthetically derived InSAR and GPS measurements of ground deformation. Given a vector  $d$  of  $m$  observations, a vector of perturbed observations can be defined as:

$$d_j = d + e_j, \quad j = 1, \dots, N \quad (23)$$

which is stored in the  $m \times N$  observation matrix:

$$D = (d_1, d_2, \dots, d_N) \quad (24)$$

The  $m \times N$  ensemble perturbation matrix, with an ensemble mean equal to zero, is defined as:

$$Y = (e_1, e_2, \dots, e_N) \quad (25)$$

The measurement error covariance matrix is expressed as:

$$C = YY^T/(N-1) \quad (26)$$

The measurement observation matrix that relates the true model state,  $\psi^t$ , to the observations  $d$  is defined by  $H$  as:

$$d = H\psi^t + e \quad (27)$$

Defining  $H$  to relate the model state to the observations is the critical step for linking observations at each time step to the correct model parameters and variables during the update analysis.

Finally, the update analysis is expressed as:

$$A^a = A + X H^T \left( H X H^T + C \right)^{-1} (D - H A) \quad (28)$$

where the  $n \times N A^a$  matrix contains the updated model state for the next model iteration. The analysis Eq. (28) is solved using Evensen (2003), Section 4.3.3 *The case with  $m \gg N$* . This approach greatly cuts down on computation, reducing the computational load from  $m^2$  to  $\sim mN$ . As the synthetic tests are pseudo-4D (2D axis-symmetric time forward models) we have run the analyses on a single processor. Tests by Evensen (1994) indicate that 100 ensembles provide a sufficient probability distribution for computing the covariance matrixes. Our own tests indicate that this number can be reduced to as few as 50 ensembles. Results presented in this study utilize 100 ensembles. Typical runs require 15–30 h depending on the number of ensembles, and length of time in the forecast (for example, we have run all of our forecasts for 100 days). However, given the nature of the MCMC approach, future efforts using EnKF analysis will take advantage of the EnKF's highly parallelizable nature and utilize cluster computing.

The EnKF approach used for investigating volcanic unrest begins by choosing a starting model for an inflating magma reservoir. The parameters derived from the starting model (Table 2) are perturbed with errors based on the uncertainty of our initial guess (i.e., an initial standard deviation of 50%). An MCMC suite of 100 ensembles is run to produce the initial ensemble matrix,  $A$  for time,  $t = 1$ . The models are forward integrated until a time at which the first set of observations becomes available. In the case of this investigation, we focus on observations of ground deformation from either GPS or InSAR. The

**Table 2**  
Starting parameters for the initial ensemble.

Parameter	Initial ensemble	
	Value	STD
$a$ (m)	500	250
$b$ (m)	500	250
$D$ (m)	4000	2000
CBP (MPa)	140	70
$\Delta V$ ( $\text{km}^3$ )	0.0016	0.0013

$a$ —horizontal radius,  $b$ —vertical radius,  $D$ —depth to center, CBP—chamber boundary pressure at the top of the chamber,  $\Delta V$ —volume increase.

observations of horizontal and vertical displacement populate the observation matrix  $D$ . At each time step where data are available, the update analysis (Eq. (28)) is calculated and a new ensemble matrix ( $A^a$ ) is produced which is then used for the next MCMC suite to calculate  $A$  for the next time step. Various strategies exist to prevent filter divergence, including inflation of the ensemble by adding noise or imposing a minimum expected error on ensemble members (Wilson et al., 2014). For this synthetic test, the standard deviation of each parameter is prevented from reducing below 10%, which has proven effective for preventing catastrophic filter divergence.

## 4. Results of the EnKF approach using synthetic data

### 4.1. Synthetic 1: spheroid

We begin our synthetic tests with an inflating spherical magma chamber, radius  $R = 750$  m, located at depth to center  $D = 3000$  m, which has expanded by  $0.0047 \text{ km}^3$  (Table 3). The EnKF analysis scheme commences by using MCMC methods to create an ensemble of 100 models varied from the starting guesses for chamber location, geometry, and volume change (Table 2). The models in the ensemble are integrated forward in time for 10 days when the first observations are available. For the first test, only InSAR data are assimilated into the EnKF analysis (Fig. 3). The InSAR data represent a dataset with infrequent observations (every 10 days), but excellent spatial coverage, in this case 100 m resolution. At day 10 the first InSAR observation of total LOS deformation is assimilated into the EnKF assimilation scheme. The ensemble predicted surface deformation fields are compared with the observed deformation and the parameters for geometry and chamber boundary pressure are updated to provide a new suite of 100 models. The new ensembles are integrated forward in time to 20 days when the next observations are available.

Fig. 3A shows how the parameter values shift toward the synthetic as new observations are assimilated. For example, the EnKF analysis initially pushes the magma chamber to deeper levels ( $D$ ), with a higher chamber boundary pressure (CBP), but then changes trajectory toward the synthetic values as it assimilates more observations. After 100 days, the estimated parameter values are approaching the synthetic values (Table 3). The best match is the value for the horizontal radius ( $a$ ), with  $\sim 4\%$  error, and the other geometrical values,  $D$  and  $b$ , are in error by  $\sim 15\%$ . CBP is not directly relatable to the synthetic since it is a function of confining pressure and therefore  $D$ . However, the total volume change is directly comparable and after 100 days of data assimilation the estimated volume changes are in error by  $\sim 35\%$ . The mean RMSE fit between the estimated surface deformation and the observed surface deformation is  $\sim 10 \text{ mm}$   $< 10\%$  error. However, the residual between the mean parameter values at 100 days and the 100 day InSAR observations produces a fit within 3 mm,  $< 2\%$  error. These results indicate that the EnKF has done a robust job of fitting the InSAR observations while dealing with the non-uniqueness of the model.

The second EnKF analysis utilizes only GPS data. The GPS data provides the three-component deformation field (north–south, east–west, up–down) with a much higher temporal resolution (every 24 h), but with a limited spatial coverage. In this case, 5 synthetic GPS

**Table 3**  
Synthetic Model 1—spherical magma chamber.

Parameter	Synthetic 1	InSAR-only after 100 d			GPS-only after 100 d			InSAR + GPS after 100 d		
	Spherical	$\mu$	$\sigma_s$	% error	$\mu$	$\sigma_s$	% error	$\mu$	$\sigma_s$	% error
$a$ (m)	750	780	80	4.0	682	65	9.1	786	78	4.8
$b$ (m)	750	638	86	14.9	620	63	17.3	663	72	11.6
$D$ (m)	3000	3436	404	15.4	3148	313	4.9	3356	352	11.9
$CBP$ (MPa)	119	148	45	—	165	46	—	113	45	—
$\Delta V$ ( $\text{km}^3$ )	0.0047	0.0064	0.0019	35.6	0.0044	0.0012	5.7	0.0044	0.0017	7.4
RMSE (mm)	—	10.6	2.7	6.4	4.4	1.8	2.7	4.7	1.5	2.9
Runtime (h)	—	18			18			18		

$a$ —horizontal radius,  $b$ —vertical radius,  $D$ —depth to center,  $CBP$ —chamber boundary pressure at the top of the chamber,  $\Delta V$ —volume increase,  $\mu$ —population mean,  $\sigma_s$ —population standard deviation, RMSE—root mean square error between modeled surface uplift and synthetic surface uplift, Runtime—indicates compute time to run the EnKF analysis to 100 days on a single processor.

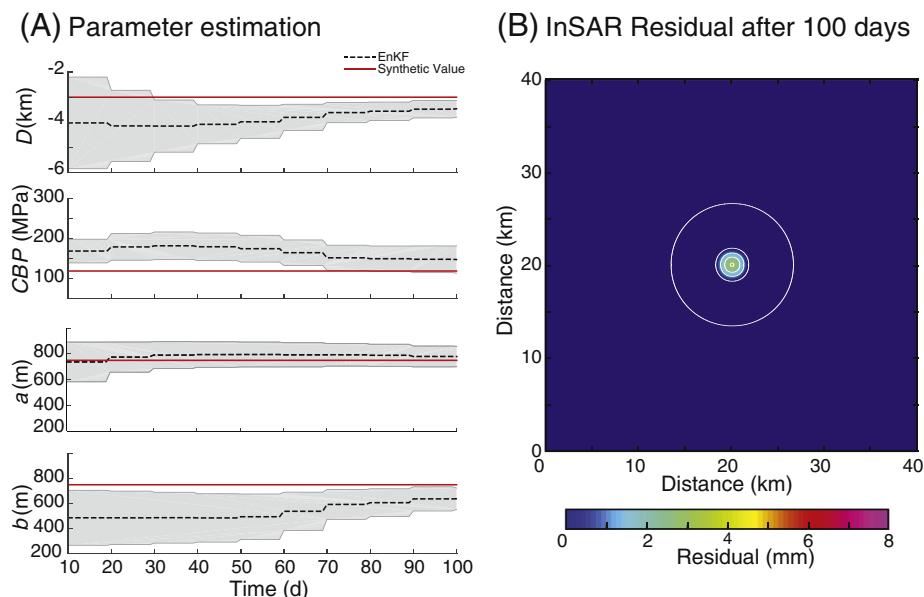
stations are spaced at 1, 3, 5, 10, and 15 km from the center of the model (Fig. 2C). The GPS observations also begin at 10 days. After 2 days of data assimilation, the parameter estimations begin to change trajectory and toward the synthetic values (Fig. 4A). The initial ensemble for  $a$  is already quite close to the actual value so holds relatively steady while the EnKF recognizes that the estimates for  $D$ ,  $CBP$ , and  $b$  are suboptimal. After an initial convergence in  $CBP$ , the EnKF also recognizes that the variation is too narrow and increases the standard deviation for the ensembles from ~15 to ~45. Alternatively, the other parameters have converged with respect to the minimum deviation of 10% set to prevent filter divergence. After 100 days of forward integration, and 90 days of assimilated GPS observations, the EnKF appears stable with model trajectories moving toward the synthetic. Aside from the estimation of the vertical radius,  $b$ , all of the parameters are estimated within 6% of the synthetic value. Fig. 4B illustrates the fit of the ensemble models calculations of surface deformation with the synthetic GPS observations at each station. Because the models are axisymmetric, the north–south and east–west components are plotted as a single radial displacement. The EnKF performs well, matching both the vertical displacement and radial displacements by adjusting model parameters.

