
Article:
Baù, D., Ferronato, M., Gambolati, G. et al. (2 more authors) (2015) Ensemble smoothing of land subsidence measurements for reservoir geomechanical characterization. International Journal for Numerical and Analytical Methods in Geomechanics, 39 (2). 207 - 228. ISSN 0363-9061

https://doi.org/10.1002/nag.2309

Ensemble Smoothing of Land Subsidence Measurements for Reservoir Geomechanical Characterization

D. Baù[a*], M. Ferronato[b], G. Gambolati[b], P. Teatini[b] and A. Alzraiee[a]

[a] Department of Civil and Environmental Engineering, Colorado State University, 1372 Campus Delivery, Fort Collins, CO, 80523-1372, United States [Now at: University of Sheffield, Civil and Structural Engineering Department, Sheffield, S1 3JD, UK]

[b] Department of Civil, Environmental and Architectural Engineering, University of Padova, Via Trieste 63, Padova, 35121, Italy.

*Corresponding author: e-mail: domenico.bau@colostate.edu; Tel: 970-491-5387; Fax: 970-491-7727

Accepted: April 2014

Reference: Int. J. Numer. Anal. Meth. Geomech. 2015; 39:207–228. DOI: 10.1002/nag.2309

Published online 19 July 2014 in Wiley Online Library:

http://onlinelibrary.wiley.com/doi/10.1002/nag.2309/abstract

Abstract

Geomechanical models are often used to predict the impact on land surface of fluid withdrawal from deep reservoirs, as well as investigating measures for mitigation. The ability to accurately simulate surface displacements, however, is often impaired by limited information on the geomechanical parameters characterizing the geological formations of interest. In this study, we employ an Ensemble Smoother (ES), a data assimilation algorithm, to provide improved estimates of reservoir parameters through assimilation of measurements of both horizontal and vertical surface displacement into geomechanical model results. The method leverages the demonstrated potential of remote sensing techniques developed in the last decade to provide accurate displacement data for large areas of the land surface.

For evaluation purposes, the methodology is applied to the case of a disk-shaped reservoir embedded in a homogeneous, isotropic and linearly elastic half space, subject to a uniform change in fluid pressure. Multiple sources of uncertainty are investigated, including the radius, R, the thickness, h, and the depth, c, of the reservoir, the pore pressure change, Δp, porous medium's vertical uniaxial compressibility, c_M, and Poisson's ratio, v, and the ratio, s, between the compressibilities of the medium during loading and unloading cycles. Results from all simulations show that the ES has the capability to effectively reduce the uncertainty associated with those parameters to which the variability, and the spatial distribution of land surface displacements are most sensitive, namely R, c, c_M and s. These analyses demonstrate that the estimation of these parameters values depends on the number of measurements assimilated and the error assigned to the measurement values.

1 Introduction

It is widely acknowledged that anthropogenic activities involving the fluid extraction from or injection into the subsurface cause a change of pore pressure followed by a volumetric deformation of the geological formations, which may be transferred up to the land surface. A well-known consequence is land subsidence due to oil or gas production from deep reservoirs, as observed, for example, in Long Beach, California [1-3], in Venezuela [4], over the North Sea Ekofisk Field, Norway [5], and in the Northern Adriatic Sea, Italy [6-10]. Groundwater withdrawals from shallow aquifers have long been recognized as another most frequent cause of land subsidence, as documented in References [11-13].

Other examples include, but are not limited to, vapor or carbon dioxide (CO_2) injection in oil-bearing geological formations to enhance oil recovery [14-17], underground water injection to mitigate man-induced land subsidence [18], or CO_2 geological sequestration as a means to reduce greenhouse gas emissions in the atmosphere [19-21]. Of related interest is gas storage and recovery (GSR), in which natural gas transported long-distance from producing countries is stored in depleted reservoirs during periods of low demand, typically warm seasons, and recovered for consumption during periods of higher demand, such as cold seasons [22-24]. In this case, the alternation of gas injection and gas extraction produces a corresponding sequence of expansion and compaction of the reservoir, which may be observed on the land surface as a periodic-type time series.

In several of the examples above, the land surface displacement has raised serious concerns in terms of human health, structural safety of buildings, economic risk, as well as environmental and hydrologic impact. For these reasons, geomechanical models are increasingly used to assess and predict the impact of fluid extraction or injection on the ground surface. In this work, we investigate the feasibility of an inverse modeling technique

to calibrate reservoir parameters by assimilation of surface displacement data. This technique is based on an ensemble smoother (ES) algorithm, a derivate of the classic Kalman Filter (KF) [25]. The ES is a Bayesian data assimilation method that, by minimizing the variance of the estimation error, merges "prior" information from a theoretical system, i.e., the mathematical model, and field data collected from the actual system in order to produce a corrected "posterior" estimate. The ES relies upon a stochastic, or Monte Carlo, simulation, in which the system uncertainty is represented by an ensemble of realizations of surface displacements obtained by "forecast" simulations using a corresponding ensemble of reservoir parameters [26]. The update of the displacement ensemble is performed by assimilating surface displacement data collected over time. The system parameters can be simultaneously updated by incorporating them into the ensemble. Doing so allows for the local displacement data to correct not only the displacement ensemble, but also the system parameters by leveraging the displacement–parameter cross-covariance estimated in the forecast simulation.

Stochastic simulation approaches based on the KF include the ES, the Ensemble Kalman Smoother (EnKS), and the Ensemble Kalman Filter (EnKF) [27] and were originally developed in oceanographic and meteorological modeling [26, 28, 29]. In the last decade, these methods have been successfully applied to hydrologic studies [30-33] and proved to be an appealing alternative to traditional optimization-based inverse modeling approaches for the estimation of system parameters [34-39], as they typically require a lower computational effort [40]. An extensive review on the use of the EnKF in reservoir engineering applications was presented by Aanonsen et al. [41]. Of related interest to this paper are the works of Chang et al. [42], Wilschut et al. [43], Iglesias and McLaughlin [44] and Emerick and Reynolds [45]. In References [42] and [43], assimilation of land subsidence data into coupled flow and geomechanical models was performed via the EnKF method, in order to assess

reservoir parameters, such as bulk compressibility [42] and fault transmissibility [43]. Iglesias and McLaughlin [44] developed a deterministic inverse method for calibrating petrophysical and poro-elastic parameters of a coupled subsurface flow-geomechanical model using surface deformation and pore pressure data. Emerick and Reynolds [45] compared the performances of the EnKF and the ES when applied to reservoir history-matching problems, and developed an iterative ES algorithm where pressure data are assimilated multiple times with decreasing measurement error.

In contrast to the EnKF, which incorporates measurements to provide an updated system state at the current time only, the ES used in this work incorporates all previous measurements and model states to compute updated displacement fields at all previous measurement times, using the spatial and temporal covariance of model results [27,46]. Since in the EnKS and EnKF schemes system updates are made sequentially, that is, at each time measurements are collected, the update algorithm must be nested into the modeling code. On the contrary, in the ES scheme the update algorithm is run only once, using all measurement data and model states at all data collection times. As such, the ES can be applied "offline" of the model simulation, which makes it a very convenient approach for estimation of system parameters [38].