In the final spherical EnKF analysis, both InSAR and GPS observations are assimilated. The initial model ensemble is integrated forward in

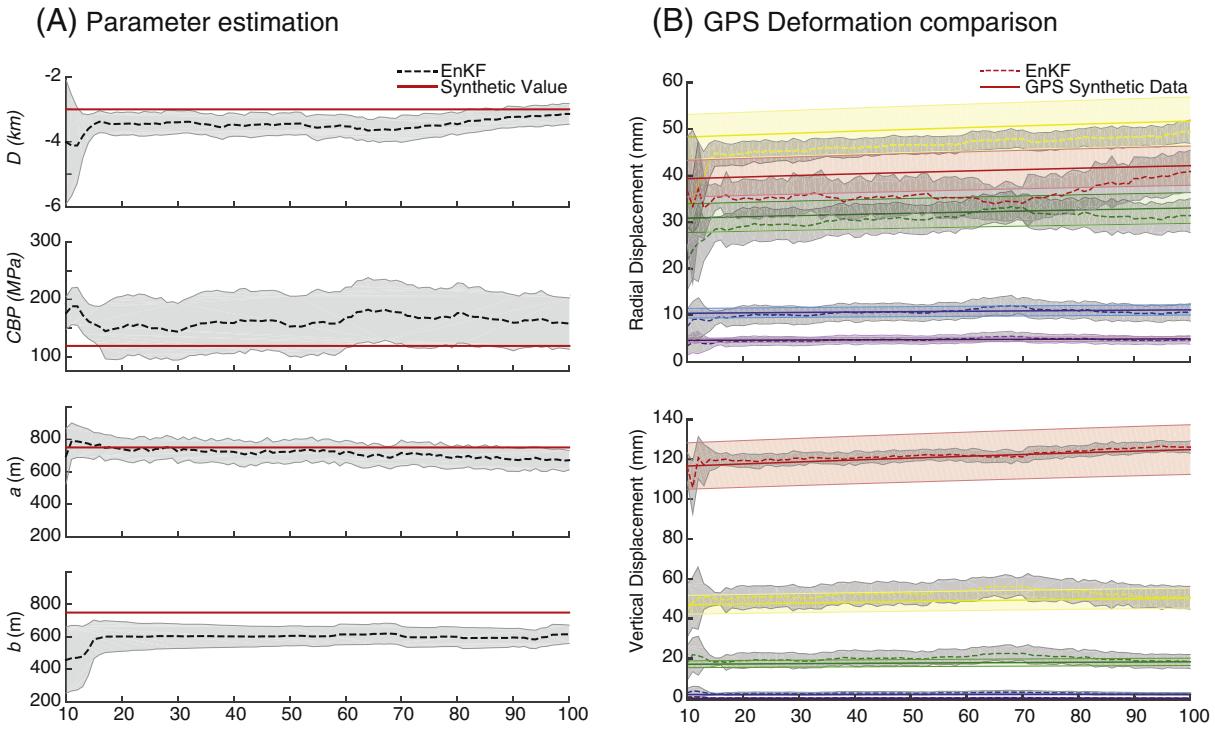
time until day 10 when both InSAR and GPS data become available. The parameters immediately begin to change their trajectories in response to the incoming observations (Fig. 5A). In this experiment we have not down-sampled the InSAR data as the EnKF weights the observations based on their estimated variances. That said, as is illustrated in Fig. 5B, the assimilation of InSAR observations results in greater errors in the predictions of surface displacement. The residual between the ensemble mean and the synthetic InSAR observation at 100 days has a less optimal fit than that of the GPS-only case (Fig. 5C). However, overall, the combined approach does a nice job of both matching the surface observations and estimating parameter values within ~10% of the synthetic values.

#### 4.2. Synthetics 2 and 3: ellipsoids

The magma chamber problem is highly non-unique as there are multiple combinations of shape, depth, and volume change that can produce the same surface deformation signal. As was illustrated in the spherical example, the EnKF can accommodate a variety of geometrical variations and appears to be able to handle the non-uniqueness of the system. In the second and third synthetic cases, surface deformation signals are produced for both a prolate and an oblate expanding spheroid



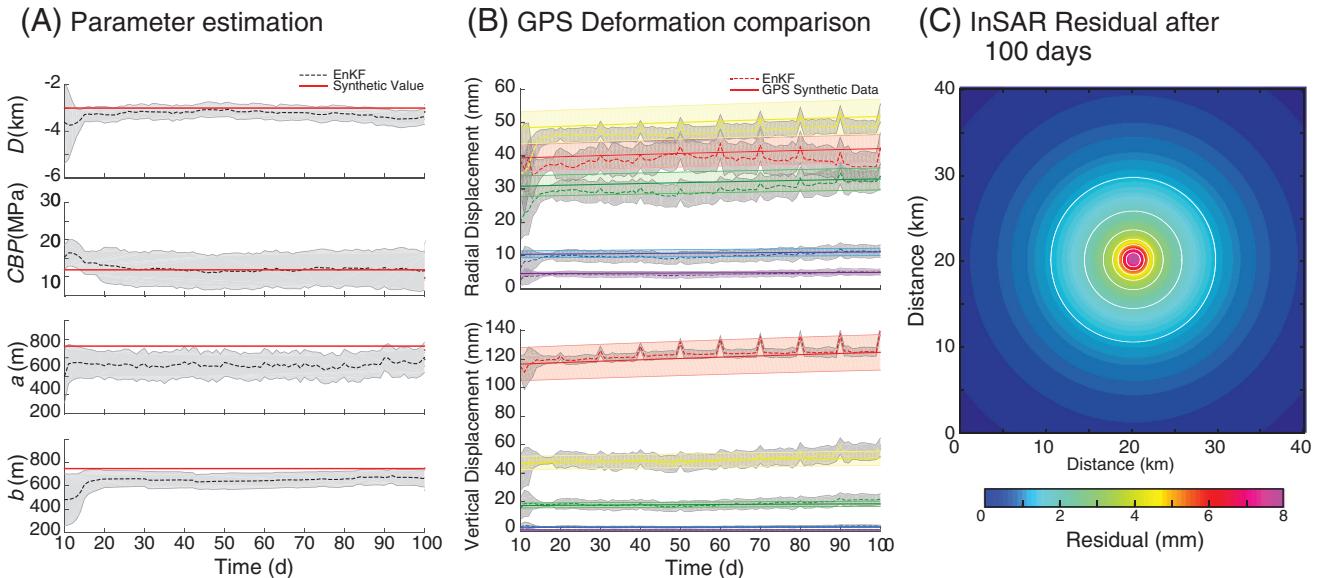
**Fig. 3.** InSAR-only data assimilation of the spherical synthetic test (Synthetic 1, Table 2). In the InSAR-only case we explore convergence with spatially dense, temporally sparse synthetic data. The forward model is run every 48 h, using 100 ensembles, and assimilation of InSAR data occurs every 10 days. The InSAR LOS deformation information has 100 m resolution and has a 10% error. (A) EnKF parameter estimation shown by black dashed line with the gray region outlining the standard deviation. The synthetic model parameters are shown by the solid red line. After 100 days of data assimilation, the EnKF analysis is providing the correct trajectory for the model convergence, but has not yet converged onto the synthetic model parameters. (B) The residual after 100 days for the mean parameter values produces a model fit to the surface observations within ~3 mm, <2% error. Contours every 1 mm. Note that the color bar scale is the same as Fig. 5A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



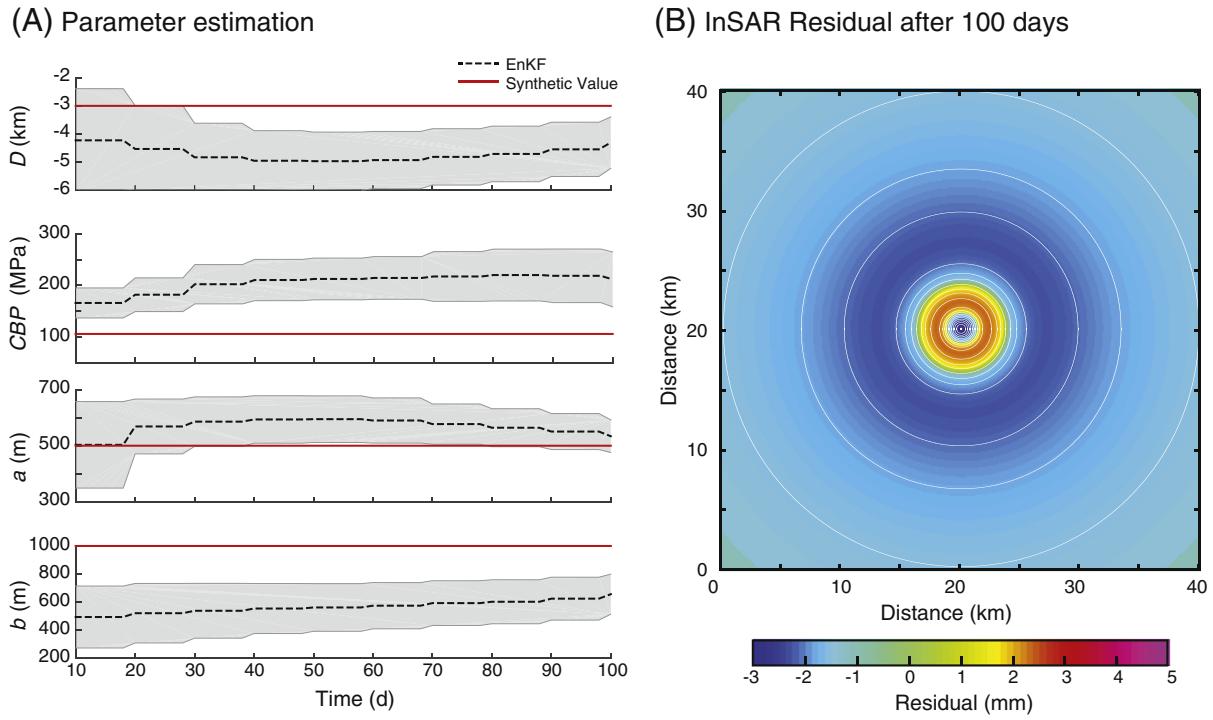
**Fig. 4.** GPS-only data assimilation for the spherical synthetic model (Synthetic Model 1 in Table 1). In the GPS-only case we explore convergence with spatially sparse, temporally dense synthetic data. The EnKF analysis uses 100 model ensembles and the GPS data are assimilated every 24 h and is assumed to have a 5% error. (A) Parameter values start to converge immediately after assimilating GPS data. After 100 days of data assimilation, the EnKF analysis is providing the correct trajectory for the model convergence, and is doing a good job of capturing the parameters used in the synthetic model. (B) Radial and vertical displacement comparison between the EnKF analysis (dashed lines, with gray region outlining the standard deviation) and the synthetic GPS data (solid line, with colored region indicating the assumed data error). After 100 days of InSAR data assimilation, there is very good convergence of the modeled deformation and observed deformation.

to see how the EnKF responds to variations in magma chamber geometry given the same starting model ensemble. Again, the approach is to first assimilate each dataset individually and then to assimilate the two data sets simultaneously.

The starting values for the initial ensemble remain the same (Table 2). For Synthetic 2: prolate spheroid, the initial guess for the horizontal radius is exactly correct. However, the vertical radius is in error by 50%. For the InSAR-only data assimilation case, the lack of observations



**Fig. 5.** Joint InSAR and GPS data assimilation for the spherical synthetic model (Synthetic Model 1 in Table 1). The EnKF analysis uses 100 model ensembles and is run every 24 h. InSAR total deformation data are assimilated every 10 days with a 10% error and GPS data are assimilated every 24 h with a 5% error. (A) After 100 days of data assimilation, the EnKF analysis is providing the correct trajectory for the model convergence, and is doing a good job of capturing the parameters used in the synthetic model. (B) Radial and vertical displacement comparison between the EnKF analysis (dashed lines, with gray region outlining the standard deviation) and the synthetic GPS data (solid line, with colored region indicating the assumed data error). InSAR data assimilation coincides with spikes in model error every 10 days due to the larger assumed observational errors with InSAR. (C) The InSAR residual after 100 days shows that the model fits the surface observations within ~8 mm, <5% error. Contours are every 1 mm. Note that the color bar scale matches Fig. 3A.



**Fig. 6.** InSAR-only data assimilation of the vertical synthetic model from Synthetic Model 2. The forward model is run every 48 h, using 100 ensembles, and assimilation of InSAR data occurs every 10 days. The InSAR LOS deformation information has 100 m resolution and has a 10% error. (A) EnKF parameter estimation shown by black dashed line with the gray region outlining the standard deviation. The synthetic model parameters are shown by the solid red line. After 100 days of data assimilation, the EnKF analysis is providing the correct trajectory for the model convergence, but has not yet converged onto the synthetic model parameters. (B) The residual after 100 days shows that the model fits the surface observations within  $\sim \pm 4$  mm, <10% error. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

results in a very slow convergence and much higher errors in the estimation of the parameters (Fig. 6A). The EnKF initially increases,  $D$ ,  $CBP$ ,  $a$ , and  $b$  in an effort to fit the surface observations while, but after assimilating 4 observations of total LOS deformation, the EnKF recognizes the incorrect trajectory of  $a$  and  $D$  and after 10 assimilations begins to decrease  $CBP$ . At the end of the 100 day run, all of the parameters are on the correct trajectory and the residual between the mean ensemble value and the observed InSAR deformation is within  $+/- 3$  mm (Fig. 6B). Even though the trajectories are correct, the overall parameter errors are still quite large after 10 InSAR assimilations (Table 4). On the other hand, the GPS-only case exhibits a very rapid convergence and after 20 days (and 10 GPS observations), all of the estimated parameters are within 10% of the synthetic value except for  $b$ , which is on the correct trajectory, but still too short (Fig. 7A). Moreover, as shown in Table 4, the EnKF estimates the magma chamber depth within 5%, as well as the volume change within 7%. After 100 days (and 90 assimilated GPS observations) there is very good fit between the synthetic and EnKF surface deformation calculations (Fig. 7B).

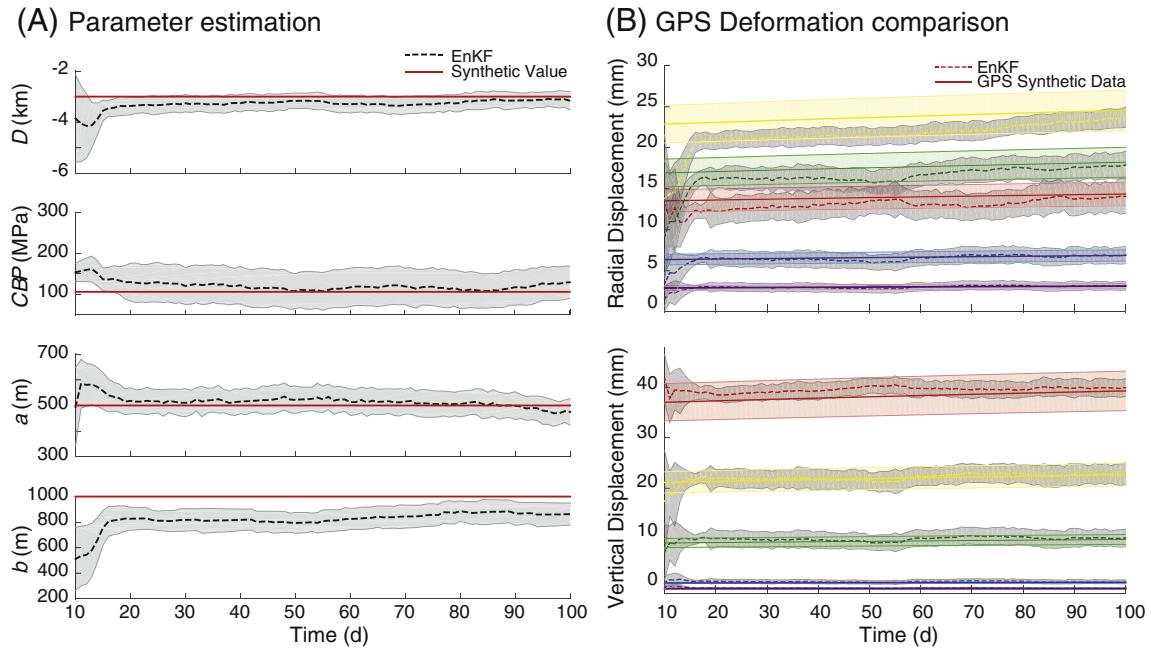
The combination of both InSAR and GPS data assimilation results in a similar convergence pattern to the GPS-only case. Except for  $b$ , all of the parameter quickly converge toward the synthetic values (Fig. 8A). However, every 10 days when InSAR data are assimilated, there is a spike in the deformation calculation (Fig. 8B). After 100 days, the deviations due to the InSAR data assimilation result in a slightly less optimal fit between the EnKF calculated deformation and the observed deformation than the GPS-only case, but a significantly better fit than the InSAR-only case. Furthermore, the residual between the mean ensemble model and the observed InSAR LOS deformation at 100 days of  $\sim 5$  mm is much less than the InSAR-only case.