Here, the applicability of the ES scheme is investigated to provide reliable estimates of the uncertain parameters characterizing, for example, the geometry of the reservoir, its geomechanical properties, and the change in pore pressure, through assimilation of sets of horizontal and vertical surface displacement data. To analyze its potential, the methodology is applied using Geertsma's analytical solution for a disk-shaped reservoir [47] in a semi-infinite homogeneous porous medium. The low computational cost associated with this solution makes it ideal to test the capabilities of the ES as a reservoir parameter estimator. To do so, we perform and present a series of preliminary tests that anticipate the future

application of the methodology to an actual reservoir system situated within the Po plain sedimentary basin, Italy, where Interferometric Synthetic Aperture Radar (InSAR) data were collected from 2003 to 2008 [24]. Note that this system will require the use of a complex 3D poro-elasto-plastic deformation model that can account for condition of spatial heterogeneity and orthotropy of key geomechanical reservoir parameters characterizing basin-scale constitutive laws.

In view of this application, the availability of surface displacement data is here hypothesized according to the characteristics and the typical errors associated with InSAR measurement surveys. In one case, we assume that surface displacement measurements from a reference reservoir system are made at a single collection time, for example, at reservoir depletion; in another, two measurements collection times are considered, at reservoir depletion, when gas pressure is at its minimum, and after gas pressure has partially recovered. In each test, the ensemble of geomechanical simulations is rerun to determine if the updated ensemble of reservoir parameters provides a parameter structure that produces results in agreement with the measurements from the reference state, i.e., to verify if the ES scheme has been successful in calibrating the model to a reasonable degree. In addition, sensitivity analyses are performed to gain insights into the influence of (i) displacement measurement errors, and (ii) number of measurements.

2 Conceptual model of land subsidence due to gas extraction

The simulation models typically used to calculate the deformation of porous media subject to a pore pressure change stem from the theory of poro-elasticity formulated by Biot [48], who studied the stress-strain relations of permeable fluid-saturated materials by coupling together the flow equation and the Cauchy equations of equilibrium applied to the solid porous skeleton. In many instances, these equations are decoupled with the flow equation

solved first and the results transferred as a strength source to the equilibrium equations. An analytical solution to the latter for a disk-shaped reservoir embedded in a homogeneous elastic half space undergoing a uniform pore pressure decline was developed by Geertsma [47]. Although this solution refers to a highly idealized setting, it is valuable for first-hand and computationally inexpensive estimates of the impact that fluid extraction or injection can exert on land surface levels.

In realistic geological settings, however, a reliable prediction of land subsidence requires appropriate numerical models able to account for the three-dimensional (3-D) geometry of the reservoir and the overburden, medium heterogeneity and anisotropy, hydro-mechanical coupling, if relevant, and the elasto-plastic behavior of porous media. The application of these sophisticated tools has often been limited by the uncertainty in the poro-mechanical parameters necessary to accurately simulate or forecast the change in ground surface levels caused by fluid injection and/or extraction. Resolving these uncertainties is of paramount importance in order to increase the reliability of results obtained with these models.

Typically, one of the most important and uncertain parameters in land subsidence modeling is the vertical uniaxial compressibility, c_M. Compressibility is traditionally estimated through either laboratory tests or in-situ measurements of the deformation of compacting fluid-bearing formations. Laboratory tests rely on tri-axial or oedometric experiments carried out on rock samples extracted from boreholes completed into the geological formation of interest. The c_M values obtained with these tests often overestimate the real medium properties due to the alteration occurring in the coring and transportation of the samples [49-53]. On the other hand, more reliable c_M values are obtained through in-situ measurements of the reservoir compaction, which can be monitored, for example, with the "radioactive marker" technique [54-56].

In this work, we use Geertsma's solution [47] as the direct model to predict land subsidence. This model (Figure 1) considers a cylindrical reference system $[r, \theta, z]$ with the vertical axis z oriented upwards, $r = \sqrt{x^2 + y^2}$, and $\theta = arctan(y/x)$, where a disk-shaped horizontal reservoir with radius R and thickness h, is centered at $r = 0$ and embedded at average depth c within a semi-infinite elastic and isotropic porous medium. The reservoir is assumed to be subject to a uniform change in pore fluid pressure, Δp. According to Geertsma's solution, the radial-symmetric horizontal and vertical displacement fields, u_r and u_z at any generic location $[r, \theta, z]$ are expressed as:

$$u_r(r,z) = \{I_1[r, \varepsilon \cdot (c+z)] + (3 - 4 \cdot \nu) \cdot I_1[r, \varepsilon \cdot (c-z)] + 2 \cdot z \cdot I_2(r, c-z)\} \cdot \frac{R \cdot c_M \cdot h \cdot \Delta p}{2} \quad (1)$$

$$u_z(r,z) = \{\varepsilon \cdot I_3[r, \varepsilon \cdot (c+z)] + (3 - 4 \cdot \nu) \cdot I_3[r, \varepsilon \cdot (c-z)] - 2 \cdot z \cdot I_4(r, c-z)\} \cdot \frac{R \cdot c_M \cdot h \cdot \Delta p}{2} \quad (2)$$

In Equations (1-2), ν is the Poisson ratio of the porous medium, and ε an integer equal to 1 if $|z| < c$ and equal to -1 otherwise. The functions I_1, I_2, I_3, and I_4 consist of the following Hankel integrals:

$$\begin{aligned} I_1(r,q) &= \int_0^\infty e^{-q\cdot\tau} \cdot \mathrm{J}_1(\tau \cdot R) \cdot \mathrm{J}_1(\tau \cdot r) \cdot d\tau \\ I_2(r,q) &= \int_0^\infty e^{-q\cdot\tau} \cdot \tau \cdot \mathrm{J}_1(\tau \cdot R) \cdot \mathrm{J}_1(\tau \cdot r) \cdot d\tau \\ I_3(r,q) &= \int_0^\infty e^{-q\cdot\tau} \cdot \mathrm{J}_1(\tau \cdot R) \cdot \mathrm{J}_0(\tau \cdot r) \cdot d\tau \\ I_4(r,q) &= \int_0^\infty e^{-q\cdot\tau} \cdot \tau \cdot \mathrm{J}_1(\tau \cdot R) \cdot \mathrm{J}_0(\tau \cdot r) \cdot d\tau \end{aligned} \quad (3)$$

where J_α is the generic first-type Bessel function of order α (0 or 1). From Equations (1-2), the surface ($z = 0$) displacement components (u_r, u_z) as a function of the radial distance, r, are given by:

$$u_r(r,0) = 2 \cdot (1-\nu) \cdot R \cdot c_M \cdot h \cdot \Delta p \cdot I_1(r,c) \tag{4}$$

$$u_z(r,0) = 2 \cdot (1-\nu) \cdot R \cdot c_M \cdot h \cdot \Delta p \cdot I_3(r,c) \tag{5}$$

Equations (4-5) show that the surface displacement components depend on the six independent parameters R, c, h, Δp, c_M, and ν, and are linearly proportional to the product $(1-\nu) \cdot c_M \cdot h \cdot \Delta p$.

3 Estimation of Geomechanical Parameters using the Ensemble Smoother

The ES algorithm adopted in this work follows a two-step forecast-update process. The forecast, or prediction, is obtained with a Monte Carlo simulation of the system state. The update, or correction, of the system occurs when available measurements are assimilated into the forecast model results.