Similar results were obtained for Synthetic 3: oblate spheroid (Table 5). In the oblate case the initial guess is exactly correct for the vertical radius, but is off by 50% for the horizontal radius. Unlike the prolate case, in which the EnKF took several steps to find the correct trajectory for  $b$ , the EnKF was able to more easily distinguish that the horizontal radius needed to be increased to 1000 m. Taking into consideration all three synthetic cases, it appears that vertical radius,  $b$ , is not

**Table 4**  
Synthetic Model 2—vertical magma chamber.

Parameter	Synthetic 2		InSAR only after 100 d			GPS only after 100 d			InSAR + GPS after 100 d		
	Vertical	$\mu$	$\sigma_s$	% error	$\mu$	$\sigma_s$	% error	$\mu$	$\sigma_s$	% error	
$a$ (m)	500	533	58	6.6	476	49	4.8	478	50	4.4	
$b$ (m)	1000	657	143	34.3	847	80	15.3	792	81	20.8	
$D$ (m)	3000	4309	917	43.6	3160	334	5.3	3277	345	9.2	
$CBP$ (MPa)	106	212	53	—	130	45	—	142	45	—	
$\Delta V$ ( $\text{km}^3$ )	0.00268	0.0037	0.0012	37.3	0.0025	0.0009	7.3	0.0025	0.0008	6.9	
RMSE (mm)	—	3.7	0.2	—	1.8	0.7	—	2	—	0.7	
Runtime (h)	—	18	—	—	17.7	—	—	18	—	—	

$a$ —horizontal radius,  $b$ —vertical radius,  $D$ —depth to center,  $CBP$ —chamber boundary pressure at the top of the chamber,  $\Delta V$ —volume increase,  $\mu$ —population mean,  $\sigma_s$ —population standard deviation, RMSE—root mean square error between modeled surface uplift and synthetic surface uplift, Runtime—indicates compute time to run the EnKF analysis to 100 days on a single processor.



**Fig. 7.** GPS-only data assimilation for Synthetic Model 2—a vertical magma chamber. The EnKF analysis uses 100 model ensembles and the GPS data are assimilated every 24 h and is assumed to have a 5% error. (A) After 100 days of data assimilation, the EnKF analysis is providing the correct trajectory for the model convergence, and is doing a good job of capturing the parameters used in the synthetic model. (B) Radial and vertical displacement comparison between the EnKF analysis (dashed lines, with gray region outlining the standard deviation) and the synthetic GPS data (solid line, with colored region indicating the assumed data error). After 100 days of InSAR data assimilation, there is very good convergence of the modeled deformation and observed deformation.

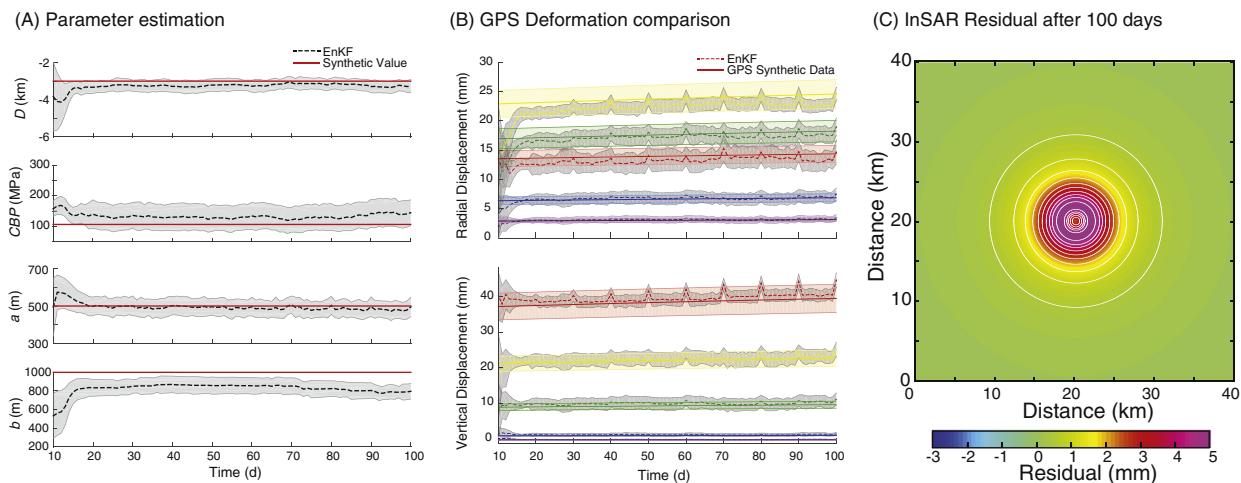
as easily resolvable as the horizontal radius,  $a$ , regardless of whether the data assimilation approach is utilizing InSAR or GPS data or both.

## 5. Discussion

### 5.1. EnKF vs. elastic inversion techniques

The classic approach to analyzing surface deformation data at an active volcano is to compute an elastic inversion using an analytical solution (Mogi, 1958; McTigue, 1987). An elastic inversion provides an estimate of the depth range, geometry (e.g., sill or sphere), and total volume change needed to produce the surface deformation signal

(Massonnet and Feigl, 1998; Battaglia et al., 2003a; Battaglia et al., 2003b; Newman et al., 2012; Parks et al., 2015). Battaglia et al. (2013) provide a software package, “dModels”, that produces elastic inversions based on either GPS or InSAR data. Using the same *a priori* values (Table 2) and rheological parameters, we have run dModels to find the optimal elastic inversion for the InSAR and GPS observations from Synthetic 1: spheroid. The best fits for both the GPS and InSAR datasets produce a source that is offset from center by ~100–200 m (Table 6). The GPS data produces the closest fit to the elastic inversion, with only a ~9% error in depth estimation and 20% error in volume change. The InSAR data inversion, which only provides a single component LOS deformation field, performs less optimally, with ~100% error in depth



**Fig. 8.** Joint InSAR and GPS data assimilation for Synthetic Model 2. The EnKF analysis uses 100 model ensembles and is run every 24 h. InSAR total deformation data are assimilated every 10 days with a 10% error and GPS data are assimilated every 24 h with a 5% error. (A) After 100 days of data assimilation, the EnKF analysis is providing the correct trajectory for the model convergence, and is doing a good job of capturing the parameters used in the synthetic model. (B) Radial and vertical displacement comparison between the EnKF analysis (dashed lines, with gray region outlining the standard deviation) and the synthetic GPS data (solid line, with colored region indicating the assumed data error). InSAR data assimilation coincides with spikes in model error every 10 days due to the larger assumed observational errors in InSAR. (C) The InSAR residual after 100 days shows that the model fits the surface observations within  $\sim \pm 2$  mm, <5% error. Note that the color bar scale is the same as Fig. 7A.

**Table 5**

Synthetic Model 3—horizontal magma chamber.

Parameter	Synthetic 3	InSAR-only after 100 d			GPS-only after 100 d			InSAR + GPS after 100 d		
	Horizontal	$\mu$	$\sigma_s$	% error	$\mu$	$\sigma_s$	% error	$\mu$	$\sigma_s$	% error
$a$ (m)	1000	1030	105	3.0	1080	107	8.0	1011	111	1.1
$b$ (m)	500	448	88	10.4	459	48	8.2	589	68	17.8
$D$ (m)	3000	3159	292	5.3	3024	304	0.8	2980	272	0.7
$CBP$ (MPa)	132	139	34	—	95	45	—	105	44	—
$\Delta V$ ( $\text{km}^3$ )	0.00767	0.0083	0.0020	8.2	0.0067	0.0032	13.3	0.0068	0.0028	11.9
$RMSE$ (mm)	—	25.1	4.9	—	11.1	3.4	—	11.7	3.3	—
<i>Runtime</i> (h)	—	19	—	—	19.5	—	—	18.5	—	—

$a$ —horizontal radius,  $b$ —vertical radius,  $D$ —depth to center,  $CBP$ —chamber boundary pressure at the top of the chamber,  $\Delta V$ —volume increase,  $\mu$ —population mean,  $\sigma_s$ —population standard deviation,  $RMSE$ —root mean square error between modeled surface uplift and synthetic surface uplift, *Runtime*—indicates compute time to run the EnKF analysis to 100 days on a single processor.

estimation and ~250% error in volume change. The InSAR data results in a dModels output that is much deeper and has a much larger volume change. Interestingly, the lower  $X^2$  for the InSAR data should suggest that this is a much better fit than the GPS inversion.

One advantage of the elastic approach is that the analytical inversion is computationally swift, and can be performed without the need for purchasing expensive software such as COMSOL or MATLAB. For monitoring agencies without access to commercial finite element software or cluster computing, the elastic inversion scheme may be the best option. One potential disadvantage is that the inversions are static and only provide a snapshot of the system state. Unlike the EnKF or other sequential data assimilation techniques, analytical inversions do not have a means for providing a model forecast to determine the dynamics of the volcanic unrest. While there are many advantages to the elastic analytical approach, numerical methods have the potential to calculate more sophisticated magma chamber dynamics, overpressurization, and eruption likelihood.

As discussed above, model ensembles for the EnKF are produced outside the data assimilation scheme. While we have used the EnKF as a sequential data assimilation approach in this investigation, it can also perform a static optimization. To use the EnKF as a static inversion, the forward integration step (Fig. 1) is removed and Eq. (28) is used on the same observations multiple times until a preferred RMSE is reached. As such, elastic models (Mogi, 1958; McTigue, 1987) can be used with the EnKF approach rather than a FEM. This static EnKF approach provides a scheme for straightforward joint inversions of multiple deformation datasets (GPS, Leveling, EDM, InSAR). Furthermore, time series datasets could be analyzed by investigating the evolution of the elastic source with time. Ultimately, one of the greatest advantages of the EnKF approach is that it is model independent. As modeling advancements are made, they can be substituted into the MCMC portion of the EnKF framework and used to create volcano forecasts from assimilation of multiple datasets.

**Table 6**  
Best-fit elastic inversion for Synthetic 1.

Parameter	Best fit to GPS		Best fit to InSAR	
	Value	% error	Value	% error
$x$ (m)	125	—	192	—
$y$ (m)	0	—	105	—
Radius (m)	500	33.33	500	33.33
$D$ (m)	2739	8.70	6147	104.90
$\Delta P$ (MPa)	28.8	—	100.8	—
$\Delta V$ ( $\text{km}^3$ )	0.0056	19.15	0.0164	248.94
$X^2_v$	0.4	—	0.1	—

Elastic inversions are carried out using dModels (Battaglia et al., 2013), for an elastic spherical source after (McTigue, 1987). The dModels input file assumes a specific radius. For simplicity we have assumed the same starting value as in the EnKF analysis.  $x$  and  $y$ —source location offset;  $D$ —depth to center;  $\Delta P$ —pressure increase (McTigue, 1987) pressure increase is not directly relatable to  $CBP$  as applied in the FEM as such % error is not calculated,  $\Delta V$ —volume increase;  $X^2_v$ — $X^2$  per degree of freedom as described in Battaglia et al. (2013).

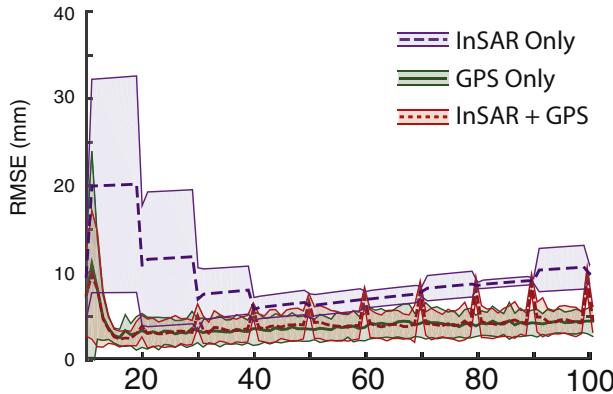
## 5.2. GPS vs. InSAR

Fig. 9 illustrates the Root Mean Squared Errors (RMSEs) between the EnKF model-predicted surface displacement and the synthetic observations. The GPS-only data assimilation cases have a consistently low RMSE after 5 days of assimilation (or the assimilation of 5 observations). Similarly, the joint InSAR and GPS case also converges rapidly producing a low RMSE after ~5 days of data assimilation. On the other hand, the InSAR-only case takes longer to converge, but has reached a minimum RMSE after assimilating 4 observations. Due to the greater observational error and single component deformation field, the InSAR-only case has a higher RMSE overall, which increases as time progresses. This increase in RMSE is also observed in the joint data assimilation case where the RMSE is minimized by the GPS observations, but then offset every 10 days by the InSAR datasets. Ultimately, when put side-by-side, the GPS-only case has done a better job of matching the surface deformation observations while estimating the parameters of the expanding magma chamber.

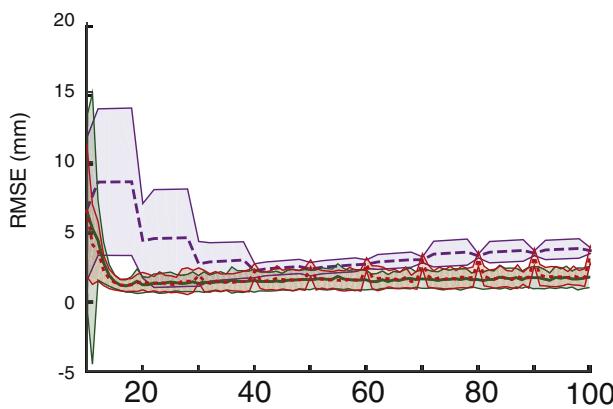
An important issue resulting from the joint data assimilation approach is the spikes in the RMSE associated with the assimilation of noisier InSAR data (Fig. 9, red signal illustrating a jump to match the InSAR RMSE). In this first EnKF implementation for volcano data, we have chosen to adapt the original EnKF formulation (Evensen, 1994). As such, the InSAR dataset, which covers a  $20 \times 20$  km area with 100 m resolution, is significantly larger than the 5 three-component GPS observations. However, strategies have been devised by other workers to resample data for multi-data stream data assimilation (Williams et al., 2009). Future efforts to further adapt the EnKF method to volcano modeling should investigate the impact of iterating between GPS stations to provide additional spatial information and down sampling InSAR data so that it does not overwhelm the signal when it is assimilated (e.g., Montgomery-Brown et al., 2015).

The synthetic results illustrate that poorer performance from InSAR data may be slightly misleading due to the simplicity of the synthetic model setup. We expect that with more complex systems the InSAR observations will be more useful for constraining magma chamber evolution. As it is, the symmetry of the system allows GPS to perform very well even though it is spatially sparse. An asymmetrical magma system will not be as non-unique as the problem investigated here. The results also show the utility of InSAR data in absence of GPS. In many cases, GPS data are simply unavailable, and while InSAR results in slower convergence, even a single data point provides important information to constrain the model. Future efforts should focus on strategies to best use InSAR including constraining observation error and devising schemes for deriving a multi-component deformation field. Furthermore, as with the static inversion, when very few InSAR time series observations are available, several EnKF iterations may be done at each time step prior to forward integration. Similarly, if InSAR observations are only available very rarely, interpolation between InSAR passes may provide additional data constraints in absence of other ground deformation signals.

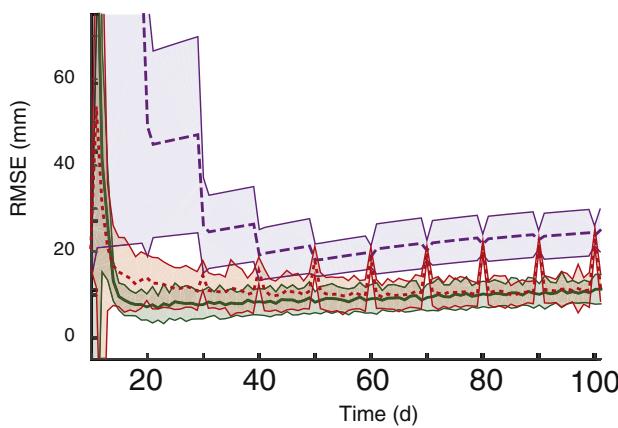
(A) Synthetic 1: Spherical



(B) Synthetic 2: Prolate



(C) Synthetic 3: Oblate



**Fig. 9.** (A) RMSE Error for the three EnKF data assimilation runs for Synthetic Model 1—the spherical synthetic case: InSAR-only data assimilation (purple dashed line with purple shading to indicate standard deviation); GPS-only data assimilation (solid green line with green shading to indicate standard deviation); Combined GPS and InSAR data assimilation (dotted red line with red shaded region to indicate standard deviation). The GPS and GPS + InSAR EnKF runs result in similar RMSE fits, with GPS-only being the best fit model. The deviation in RMSE every 10 days is due to the assimilation of the InSAR LOS total deformation data, which has a greater error than the GPS data containing the two-component deformation field (vertical and radial). (B) RMSE Error for the three EnKF data assimilation runs for Synthetic Model 2: Prolate. (C) RMSE Error for the three EnKF data assimilation runs for Synthetic Model 3: Oblate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Another advantage of initial InSAR observations of a deforming volcano is that the EnKF can be used to distinguish when model errors are resulting from poor data constraints. This is very beneficial for developing a survey plan during the onset of volcanic unrest. The broad spatial coverage provided by InSAR allows for a regional assessment of the deformation field and the opportunity to develop future data collection strategies. For example, several EnKF scenarios can be run to determine which instrument locations will provide the most significant benefit to monitoring the volcanic unrest. This is particularly useful when resources are limited. Furthermore, as data assimilation strategies are developed for other datasets (such as gravity, seismicity, tomography, and magnetotellurics) a quick EnKF assessment will provide information for optimal survey design for complimentary datasets.