3.1 Forecast step

The forecast step involves simulating an ensemble of model state $\boldsymbol{U^f}$ based on the solution to the mathematical model, $\boldsymbol{\Phi}$, which generally depends upon uncertain system parameters, $\boldsymbol{P}$, forcing terms, $\boldsymbol{p}$, initial conditions, $\boldsymbol{U}_0$, and boundary conditions, $\boldsymbol{b}$:

$$\boldsymbol{U}_t^f = \boldsymbol{\Phi}(\boldsymbol{P}; \boldsymbol{U}_0; \boldsymbol{p}; \boldsymbol{b}) \tag{6}$$

The parameters $\boldsymbol{P}$ are assumed to be time-independent, whereas both $\boldsymbol{p}$ and $\boldsymbol{b}$ can be time-dependent. The system state $\boldsymbol{U}_t^f$ represents the ensemble of systems states at a generic time t generated through stochastic simulation of the model $\boldsymbol{\Phi}$, using a corresponding ensemble of uncertain parameters and variables generated according to pre-established statistical

distributions. Typically, $\boldsymbol{U}_t^f$ can be represented by a $n_f \times n_{MC}$ matrix, where n_f denotes the number of degrees of freedom used to characterized the system state and n_{MC} the number of realizations in the ensemble.

For the ideal subsurface system described in Section 2, system parameters $\boldsymbol{P}$ include the radius R, the thickness h and the depth c of the disk-shaped reservoir, and the geomechanical parameters c_M and ν; forcing terms $\boldsymbol{p}$ include the change in pore fluid pressure, Δp, within the reservoir. The solution to the mathematical model $\boldsymbol{\Phi}$ is obtained by employing Equations (4-5) to calculate the surface displacements, $u_r(r,0)$ and $u_z(r,0)$, over a regular grid made up by n nodes. Note that $n_f = 2 \cdot n$ since the displacement at each node is characterized by two components. The horizontal and vertical surface displacement fields are denoted by the vectors $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$. Uncertainty in the model state may be established by providing random values to $\boldsymbol{P}$, $\boldsymbol{U}_0$, $\boldsymbol{p}$, and $\boldsymbol{b}$ for each member in the surface displacement simulation ensemble. In this work, the only random variables taken into consideration are the six reservoir parameters R, c, h, Δp, c_M, and ν.

3.2 Update Step

In this step, the forecasted ensemble $\boldsymbol{U}_t^f$ established through Equation (6) is corrected, or updated, using m field measurement data through the following algorithm:

$$\boldsymbol{U}_t^u = \boldsymbol{U}_t^f + \boldsymbol{K}_t \cdot \left(\boldsymbol{D}_t - \boldsymbol{H} \cdot \boldsymbol{U}_t^f\right) \tag{7}$$

where $\boldsymbol{U}_t^u$ $[n_f \times n_{MC}]$ is the updated ensemble, and $\boldsymbol{H}$ $[m \times n_f]$ is a matrix that maps measurement locations into the model grid, so that the product $\boldsymbol{H} \cdot \boldsymbol{U}_t^f$ $[m \times n_{MC}]$ holds the ensemble of model results at measurement locations. The matrix $\boldsymbol{D}_t$ $[m \times n_{MC}]$ holds the measurement data perturbed using an ensemble of Gaussian noises, stored in a matrix $\boldsymbol{E}$

$[m \times n_{MC}]$ representing the measurement random error. If the measurements are error-free, all n_{MC} columns of $\boldsymbol{D}_t$ are equal to the data.

At the right-hand side of Equation (7), the residual $\boldsymbol{D}_t - \boldsymbol{H} \cdot \boldsymbol{U}_t^f$ defines the deviation between the forecasted state and the true state at measurement locations. This residual forms the basis for correcting the forecast ensemble. The degree of this correction depends upon the uncertainty of both the forecast ensemble and the measurement data, which is contained in the Kalman Gain matrix $\boldsymbol{K}_t$ $[n_f \times m]$:

$$\boldsymbol{K}_t = \boldsymbol{C}^f \boldsymbol{H}^T (\boldsymbol{H} \boldsymbol{C}^f \boldsymbol{H}^T + \boldsymbol{R})^{-1} \tag{8}$$

where $\boldsymbol{C}^f$ $[n_f \times n_f]$ is the forecast error covariance matrix and $\boldsymbol{R}$ $[m \times m]$ is the measurement error covariance matrix. These two matrices are defined as:

$$\text{(a)} \qquad \boldsymbol{C}^f = \frac{\left(\boldsymbol{U}_t^f - \overline{\boldsymbol{U}}\right)\left(\boldsymbol{U}_t^f - \overline{\boldsymbol{U}}\right)^T}{n_{MC} - 1} \qquad \text{(b)} \qquad \boldsymbol{R} = \frac{\boldsymbol{E}\boldsymbol{E}^T}{n_{MC} - 1} \tag{9}$$

where each column of $\overline{\boldsymbol{U}}$ $[n_f \times n_{MC}]$ holds the average value of the ensemble of each node displacement. Hence the matrices $\boldsymbol{C}^f$ and $\boldsymbol{R}$ contain the spread of the model values and the measurement values, respectively.

It may be shown that, in Equation (7), if the spread of the measurement values ($\boldsymbol{R}$) is small compared to the spread of the model values ($\boldsymbol{H} \boldsymbol{C}^f \boldsymbol{H}^T$), i.e., the measurement value is "trusted" more, then the values contained in the Kalman Gain $\boldsymbol{K}_t$ approach 1.0, and the residual between the model values and the measurement values, $\boldsymbol{D}_t - \boldsymbol{H} \cdot \boldsymbol{U}_t^f$, is weighted more heavily in correcting the model value to approach the measurement value. Conversely, if the spread of the measurements is large with respect to the spread of the model values, then the residual receives little weight in correcting the model value, which remains similar to the forecast estimate. It must be observed that, since the coefficients of $\boldsymbol{C}^f$ quantify the spatial correlation between displacements at distinct node locations, the ES algorithm has the effect

of spreading information on the system from measurement locations to non-measurement locations [39]. In this study, $\boldsymbol{D}_t$ contains the perturbed measurements of the surface displacement vector (u_r, u_z) at a number of pre-established locations.

3.3 Coupled Update of Surface Displacement and Reservoir Parameters

Since the surface displacement distributions $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ are dependent on the uncertain reservoir parameters R, c, h, Δp, c_M, and ν, field measurements of displacement may be "inverted" to provide improved information about these parameters. Within the ES framework, this is accomplished by including the ensemble of the reservoir parameters into the forecast matrix $\boldsymbol{U}_t^f$. With this approach, the j-th column of $\boldsymbol{U}_t^f$ will have the structure $\left[\boldsymbol{u}_r^{(j)}, \boldsymbol{u}_z^{(j)}, R^{(j)}, c^{(j)}, h^{(j)}, \Delta p^{(j)}, c_M^{(j)}, \nu^{(j)}\right]^{\mathrm{T}} (j = 1,2, \dots, n_{MC})$. The size of $\boldsymbol{U}_t^f$ is thus $(n_f + n_p) \times n_{MC}$, where n_p is the number of uncertain reservoir parameters (equal to six in this case).

By using forecast simulations, this approach yields a matrix $\boldsymbol{C}^f$ $\left[(n_f + n_p) \times (n_f + n_p)\right]$ that contains not only the spatial covariance between displacements at different locations, but also the spatial cross-covariance sub-matrices between displacement distributions $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ and the five reservoir parameters, R, c, h, Δp, c_M, and ν, which ultimately allows field measurements to correct the forecast parameter values as well as the forecast displacement ensembles.