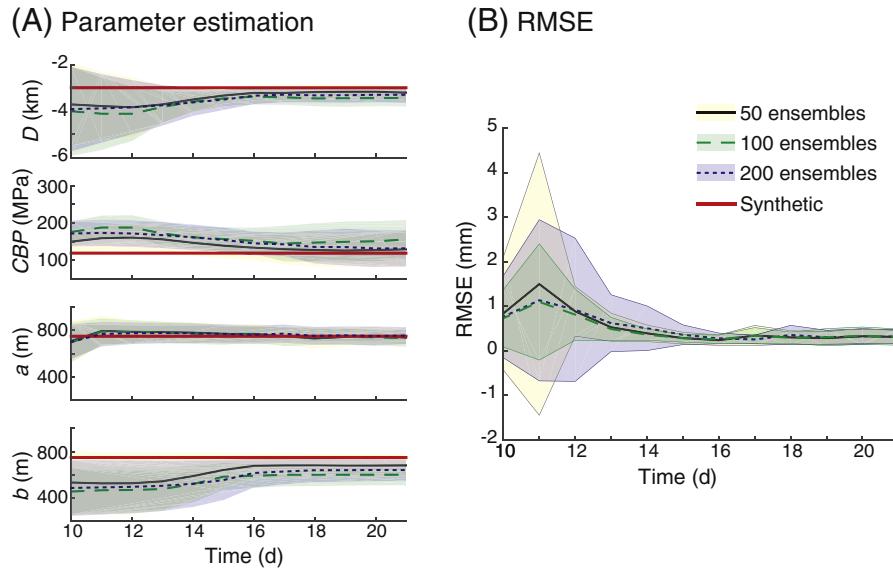
### 5.3. Potential limitations of EnKF in volcano monitoring

The goal of this investigation is to test the EnKF as a potential sequential data assimilation technique for forecasting volcanic unrest. The synthetic tests indicate that the EnKF method can be used in this regard. However, the current EnKF framework developed in this investigation is not yet practical for real time assessment of volcanic hazards. Specific roadblocks include computational run time, storage space for data and model ensembles, and the cost of licensing commercial software.

Computational run time for the EnKF analysis step (Eq. (28)) is significantly less than the Kalman filter and Extended Kalman filter approaches. The primary computational expense is the runtime of the 4D multiphysics FEMs. Specifically, hundreds of FEMs are run at each time step in the EnKF analysis scheme. Each model may take several minutes to hours to complete, depending on model size and resolution. When developing the EnKF, Evensen (1994) tested ensemble sizes in order to determine how many ensembles are needed to provide the statistics for a “healthy” covariance matrix. As the number of ensembles approaches infinity, the resolution converges to the classic Kalman filter approach. Since the number of ensembles directly affects computation time, finding a minimum that provides good performance is important. Evensen (1994) shows that 100 ensembles provides solid convergence and 500 performs better, but not significantly more so. In our tests, we compared cases with 50, 100, and 200 ensembles (Fig. 10) and do not find a significant improvement between these three ensemble sizes. Given the computation time of a 4D FEM and the lack of significant improvement in 200 ensembles, we chose to run this investigation with 100 ensembles, consistent with previous researchers (Evensen, 1994). However, future improvements including parallelization of the FEM and EnKF will allow for testing larger ensemble suites.

A pseudo-4D COMSOL FEM of an evolving magma chamber (as illustrated in the above numerical experiments) requires 1–3 GB of memory, depending on the size of the model space and resolution. After each EnKF analysis, the ensemble members are overwritten. However, local storage space must be available for the number of ensemble members in the analysis. While this is not an issue when running the EnKF on a workstation or cluster, it limits its applicability on small personal computers. As mentioned above, using an analytical approach would be much swifter and requires significantly less storage space than an FEM. However, while the elastic analytical approaches provide important analyses of deformation data, their inability to calculate magma chamber dynamics and estimate pressurization of the magma chamber limits their application for volcano forecasting.

Finally, as mentioned above, the cost of purchasing commercial software such as MATLAB and COMSOL may be prohibitive. These limitations, when taken as a whole, indicate that the largest roadblock to effectively using the EnKF for volcano forecasting is likely the COMSOL-MATLAB approach. As further advancements are made, using community developed FEM software and high-level computing programming languages such as Python will be critical for creating a data



**Fig. 10.** RMSE comparison of Synthetic 1: Spheroid, with 50 ensembles—black solid line (mean) and yellow shaded region (standard deviation), 100 ensembles—green dashed line with green shaded region, and 200 ensembles—blue dotted line with blue shaded region. (A) Parameter estimation for each of the cases is fairly consistent. The 50 ensemble case has slightly higher spread in the estimated parameter values, but all three are converging at a similar rate onto the synthetic value (red solid line). (B) The RMSE convergence is similar between the three ensemble sizes. Increasing the ensemble size results in a tighter spread in the RMSE, but very similar convergence rates. While this simple synthetic investigation suggests that 50 ensembles may be adequate, previous investigations suggest that 100 ensembles is a better minimum (Evensen, 1994; Wilson and Özkan-Haller, 2012). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

assimilation framework that is broadly transferable to the global volcano monitoring community.

## 6. Conclusions

The Ensemble Kalman Filter (EnKF) has been shown by several previous workers to be a powerful model-data fusion technique for assimilating multiple, disparate data streams into highly non-linear models (e.g., Evensen, 1994; Evensen, 2009a, 2009b). The EnKF overcomes the computational and linearity issues of the Kalman filter and Extended Kalman Filter (EKF), by using MCMC methods to calculate the error covariance matrix (Evensen, 1994). In this study, the EnKF technique provides a sequential data assimilation framework to assess volcanic unrest. The goal of this investigation is to provide a foundation for future efforts in volcano forecasting and a method for utilizing the wide range of observations collected at active volcanic systems.

To investigate the applicability of the EnKF approach, a series of numerical experiments has been run on synthetically generated observations of ground deformation. The synthetic tests for three hypothetical volcano scenarios indicate that the EnKF approach is capable of capturing the time variant dynamics of an actively deforming volcano through the assimilation of InSAR and GPS. After 100 days of data analysis, fits between model predicted surface deformation and observations are minimized to <5% relative error and parameters are estimated in most cases to <10%. These results compare very favorably to parameter estimations from an elastic analytical inversion, which resulted in errors of ~10–250%. Moreover, the use of the EnKF approach with FEMs opens the door for future efforts to calculate pressurization and failure potential of an inflating magma chamber. Furthermore, the EnKF is not model specific so the framework outlined in this investigation can be used with any preferred volcano model (e.g., analytical, elastic, visco-elastic, FEM, etc.).

Results further indicate that temporally rich GPS datasets with a three-component deformation field provide better constraints and more rapid model convergence than InSAR LOS observations. Even though the magma chambers are symmetric, and the problem is highly non-unique, the EnKF performs well with spatially sparse information. On the other hand, temporally sparse InSAR data are slower to converge

due to the lack of observational inputs, the single component LOS total deformation field, and the larger errors associated with InSAR data. However, overall, the EnKF performs robustly and indicates a high potential for successful applications to forecast volcanic unrest at natural systems.

## Acknowledgments

Utilizing the EnKF approach to investigate volcano unrest was motivated by discussions with G. Wilson and T. Özkan-Haller. We would also like to acknowledge helpful discussions with R. Denlinger, J. Pallister, and F. Amelung. This manuscript greatly benefitted from careful edits by L. Wilson. Development of data assimilation methods for monitoring active volcanoes using InSAR is funded by NASA (13-ESI13-0034).