In contrast to other sequential filtering methods, such as the EnKF and the EnKS [27], the ES has the ability of including all displacement distributions and measurement data up until the final measurement collection time t_F, at which the ES update algorithm is run to provide updated system states at all previous collection times. This allows the update algorithm to be run only once when the model simulation has ended, rather than embedding a sequential

update algorithm into the modeling code [27]. At time t_F, the forecast matrix $\widetilde{\boldsymbol{U}}_{t_F}{}^{f}$ thus includes the ensemble of displacements at all collection times and the ensemble of reservoir parameters:

$$\widetilde{\boldsymbol{U}}_{t_F}^{f} = \left[\boldsymbol{U}_{t_{m1}}^{f}, \boldsymbol{U}_{t_{m2}}^{f}, \dots, \boldsymbol{U}_{t_F}^{f}\right]^{T} \qquad \left(n_{t_m} \cdot n_f + n_p\right) \times n_{MC} \tag{10}$$

where n_{t_m} is the number of times at which measurements are collected. The forecast covariance matrix $\widetilde{\boldsymbol{C}}_{t_F}{}^{f}$, obtained by substituting $\boldsymbol{U}_t^f$ with $\widetilde{\boldsymbol{U}}_{t_F}{}^{f}$ in Equation (9a), now contains both spatial covariance terms and temporal covariance terms between grid nodes at different data collection times.

The measurement matrix $\widetilde{\boldsymbol{D}}_{t_F}$ holds perturbed measurements from all data collection times:

$$\widetilde{\boldsymbol{D}}_{t_F} = \left[\boldsymbol{D}_{t_{m1}}, \boldsymbol{D}_{t_{m2}}, \dots, \boldsymbol{D}_{t_F}\right]^{\boldsymbol{T}} \qquad \left(n_{t_m} \cdot m\right) \times n_{MC} \tag{11}$$

The measurement error covariance matrix $\widetilde{\boldsymbol{R}}_{t_F}$ is also established using the perturbations for each of the measurement values for each of the n_{t_m}collection times (see Equation (9b)). By inserting $\widetilde{\boldsymbol{U}}_{t_F}^{f}$ and $\widetilde{\boldsymbol{D}}_{t_F}$ into Equation (7) and $\widetilde{\boldsymbol{C}}_{t_F}^{f}$and $\widetilde{\boldsymbol{R}}_{t_F}$ into Equation (8), the updated system state matrix $\widetilde{\boldsymbol{U}}_{t_F}^{u}$ contains updated model states for each measurement time and also updated system parameters if these are included in the forecast matrix $\widetilde{\boldsymbol{U}}_{t_F}^{f}$.

Note that from a Bayesian perspective, the ES algorithm aims at estimating the joint "posterior" probability distribution function (PDF), PDF$[(\boldsymbol{u}_r, \boldsymbol{u}_z, R, c, h, \Delta p, c_M, \nu)|(\boldsymbol{D}, \boldsymbol{I})]$, of the augmented model states (*i.e.*, the displacement fields $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ at all measurement times, and the reservoir parameters $R, c, h, \Delta p, c_M$, and ν) conditional to the collected surface displacement data $\boldsymbol{D}$ and to the prior information $\boldsymbol{I}$ hypothesized for the reservoir parameters.

In summary, the data assimilation procedure developed is based on the following sequence of steps:

a. Generate the forecast ensemble of reservoir parameters $\left[R^{(j)}, c^{(j)}, h^{(j)}, \Delta p^{(j)}, c_M^{(j)}, \nu^{(j)}\right] (j = 1,2, \dots, n_{MC})$ according to prescribed independent prior statistics for each parameter.

b. Run the geomechanical model (Equation 4-5) for each realization of the forecast parameter ensemble generated in **a.,** to obtain the forecast displacement ensemble $\left[\boldsymbol{u}_r^{(j)}, \boldsymbol{u}_z^{(j)}\right] (j = 1,2, \dots, n_{MC})$ at each of the data collection times $t_{m1}, t_{m2}, \dots, t_F$.

c. Assemble the forecast matrix $\widetilde{\boldsymbol{U}}_{t_F}^f$ according to Equation (10), and the measurement matrix $\widetilde{\boldsymbol{D}}_{t_F}$ according to Equation (11).

d. Calculate the update matrix $\widetilde{\boldsymbol{U}}_{t_F}^u$ using Equations (7-9).

The algorithm used to solve Equation (7) within the ES framework is based on the procedure presented in Reference [57], which provides an efficient numerical strategy for updating the system state.

3.4 Ensemble Smoothing Performance

In order to quantify the effectiveness of the ES algorithm in updating the forecasted model states, a procedure is followed whereby "synthetic" measurement data are collected from the "true" reference state. The reference state is obtained with the land subsidence model using the "true" parameter values. Doing so allows for computing the deviation from the "true" system state on a node-by-node basis for both the forecasted and updated system states, which quantifies the degree to which the forecasted state is corrected. This correction

is analyzed by the two associated global parameters *AE* (absolute error) and *AEP* (average ensemble precision) [39]:

$$AE(\boldsymbol{U}) = \frac{1}{n_{MC} \cdot n} \cdot \sum_{j=1}^{n_{MC}} \sum_{i=1}^{n} \left| U_{i,j} - U_{i,true} \right| \tag{12}$$

$$AEP(\boldsymbol{U}) = \frac{1}{n_{MC} \cdot n} \cdot \sum_{j=1}^{n_{MC}} \sum_{i=1}^{n} \left| U_{i,j} - \overline{U}_i \right| \tag{13}$$

where $\overline{U}_i$ is the ensemble mean at the i-th node, $U_{i,true}$ is the "true" value at the i-th node node, and $U_{i,j}$ is the variable value at the i-th node in the j-th realization of the ensemble. Equation (12) provides a measure of the deviation between the model state and the reference state, whereas Equation (13) provides a measure of the spread of the values around the ensemble mean of the model state. Lower values of AE correspond to a model state that is approaching the reference state, whereas lower values of AEP signify reduced uncertainty in the model state. Calculating the difference between performance parameters of the forecasted and updated model states assesses the performance of the update algorithm.

4 Results

Equations (4-5) indicate that the distribution and the magnitude of the surface displacements, $u_r(r,0)$ and $u_z(r,0)$ is strongly related to the value of reservoir parameters R, c, h, Δp, c_M, and ν. Consequently, the uncertainty on these parameters inevitably affects the reliability of the displacement simulation. In this section, we analyze quantitatively several aspects of the performance of the ES in reducing parameter uncertainty based on measurements of land surface displacement.

4.1 Forecast Ensemble of Model States

We consider the reference case of a disk-shaped reservoir characterized by the uncertain parameters R, c, h, Δp, c_M, and ν. In addition, we assume that the "prior" uncertainty on these parameters can be modeled using a stochastic conceptual model, according to which each parameter is represented as a statistical variable characterized by a predefined PDF. The PDFs of R, c, h, Δp, c_M, and ν are given in Table 1. Gaussian-normal PDFs are hypothesized for the radius R, the depth c, and the thickness h of the reservoir, as well as for the fluid pressure change Δp; a log-normal distribution is assumed for the medium compressibility c_M; and a uniform distribution is adopted for Poisson's ratio ν.