## References

- Allen, J.I., Eknes, M., Evensen, G., 2003. An Ensemble Kalman Filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea. *Ann. Geophys.* 21, 399–411.
- Battaglia, M., Segall, P., Murray, J., Cervelli, P., Langbein, J., 2003a. The mechanics of unrest at Long Valley caldera, California: 1. Modeling the geometry of the source using GPS, leveling and two-color EDM data. *J. Volcanol. Geotherm. Res.* 127, 195–217. [http://dx.doi.org/10.1016/s0377-0273\(03\)00170-7](http://dx.doi.org/10.1016/s0377-0273(03)00170-7).
- Battaglia, M., Segall, P., Roberts, C., 2003b. The mechanics of unrest at Long Valley caldera, California. 2. Constraining the nature of the source using geodetic and micro-gravity data. *J. Volcanol. Geotherm. Res.* 127, 219–245. [http://dx.doi.org/10.1016/s0377-0273\(03\)00171-9](http://dx.doi.org/10.1016/s0377-0273(03)00171-9).
- Battaglia, M., Cervelli, P.F., Murray, J.R., 2013. dMODELS: A MATLAB software package for modeling crustal deformation near active faults and volcanic centers. *J. Volcanol. Geotherm. Res.* 254, 1–4. <http://dx.doi.org/10.1016/j.jvolgeores.2012.12.018>.
- Bertino, L., Evensen, G., Wackernagel, H., 2003. Sequential data assimilation techniques in oceanography. *Int. Stat. Rev.* 71, 223–241.
- Biggs, J., Ebmeier, S.K., Aspinall, W.P., Lu, Z., Pritchard, M.E., Sparks, R.S.J., Mather, T.A., 2014. Global link between deformation and volcanic eruption quantified by satellite imagery. *Nat. Commun.* 5. <http://dx.doi.org/10.1038/ncomms4471>.
- Brusdal, K., Brankart, J.M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P.J., Dombrowsky, E., Verron, J., 2003. A demonstration of ensemble-based assimilation methods with a layered OCCM from the perspective of operational ocean forecasting systems. *J. Mar. Syst.* 40, 253–289. [http://dx.doi.org/10.1016/s0924-7963\(03\)00021-6](http://dx.doi.org/10.1016/s0924-7963(03)00021-6).
- Cannava, F., Camacho, A.G., Gonzalez, P.J., Mattia, M., Puglisi, G., Fernandez, J., 2015. Real time tracking of magmatic intrusions by means of ground deformation modeling during volcanic crises. *Sci. Rep.* 5, 10970. <http://dx.doi.org/10.1038/srep10970>.