INSERT TABLE 1 HERE

Although the assumption of these statistical distributions is somewhat arbitrary, the prescribed degrees of uncertainty are reasonably realistic. For example, current exploration methods allow for identifying the geometry of a gas reservoir with a relatively good accuracy. Gas pressure change values can be also measured with some fairly good level of reliability. By contrast, the porous medium compressibility and Poisson's ratio in the reservoir are typically the most uncertain parameters, given the technical challenges associated with both in-situ and lab tests [53, 54, 56]. On this basis, c_M is here characterized by a large variability, over a few orders of magnitude, thus justifying the choice of a log-normal distribution with a log-standard deviation $\sigma_{\log c_M}$ equal to 1 (Table 1). A similar assumption was made by Ferronato et al. [58].

On the other hand, the variability of ν is much more restricted. While thermodynamic conditions for linearly elastic and isotropic materials require ν to be between -1 and 0.5 [59], its value is strictly positive in typical porous media [60]. Therefore, ν is here simulated as a uniformly distributed random variable spanning the interval (0,0.5) (Table 1).

Using the statistics of Table 1, an ensemble of n_{MC} =1,000 realizations of the reservoir parameters R, c, h, Δp, c_M, and ν is generated. No correlation, or statistical dependence, is assumed among these parameters. The size of this ensemble is chosen based on preliminary analyses (not shown here), which indicated that the ES results do not vary significantly beyond this value of n_{MC}.

The generated ensemble is then used in a stochastic Monte Carlo simulation to obtain the corresponding forecast ensemble of the surface displacements, $u_r(r,0)$ and $u_z(r,0)$. Figure 2 shows the semi-log plots of the forecast ensembles for horizontal (Figure 2a) and vertical (Figure 2b) surface displacement vs. the radial distance from the center of the reservoir.

INSERT FIGURE 2 HERE

The results of this stochastic simulation are used to establish the forecast ensemble of model states, $\left[\boldsymbol{u}_r^{(j)}, \boldsymbol{u}_z^{(j)}, R^{(j)}, c^{(j)}, h^{(j)}, \Delta p^{(j)}, c_M^{(j)}, \nu^{(j)}\right]^{\mathrm{T}}$ $(j = 1,2,\dots,n_{MC})$, assembled into the forecast matrix $\boldsymbol{U}_t^f$. Both horizontal and vertical displacements are computed on $n = 301$ nodes, with a uniform spacing of 10 m. Since $n_f = 2 \cdot n$, in the case of assimilation of data collected at one single time ($n_{t_m} = 1$) the size of $\boldsymbol{U}_t^f$ is $\left(n_{t_m} \cdot n_f + n_p\right) \times n_{MC} = (1 \cdot 2 \cdot 301 + 6) \times 1{,}000 = 608 \times 1{,}000$.

Note that ensemble based algorithms, such as the ES, are developed as an extension of the classic KF to non-linear models. The KF constitutes a best linear unbiased estimator (BLUE) if applied to linear systems with model and measurement errors unbiased, uncorrelated and normal-distributed. Under these hypotheses, Equation (7) provides the "maximum-likelihood" solution, which maximizes the posterior probability of the model states conditional to all measurements collected [28] and the hypothesized prior information on system parameters. Since in this case the model (Equations 4-5) is generally non-linear, the ES solution can be sub-optimal. However, this sub-optimality may be reduced if all

requirements of the KF other than model linearity are met. It is thus suitable to require that: *(i)* all components of the model state vector $[\boldsymbol{u}_r, \boldsymbol{u}_z, R, c, h, \Delta p, c_M, v]$ be Gaussian stochastic variables; *(ii)* all measurement errors for $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ be standard-normal distributed (see Section 4.3). To meet these conditions, adequate transformations must be applied to the state variables that are not Gaussian. In the present case, the reservoir parameters R, c, h and Δp are already normal-distributed and thus require no transformation. Instead, c_M is log-transformed, whereas v is first scaled to a standard uniform distribution, $U(0,1)$, and then transformed by using the inverse function, denoted Ψ^{-1}, of the standard normal CDF [61].

In order to identify the statistical distributions of $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ a fit analysis is carried out (Figure 3), consisting of a linear regression of the dataset $\left[f_\Phi(u_j); \Phi^{-1}\left(\frac{j-0.5}{n_{MC}}\right)\right]$ $(j = 1,2,\dots,n_{MC})$ where u_j is the surface displacement sampled by stochastic simulation, i.e., either u_r or u_z, Φ^{-1} is the inverse of the theoretical CDF, numerically computed, and f_Φ is a function of u_j that depends on the type of CDF under consideration. The fit analysis is conducted by testing different types of CDF. The "goodness" of fit for the sampled distribution is assessed based on the accuracy with which the regression line approximates the sample. In Figure 3, the statistical distribution that provides the best fit is log-normal for both u_r and u_z. In practice, this shows that the forecast ensemble of the displacements must be log-transformed to meet the theoretical conditions required by the ES.

INSERT FIGURE 3 HERE

4.2 Reference State and Measurement Collection

In addition to the n_{MC} realizations of $[\boldsymbol{u}_r, \boldsymbol{u}_z, R, c, h, \Delta p, c_M, v]$ forming the forecast ensemble, another random realization is created to establish a "true" reference state of the system. The reference state is created to: (a) provide a displacement spatial distribution from which measurements can be collected; (b) provide a system state against which both the

forecasted and updated ensemble of model states can be compared, in order to assess the performance of the ES update algorithm (Equation 12).

The reservoir parameters for the reference system are indicated in the rightmost column of Table 1. These parameters are used to obtain the "true" reference fields for $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ from which measurements of surface displacement are collected. The "true" reference fields are represented by the solid-black profiles plotted in Figure 2. Since the $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ fields are log-normally distributed (Figure 3), the measurements data are log-transformed before use in the ES update algorithm. Each of the measurement log-values is then perturbed using a prescribed standard deviation, σ. Note that the logarithmic transformation has also the effect of producing an ensemble of perturbed measurements that maintains the same sign as the measurement values being considered. The perturbed measurements are used to populate the error matrix $\boldsymbol{E}$ and, through Equation (9b), calculate the measurement error covariance matrix $\boldsymbol{R}$. The effects on the ES performance of the number of measurements collected and of the intensity of measurement errors are investigated in Section 4.4.

4.3 Analysis of Updated Ensembles

The basic set of data assimilation scenarios consists of updating the forecast ensembles of $\log(\boldsymbol{u}_r)$ and $\log(\boldsymbol{u}_z)$, R, c, h, Δp, $\log(c_M)$, and $\Psi^{-1}(\nu)$ using measurements of surface horizontal and vertical displacement collected from the five locations shown in Figure 2. Observe that, since two measurements are collected at each location, u_r and u_z, the total number of data, m, is equal to 10. We assume the log-transformed measurements to be affected by an error corresponding to a standard deviation $\sigma = 0.05$. The results obtained after applying the ES algorithm (Equation 7) are summarized in Figures 4 and 5.

INSERT FIGURE 4 AND 5 HERE

The subpanels in Figure 4 show the profiles of the sampled CDFs of the six reservoir parameters R, c, h, Δp, c_M, and ν obtained before and after the ES update. Each CDF profile is obtained by sorting the ensemble of the n_{MC} values of the considered reservoir parameter, for example c_M, in ascending order, $c_{M,1} \leq c_{M,2} \leq \cdots \leq c_{M,n_{MC}}$. The CDF value associated to $c_{M,j}$ $(j = 1,2, \ldots, n_{MC})$ is then calculated as [62]:

$$\mathrm{CDF}(c_{M,j}) = \frac{j - 0.5}{n_{MC}} \tag{12}$$

In each subpanel of Figure 4, the "true" reference value of the corresponding reservoir parameter is also shown. The spread of the CDF profiles around the reference value gives a straightforward assessment of the uncertainty associated with the parameter estimation before and after assimilation of the displacement data. One may observe that the ES produces a strong reduction in uncertainty for c_M (Figure 4e) and a low - yet non-negligible - reduction in uncertainty for R and c (Figure 4c). However, no improvement is observed for all other parameters.

The primary cause of this effect may be explained by the influence that the uncertainty in the reservoir parameters has on the intensity of the land surface displacements. A sensitivity analysis conducted on the five parameters R, c, h, Δp, c_M, and ν, whose results are not shown here, reveals that the intensity of the ground surface displacements is mostly sensitive to the porous medium compressibility c_M since this is the parameter with the largest degree of prior uncertainty (see Table 1). On the other hand, the radius R and the depth c of the reservoir have a direct effect on the shape of the displacement profile. In this situation, the ES smoothing algorithm proves effective to condition c_M and, to a minor extent, R and c.

The performance of the ES algorithm can be also quantified by analyzing the reduction, with respect to the forecast stage (see Table 2, Scenario 0), of the two coefficients AE and AEP (Equations 12-13) for the updated ensembles of $\log(\boldsymbol{u}_r)$ and $\log(\boldsymbol{u}_z)$, R, c, h, Δp,

$\log(c_M)$, and $\Psi^{-1}(\nu)$ reported in Table 2 for Scenario 2. AE and AEP are reduced: for R, from 83.9 and 81.2 (Scenario 0) to 70.1 and 61.5 (Scenario 2), respectively; for c, from 107.3 and 76.0 (Scenario 0) to 63.7 and 50.3 (Scenario 2), respectively; and for $\log(c_M)$ from 0.944 and 0.820 (Scenario 0), to 0.149 and 0.125 (Scenario 2), respectively. No significant reduction is observed for the other reservoir parameters.

A well-known inconsistency of ensemble-based KF methods, when applied to non-linear processes, lies in their tendency to provide updated state values that are not coherent with the physical laws behind the model equations. While this may be an important concern when updating system response variables, it is a secondary issue when the goal of data assimilation is to estimate system parameters, as in this case. Nevertheless, a simple way to cope with this issue is to apply the updated ensemble of parameters $\left[R^{(j)}, h^{(j)}, c^{(j)}, \Delta p^{(j)}, c_M^{(j)}, \nu^{(j)}\right]$ $(j = 1,2, \dots, n_{MC})$ in an "a posteriori" stochastic simulation using Equations (4-5) to produce physically consistent updated ensembles of the surface displacement $\left[\boldsymbol{u}_r^{(j)}, \boldsymbol{u}_z^{(j)}\right]$ $(j = 1,2, \dots, n_{MC})$. These ensembles are plotted in Figure 5. The spread of these ensembles around the noisy measurements is remarkably small. As reported in Table 2, the AE and AEP values for the updated surface displacement ensembles are greatly reduced, for $\log(\boldsymbol{u}_r)$, from 1.016 and 0.833 (Scenario 0), to 0.017 and 0.017 (Scenario 2), respectively, and for $\log(\boldsymbol{u}_z)$, from 1.011 and 0.836 (Scenario 0), to 0.018 and 0.017 (Scenario 2), respectively. Figure 5c shows the corresponding prior (forecast) and posterior (update) CDFs of the lumped parameter $(1-\nu) \cdot c_M \cdot h \cdot \Delta p$, which is linearly proportional to the intensity of the surface displacement components, u_r and u_z, according to Equations (4-5). The comparison of these CDFs indicates that the ES is able to greatly condition this parameter around its deterministic reference value. The reduced spread of the displacement ensembles given in Figures 5a and 5b may thus be explained by the substantially small variability of the posterior distribution of

$(1-\nu)\cdot c_M \cdot h \cdot \Delta p$ (Figure 5c) and the variability of the updated reservoir parameters R and c (Figures 4a and 4b), as expressed by Equations (4-5).

INSERT TABLE 2 HERE

The results presented in Figure 4 seem to suggest that the ES is unable to significantly reduce, for example, the uncertainty in the radius of the reservoir R. This may constitute a significant drawback, given that, in a real-world situation in which the Geertsma's analytical solution (Equations 1-2) was used to evaluate the magnitude of the land surface displacement, the reservoir radius would be the parameter representing the average spatial extent of the area interested by a reduction in pore pressure.

To investigate this issue an ad-hoc data assimilation test is carried out, in which the prior uncertainty in the radius of the reservoir R is assumed to be much larger than previously assumed (see Table 1). The reservoir radius R is thus sampled from a uniform PDF characterized by a minimum value of 500 m and a maximum value of 5000 m. Similar to what is done for ν (Section 4.1), R needs to be scaled to a standard uniform distribution, $U(0,1)$, and then transformed to standard normal distribution before use in the ES algorithm.

In this test, all other statistical parameters, the true reference state, the measurement locations and the measurement error remain unchanged. The results of the data assimilation are summarized in Figure 6, which shows the CDFs of R and c_M before and after the assimilation of the land surface displacement data (Figure 2). In each subpanel, the "true" reference value of the corresponding parameter is also shown. These results indicate that, in this case, the ES is able to strongly reduce the uncertainty not only on the medium compressibility c_M, but also on the reservoir radius R. One may observe that, according to Geertsma's model (Equations 1-2), both horizontal and vertical displacements depend on c_M through a linear relationship, and on R through a non-linear relationship. Since in this case the statistical variability of the two parameters is assumed to be high, the sensitivities of the

land displacements to the values of c_M and R are comparable to one another and the ES effectively "distributes" the reduction in uncertainty on both parameters. Figure 6 does not report the CDFs for the reservoir parameters c, h, Δp, and ν, since no substantial difference is observed with the corresponding CDFs presented in Figure 4. Note, however, that a result qualitatively similar to that shown in Figure 6 would be obtained if a much larger degree of prior uncertainty was hypothesized for either h or Δp, since land surface displacements (Equations 4-5) are linear with respect to the product $c_M \cdot h \cdot \Delta p$.

INSERT FIGURE 6 HERE

4.4 Effect of Measurement Errors and Number of Measurement Data

Supplementary data assimilation scenarios are investigated in order to quantify the influence of the measurement error and the number of measurements on the reservoir parameter ensemble updated with the ES algorithm. The results of these tests are summarized in Table 2 and Figure 7.

INSERT FIGURE 7 HERE

In Scenarios 1-4 (Table 2), the measurement error for the log-transformed displacement data is progressively decreased assuming standard deviation values σ equal to 0.5, 0.05, 0.01, and 0.0001. In Table 2, the AE and AEP metrics for the displacement distributions $\log(\boldsymbol{u}_r)$ and $\log(\boldsymbol{u}_z)$ exhibit a general decreasing trend with decreasing measurement errors. With respect to the forecast stage (Scenario 0), the same metrics for the reservoir parameters R, c and c_M are significantly reduced for low values of σ, whereas they remain substantially unchanged for h, Δp, and ν. Figure 7a, 7c, and 7e show the prior and posterior CDFs for the reservoir radius R, the reservoir average depth c, and the compressibility c_M, respectively, obtained after assimilating surface displacement data with the σ values corresponding to the Scenarios 1-4 presented in Table 2.