- Carbone, D., Currenti, G., Del Negro, C., 2007. Elastic model for the gravity and elevation changes before the 2001 eruption of Etna volcano. *Bull. Volcanol.* 69, 553–562. <http://dx.doi.org/10.1007/s00445-006-0090-5>.
- Currenti, G., Williams, C.A., 2014. Numerical modeling of deformation and stress fields around a magma chamber: constraints on failure conditions and rheology. *Phys. Earth Planet. Inter.* 226, 14–27. <http://dx.doi.org/10.1016/j.pepi.2013.11.003>.
- Currenti, G., Del Negro, C., Ganci, G., 2007. Modelling of ground deformation and gravity fields using finite element method: an application to Etna volcano. *Geophys. J. Int.* 169, 775–786. <http://dx.doi.org/10.1111/j.1365-246X.2007.03380.x>.
- Currenti, G., Del Negro, C., Ganci, G., 2008. Finite element modeling of ground deformation and gravity field at Mt. Etna. *Ann. Geophys.* 51, 105–119.
- Currenti, G., Napoli, R., Del Negro, C., 2011a. Toward a realistic deformation model of the 2008 magmatic intrusion at Etna from combined DInSAR and GPS observations. *Earth Planet. Sci. Lett.* 312, 22–27. <http://dx.doi.org/10.1016/j.epsl.2011.09.058>.
- Currenti, G., Napoli, R., Di Stefano, A., Greco, F., Del Negro, C., 2011b. 3D integrated geo-physical modeling for the 2008 magma intrusion at Etna: constraints on rheology and dike overpressure. *Phys. Earth Planet. Inter.* 185, 44–52. <http://dx.doi.org/10.1016/j.pepi.2011.01.002>.
- Currenti, G., Solaro, G., Napoli, R., Pepe, A., Bonaccorso, A., Del Negro, C., Sansosti, E., 2012. Modeling of ALOS and COSMO-SkyMed satellite data at Mt Etna: implications on relation between seismic activation of the Pernicana fault system and volcanic unrest. *Remote Sens. Environ.* 125, 64–72. <http://dx.doi.org/10.1016/j.rse.2012.07.008>.
- Del Negro, C., Currenti, G., Scandura, D., 2009. Temperature-dependent viscoelastic modeling of ground deformation: application to Etna volcano during the 1993–1997 inflation period. *Phys. Earth Planet. Inter.* 172, 299–309.
- Evensen, G., 1992. Using the extended kalman filter with multilayer quasi-geostrophic ocean model. *J. Geophys. Res. Oceans* 97, 17905–17924. <http://dx.doi.org/10.1029/92jc01972>.
- Evensen, G., 1993. Open Boundary-conditions for the Extended Kalman Filter with a quasi-geostrophic ocean model. *J. Geophys. Res.-Oceans* 98 (C9), 16529–16546. <http://dx.doi.org/10.1029/93jc01365>.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans* 99, 10143–10162. <http://dx.doi.org/10.1029/94jc00572>.
- Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.* 53, 343–367. <http://dx.doi.org/10.1007/s10236-003-0036-9>.
- Evensen, G., 2009a. *Data Assimilation: The Ensemble Kalman Filter*. Second Edition. Springer, Heidelberg.
- Evensen, G., 2009b. The Ensemble Kalman Filter for combined state and parameter estimation—Monte Carlo techniques for data assimilation in large systems. *IEEE Control. Syst. Mag.* 29, 83–104. <http://dx.doi.org/10.1109/mcs.2009.932223>.
- Fukuda, J., Higuchi, T., Miyazaki, S., Kato, T., 2004. A new approach to time-dependent inversion of geodetic data using a Monte Carlo mixture Kalman filter. *Geophys. J. Int.* 159, 17–39. <http://dx.doi.org/10.1111/j.1365-246X.2004.02383.x>.
- Gerbault, M., 2012. Pressure conditions for shear and tensile failure around a circular magma chamber; insight from elasto-plastic modelling. *Faulting, Fracturing and Igneous Intrusion in the Earth's Crust.* 367 pp. 111–130. <http://dx.doi.org/10.1144/sp367.8>.
- Gerbault, M., Cappa, F., Hassani, R., 2012. Elasto-plastic and hydromechanical models of failure around an infinitely long magma chamber. *Geochem. Geophys. Geosyst.* 13, <http://dx.doi.org/10.1029/2011gc003917>.
- Gordon, N.J., Salmond, D.J., Smith, A.F.M., 1993. Novel-approach to nonlinear non-gaussian Bayesian state estimation. *IEE Proc. F Radar Signal Process.* 140, 107–113.
- Gregg, P.M., de Silva, S.L., Grosfils, E.B., Parmigiani, J.P., 2012. Catastrophic caldera-forming eruptions: thermomechanics and implications for eruption triggering and maximum caldera dimensions on Earth. *J. Volcanol. Geotherm. Res.* 241–242, 1–12. <http://dx.doi.org/10.1016/j.jvolgeores.2012.06.009>.
- Gregg, P.M., de Silva, S.L., Grosfils, E.B., 2013. Thermomechanics of shallow magma chamber pressurization: implications for the assessment of ground deformation data at active volcanoes. *Earth Planet. Sci. Lett.* 384, 100–108. <http://dx.doi.org/10.1016/j.epsl.2013.09.040>.
- Gregg, P.M., Grosfils, E.B., de Silva, S.L., 2015. Catastrophic caldera-forming eruptions II: the subordinate role of magma buoyancy as an eruption trigger. *J. Volcanol. Geotherm. Res.* 305, 100–113. <http://dx.doi.org/10.1016/j.jvolgeores.2015.09.022>.
- Grewal, M.S., Andrews, A.P., 2008. *Kalman Filtering: Theory and Practice Using MATLAB®*, 3rd edition John Wiley & Sons, Inc., Hoboken, New Jersey.
- Grosfils, E.B., 2007. Magma reservoir failure on the terrestrial planets: assessing the importance of gravitational loading in simple elastic models. *J. Volcanol. Geotherm. Res.* 166, 47–75. <http://dx.doi.org/10.1016/j.jvolgeores.2007.06.007>.
- Grosfils, E.B., McGovern, P.J., Gregg, P.M., Galgana, G.A., Hurwitz, D.M., Long, S.M., Chestler, S., 2015. Elastic models of magma reservoir mechanics: A key tool for understanding planetary volcanism, in *Volcanism and Tectonism Across the Inner Solar System*. In: Platz, T., Massironi, M., Byrne, P., Hiesinger, H. (Eds.), Geological Society. Special Publications, London, pp. 139–158 <http://dx.doi.org/10.1144/SP401.19>.
- Hickey, J., Gottsmann, J., 2014. Benchmarking and developing numerical Finite Element models of volcanic deformation. *J. Volcanol. Geotherm. Res.* 280, 126–130. <http://dx.doi.org/10.1016/j.jvolgeores.2014.05.011>.
- Hickey, J., Gottsmann, J., del Potro, R., 2013. The large-scale surface uplift in the Altiplano-Puna region of Bolivia: a parametric study of source characteristics and crustal rheology using finite element analysis. *Geochem. Geophys. Geosyst.* 14, 540–555. <http://dx.doi.org/10.1002/ggge.20057>.
- Julier, S.J., Uhlmann, J.K., 2004. Unscented filtering and nonlinear estimation. *Proc. IEEE* 92, 401–422. <http://dx.doi.org/10.1109/jproc.2003.823141>.
- Julier, S., Uhlmann, J., Durrant-Whyte, H.F., 2000. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Autom. Control* 45, 477–482. <http://dx.doi.org/10.1109/9.847726>.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 35–45.
- Kalman, R.E., Bucy, R.S., 1961. New results in linear filtering and prediction theory. *J. Basic Eng.* 83, 95–108.
- Larson, K.M., Cervelli, P., Lisowski, M., Miklius, A., Segall, P., Owen, S., 2001. Volcano monitoring using the Global Positioning System: filtering strategies. *J. Geophys. Res. Solid Earth* 106, 19453–19464. <http://dx.doi.org/10.1029/2001jb000305>.
- Lisaeter, K.A., Evensen, G., Laxon, S., 2007. Assimilating synthetic CryoSat sea ice thickness in a coupled ice-ocean model. *J. Geophys. Res. Oceans* 112, C07023. <http://dx.doi.org/10.1029/2006jc003786>.
- Massonet, D., Feigl, K.L., 1998. Radar interferometry and its application to changes in the earth's surface. *Rev. Geophys.* 36, 441–500. <http://dx.doi.org/10.1029/97rg03139>.
- McGuire, J.J., Segall, P., 2003. Imaging of aseismic fault slip transients recorded by dense geodetic networks. *Geophys. J. Int.* 155, 778–788. <http://dx.doi.org/10.1111/j.1365-246X.2003.02022.x>.
- McTigue, D.F., 1987. Elastic stress and deformation near a finite spherical magma body: resolution of a point source paradox. *J. Geophys. Res.* 92, 15977–15990.
- Miyazaki, S., McGuire, J.J., Segall, P., 2003. A transient subduction zone slip episode in southwest Japan observed by the nationwide GPS array. *J. Geophys. Res. Solid Earth* 108, 2087. <http://dx.doi.org/10.1029/2001jb000456>.
- Miyazaki, S., Segall, P., McGuire, J.J., Kato, T., Hatanaka, Y., 2006. Spatial and temporal evolution of stress and slip rate during the 2000 Tokai slow earthquake. *J. Geophys. Res. Solid Earth* 111, B03409. <http://dx.doi.org/10.1029/2004jb003426>.
- Miyazaki, S., McGuire, J.J., Segall, P., 2011. Seismic and aseismic fault slip before and during the 2011 off the Pacific coast of Tohoku Earthquake. *Earth Planets Space* 63, 637–642. <http://dx.doi.org/10.5047/eps.2011.07.001>.
- Mogi, K., 1958. Relations of the eruptions of various volcanoes and the deformations of the ground surface around them. *Bull. Earthq. Res. Inst. Univ. Tokyo* 36, 99–134.
- Montgomery-Brown, E.K., Wicks, C.W., Cervelli, P.F., Langbein, J.O., Svarc, J.L., Shelly, D.R., Hill, D.P., Lisowski, M., 2015. Renewed inflation of Long Valley Caldera, California (2011 to 2014). *Geophys. Res. Lett.* 42, 5250–5257. <http://dx.doi.org/10.1002/2015GL064338>.
- Murray, J.R., Segall, P., 2005. Spatiotemporal evolution of a transient slip event on the San Andreas fault near Parkfield California. *J. Geophys. Res. Solid Earth* 110. <http://dx.doi.org/10.1029/2005jb003651>.
- Natvik, L.J., Evensen, G., 2003. Assimilation of ocean colour data into a biochemical model of the North Atlantic—Part 1 Data assimilation experiments. *J. Mar. Syst.* 40, 127–153. [http://dx.doi.org/10.1016/s0924-7963\(03\)00016-2](http://dx.doi.org/10.1016/s0924-7963(03)00016-2).
- Newman, A.V., Stiros, S., Feng, L.J., Psimoulis, P., Moschas, F., Saltogianni, V., Jiang, Y., Papazachos, C., Panagiotopoulos, D., Karagianni, E., Vamvakaris, D., 2012. Recent geodetic unrest at Santorini Caldera, Greece. *Geophys. Res. Lett.* 39, L06309. <http://dx.doi.org/10.1029/2012gl051286>.
- Ohtani, R., McGuire, J.J., Segall, P., 2010. Network strain filter: a new tool for monitoring and detecting transient deformation signals in GPS arrays. *J. Geophys. Res. Solid Earth* 115, B12418. <http://dx.doi.org/10.1029/2010jb007442>.
- Parks, M.M., Moore, J.D.P., Papanikolaou, X., Biggs, J., Mather, T.A., Pyle, D.M., Raptakis, C., Paradissis, D., Hooper, A., Parsons, B., Nomikou, P., 2015. From quiescence to unrest: 20 years of satellite geodetic measurements at Santorini volcano, Greece. *J. Geophys. Res. Solid Earth* 120, 1309–1328. <http://dx.doi.org/10.1002/2014jb011540>.
- Schmidt, S.F., 1966. Application of state-space methods to navigation problems. In: Leondes, C. (Ed.), *Advances in Control Systems*. Academic Press, New York, NY, pp. 293–340.
- Segall, P., 2013. *Volcano deformation and eruption forecasting. Remote Sensing of Volcanoes and Volcanic Processes: Integrating Observation and Modelling.* 380 pp. 85–106.
- Segall, P., Matthews, M., 1997. Time dependent inversion of geodetic data. *J. Geophys. Res. Solid Earth* 102, 22391–22409. <http://dx.doi.org/10.1029/97jb01795>.
- Seiler, A., Evensen, G., Skjervheim, J.A., Hove, J., Vabo, J.G., 2001. Using the EnKF for history matching and uncertainty quantification of complex reservoir models. In: Biegler, L., Biros, G., Ghattas, O., Heinkenschloss, M., Keyes, D., Mallick, B., Tenorio, L., van Bloemen Waanders, B., Wilcox, K. (Eds.), *Large-Scale Inverse Problems and Quantification of Uncertainty*. John Wiley & Sons, Ltd, Chichester, West Sussex, United Kingdom, p. 388.
- Skjervheim, J.A., Evensen, G., Aanonsen, S.I., Ruud, B.O., Johansen, T.A., 2007. Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter. *SPE J.* 12, 282–292.
- van Leeuwen, P.J., Evensen, G., 1996. Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weather Rev.* 124, 2898–2913. [http://dx.doi.org/10.1175/1520-0493\(1996\)124<2898:daaimi>2.0.co;2](http://dx.doi.org/10.1175/1520-0493(1996)124<2898:daaimi>2.0.co;2).
- Williams, M., Schwarz, P.A., Law, B.E., Irvine, J., Kurpius, M.R., 2005. An improved analysis of forest carbon dynamics using data assimilation. *Glob. Chang. Biol.* 11, 89–105. <http://dx.doi.org/10.1111/j.1365-2486.2004.00891.x>.
- Williams, M., Richardson, A.D., Reichstein, M., Stoy, P.C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D.Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C., Wang, Y.-P., 2009. Improving land surface models with FLUXNET data. *Biogeosci. Discuss.* 6, 2785–2835. <http://dx.doi.org/10.5194/bgd-6-2785-2009>.
- Wilson, G., Özkan-Haller, H.T., 2012. Ensemble-based data assimilation for estimation of river depths. *J. Atmos. Ocean. Technol.* 29, 1558–1568. <http://dx.doi.org/10.1175/jtech-d-12-00014.1>.
- Wilson, G.W., Özkan-Haller, H.T., Holman, R.A., 2010. Data assimilation and bathymetric inversion in a two-dimensional horizontal surf zone model. *J. Geophys. Res. Oceans* 115, C12057. <http://dx.doi.org/10.1029/2010jc006286>.
- Wilson, G.W., Özkan-Haller, H.T., Holman, R.A., Haller, M.C., Honegger, D.A., Chickadel, C.C., 2014. Surf zone bathymetry and circulation predictions via data assimilation of remote sensing observations. *J. Geophys. Res. Oceans* 119, 1993–2016. <http://dx.doi.org/10.1002/2013jc009213>.