When analyzing the results in Table 2 and Figure 7 it becomes apparent that, when land surface displacement data are characterized by low measurement errors as, for example, in Scenarios 3 and 4, the ES algorithm is able to strongly reduce the uncertainty not only on c_M, as observed in Scenario 2 (Figure 4e), but also on R and c. In practice this shows that, in terms of parameter estimation by assimilation of surface displacement data and given the assumed prior statistics (Table 1), R and c are more sensitive to measurement errors than c_M. On the other hand, the uncertainty on parameters such as h, Δp, and ν is not reduced since the sensitivity of the land displacement distribution to their value is much smaller than for R, c and c_M.

The last three rows of Table 2 summarize the results of analyses conducted to study the effect of the number of measurements. Scenario 5 represents the case in which surface displacements are collected from a single location ($m = 2$), situated at the intermediate point of the five locations shown in Figure 2. Scenario 2 represents the base case described above, in which five measurement locations are sampled (Figure 2). With respect to Scenario 2, Scenario 6 considers three additional measurement locations positioned at radial distances of 500 m, 1650 m, and 2600 m. The AE and AEP values provided in Table 2 for Scenario 5 ($m = 2$) indicate the ES is able to significantly reduce the uncertainty in the surface displacements by assimilating the surface displacements from a single location adequately chosen. A pronounced improvement is observed for Scenario 2 ($m = 10$). Scenario 6 ($m = 16$) is slightly better but substantially equal to Scenario 2, which indicates that the increase of information provided by further measurements is progressively fading.

As observed above, the only reservoir parameters whose uncertainty is reduced upon assimilation of surface displacement data are R, c and c_M. Figures 7b, 7d and 7f show the forecasted and the updated sample CDFs of R, c and c_M, respectively, obtained in Scenarios 2 ($m = 10$), 5 ($m = 2$), and 6 ($m = 16$). Comparison of these profiles indicates that the ES is

capable to identify the reservoir compressibility c_M even with the assimilation of a few measurements (Figure 7f). The reduction in the prior uncertainty on the reservoir radius R is negligible in Scenario 5 ($m = 2$), whereas some slight improvement is observed in Scenarios 2 ($m = 10$), and 6 ($m = 16$) (Figure 7b). The reduction in the prior uncertainty on the reservoir depth c remains negligible in Scenario 5 ($m = 2$), but becomes quite significant in Scenarios 2 ($m = 10$), and 6 ($m = 16$) (Figure 7d). In each case, as the number of measurements increases the reduction in parameter uncertainty becomes smaller.

4.5 Updating of reservoir parameters during loading/unloading sequences.

A final test is carried out to investigate the efficiency of the ES algorithm for geomechanical characterization during loading/unloading events. These events are characterized by a reduction of the pore fluid pressure in the reservoir, which produces a settlement of the ground surface, followed by a fluid pressure recovery with an ensuing surface rebound. These rebound effects are observed, for example, over reservoirs used for GSR [24], or reservoirs where the fluid pressure drop during gas production is followed by a re-pressurization sustained by the dynamics of aquifers in contact with the reservoir, which may keep entering the original gas region long after the reservoir is abandoned [63]. Under these conditions, the porous medium typically exhibits a hysteretic behavior, with the compressibility of the unloading stage, $c_{M,u}$, reduced with respect to the compressibility of the loading stage, $c_{M,l}$, by a ratio s ($\geq$1), that is, $c_{M,u} = c_{M,l}/s$ [54].

In this test, a reference system with the "true" reservoir parameters reported in Table 1 is considered. These parameters are the same as in the test of Section 4.3. However, the reservoir undergoes a pressure recovery in the unloading phase, Δp_u, equal to half the pressure decrease in the loading phase, Δp_l, while the medium compressibility in the

unloading phase, $c_{M,u}$, is $s = 4.26$ times smaller than the compressibility during the loading phase, $c_{M,l}$ (see the last row of Table 1).

A forecast ensemble of n_{MC} =1,000 states is generated through stochastic simulation of Geertsma's analytical solution. In this case, the reservoir parameters R, h, c, Δp_l, Δp_u, $c_{M,l}$, ν and s are assumed to be uncertain and characterized by the prior PDFs given in Table 1. The ratio s is assumed to fit to a uniform distribution $U(1,10)$. Note that since two assimilation times are considered ($n_{t_m} = 2$) and eight are the uncertain reservoir parameters ($n_p = 8$), the size of $\boldsymbol{U}_t^f$ is $\left(n_{t_m} \cdot n_f + n_p\right) \times n_{MC} = (2 \cdot 2 \cdot 301 + 8) \times 1{,}000 = 1{,}212 \times 1{,}000$, which is about twice as large as in the case considered in Section 4.2.

INSERT FIGURES 8 AND 9 HERE

The ES update results are summarized in Figures 8 and 9. The forecast ensemble of surface displacement distributions $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ in both loading and unloading phases are obtained by applying Equations (4-5) with pressure change and compressibility equal to Δp_l and $c_{M,l}$ during loading, and equal to Δp_u and $c_{M,u}$ during unloading. In this case, the log-transformed measurements are affected by an error quantified by a standard deviation $\sigma = 0.01$. Measurement collection is simulated from the "true" surface displacement fields $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ obtained by applying Geertsma's model with the reservoir parameters given in Table 1. Surface displacement values are collected from the same five locations shown in Figure 2 during both loading and unloading stages, thus assimilating m=20 data overall.

Figure 8 displays the results of the fit analyses (as in Section 4.3) performed on the forecast and update ensembles of the reservoir parameters R, h, c, Δp_l, Δp_u, $c_{M,l}$, ν and s. Similar to what observed in Section 4.3, the ES proves effective in reducing the spread of the prior statistical distributions of R, c, $c_{M,l}$ and s, whereas this spread remains substantially unchanged for h, Δp_l, Δp_u and ν.

Figures 8a and 8b show that the ES is able to narrow significantly the uncertainty on R and c around their respective true value. In Figure 8c, one may note the remarkable effectiveness with which the ES reduces the variability of $c_{M,l}$ around its true value (Table 1). In addition, even after updating, the statistical distributions of R, c and $c_{M,l}$ remain of the same type hypothesized for their prior.

Figure 8d provides the results of the fit analysis for the ratio s, based on the assumption that both forecasted and updated samples are uniformly distributed. While this hypothesis is obviously met for the forecast ensemble, generated on the same assumption, it is not verified for the update ensemble, which indicates that in this case the ES is not able to preserve the PDF type originally prescribed for the ratio s. More importantly, however, the ES is able to provide an updated s with a spread that is significantly reduced with respect to the forecast ensemble. In particular, Figure 8d shows that the updated sample of s is more clustered around its median value (the 50th percentile, where the regression line intersects the abscissa), which is very close to its true value $s = 4.26$ (Table 1).

Figure 9 presents the results obtained with an "a posteriori" stochastic simulation using Equations (4-5) with the updated ensemble of reservoir parameters $\left[R^{(j)}, h^{(j)}, c^{(j)}, \Delta p_l^{(j)}, \Delta p_u^{(j)}, c_{M,l}^{(j)}, \nu^{(j)}, s^{(j)}\right]$ $(j = 1,2,\dots,n_{MC})$ (Figure 8). Figures 9a and 9b show the ensembles of surface displacements $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$, respectively, for the loading phase ($c_M^{(j)} \equiv c_{M,l}^{(j)}$), whereas Figures 9c and 9d display the same ensembles for the unloading phase (with $c_M^{(j)} \equiv c_{M,u}^{(j)} = c_{M,l}^{(j)}/s^{(j)}$). It is interesting to observe that the spread of both the $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ ensembles around the noisy measurements is remarkably small during loading (Figures 9a and 9b), while it is larger during unloading (Figures 9c and 9d). This is the because the ES proves more effective in reducing the larger uncertainty on $c_{M,l}$ (Figure 8c) than it does on s

(Figure 8d). However, this difference is not as pronounced as Figure 9 would indicate at first glance, since a logarithmic scale is used to plot the displacements.

5 Discussion

In general terms, the effectiveness of the ES (Equations 7-8) in reducing the uncertainty on reservoir parameters based on a given number measurement of surface displacement depends on two major factors: (*i*) the magnitude of measurement errors, which is accounted for in the matrix $\boldsymbol{R}$ (Equation 9b) at the "denominator" of the Kalman gain matrix $\boldsymbol{K}_t$ (Equations 8); (*ii*) the values of cross-correlation between displacements and parameters, which are found in the correlation matrix $\boldsymbol{C}^f\boldsymbol{H}^T$ at the "numerator" of $\boldsymbol{K}_t$. The latter are essentially related to (a) the partial derivatives of the displacements with respect to the parameter and (b) the "prior" uncertainty assigned to the parameter, quantified by the spread of the parameter PDF. Larger cross-correlation values typically signify a better ability of the ES to identify a given parameter based on displacement measurements.

The results presented in the previous sections indicate that, by merging surface displacement data into Geertsma's solution, the ES can successfully exploit the correlation between the parameters R, c and c_M and the state vectors $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ to condition the parameter ensembles around their "true" reference value. This "conditioning" is not as effective for reservoir parameters such as h, Δp and ν. Indeed, because of the prior statistical variability hypothesized for h, Δp and ν and the structure of Geertsma's model (Equations 1-2), the correlations of these parameters to $\boldsymbol{u}_r$ and $\boldsymbol{u}_z$ result negligible when compared to the same correlations for R, c and c_M. In other words, R, c and c_M affect the variability of the surface displacement prediction much more than h, Δp and ν.

These effects may be substantiated, for example, by inspecting the coefficients of the Kalman gain matrix $\boldsymbol{K}_t$. Table 3 reports, for Scenario 3 (m =10, $\sigma = 0.01$), the coefficients

of $\boldsymbol{K}_t$ for the rows corresponding to the six reservoir parameters, R, h, c, Δp, $\log(c_M)$, and $\Psi^{-1}(v)$, and the columns corresponding to the log-transformed surface displacements, $\log(u_r)$ and $\log(u_z)$, at the locations $r =$ 1,100 m and $r = 0$ m, where they reach their maximum value, respectively. Table 3 shows that, for at least one of the considered displacement components, the magnitude of the coefficients associated with R, c and $\log(c_M)$ is significantly larger than magnitude of the coefficients associated with h, Δp and $\Psi^{-1}(v)$, and thus explains why the corrections of R, c and c_M achieved by assimilating surface displacements results much more pronounced than the corrections of h, Δp and v.

INSERT TABLE 3 HERE

As explained in Section 3.2, if measurement errors affecting surface displacements are large, then the ES is not able to correct the forecast ensemble and the updated ensemble remains similar to the forecast. This is confirmed by the results obtained in Scenarios 2, 5 and 6 (Figures 7b, 7d and 7f), which show that the ability of the ES to condition the reservoir parameters R, c and c_M strongly depends on measurement errors (Table 2). While the identification of R and c requires the availability of low-noise measurements (Figures 7a and 7c), c_M is less sensitive to measurement errors because its correlation to the displacements is larger given its large prior uncertainty. In this respect, the results in Figure 7c show that even when the log-standard deviation of the measurement error σ is as large as 1, the ES still maintains its ability to reduce the prior uncertainty on c_M.

The number of collected measurements is also relevant to the effectiveness of ES. In the investigated scenarios, five adequately chosen measurement locations seem enough to achieve a fairly good estimation of R, c and c_M (Figures 7b, 7d and 7f) and, consequently, a more reliable prediction of the surface displacements. However, many more measurement locations will be likely necessary in real-world systems in which reservoir parameters are spatially heterogeneous and where the prediction of land motion requires the use of numerical

models by far more complex than Geerstma's. In this instance, it is also expected that the influence of measurement errors will become more significant, with low quality measurement data carrying a limited power to condition reservoir geomechanical parameters.

The strength of the ES algorithm, which is devised to improve the reliability of the modeling of dynamic systems, is best exploited in the simulation of reservoirs subject to time-dependent fluid pressure changes. For example, the simulation of loading/unloading sequences presented in Section 4.5 shows the great benefit that may be achieved from collecting and assimilating time-laps surface displacements at predefined locations. In practice, this has the effect of increasing the total number m of available measurements in the spatio-temporal space, so that data collected at a given time have the ability to condition the model response not only at that time, but also at other past and future collection times. This effect is particularly instrumental in assessing the hysteretic properties of the porous medium compressibility. Indeed, although the conditioning of the loading-phase compressibility, $c_{M,l}$ (Figure 8c) seems more effective than the conditioning of the ratio $s = c_{M,l}/c_{M,u}$ (Figure 8d), it is expected that assimilation of surface displacement data over several loading/unloading cycles will produce a more pronounced reduction of the statistical variability of $c_{M,u}$ and consequently of the ratio s.

Finally, we would like to remark that the effectiveness of the ES data assimilation algorithm heavily relies on the correlation between reservoir parameters and surface displacement, which is estimated using the forecast model stochastic simulation. Implicitly, this requires the model to be able to reproduce the physics of the processes that cause the surface displacements being measured. It is obvious that, if the observed surface displacement is the compound effect of multiple causes, such as natural consolidation, surface loading, or water withdrawal from shallow aquifers, the parameter estimation obtained from the ES framework can have the effect of providing fictitious parameter values

making the model explain the observed data. In principle, this can be dealt with by re-formulating the data assimilation equations to account for model uncertainty and bias in addition to parameter uncertainty. Although a very important topic, model uncertainty and bias in data assimilation is outside the scopes of this work.

6 Conclusions

In this study, an ES, i.e., a statistical data assimilation algorithm that merges measurement data into uncertain model-produced values within a Bayesian framework, has been applied for reservoir geomechanical characterization in developed gas fields. The measurement data consisted of land surface horizontal and vertical displacements collected at a number of points at prescribed collection times.

The ideal system under consideration was a horizontal disk-shaped reservoir subject to a uniform fluid pressure change and embedded in a semi-infinite homogeneous and isotropic porous medium, for which an analytical solution in terms of horizontal and vertical displacement is available from the work of Geertsma [47]. The use of this simple direct model was embraced for merely demonstrative purposes, in order to test the potential of the ES algorithm [27] to assimilate surface displacement data that can be obtained, for example, from InSAR surveys.

Given the low computational cost associated with Geertsma's model, an ensemble of 1,000 realizations of reservoir parameters was used to obtain a corresponding ensemble of land surface displacements. Measurement data consisting of surface horizontal and vertical displacement at prescribed locations were used to condition the generated parameter ensembles, with far superior results achieved in reducing the variability prescribed to parameters that have a larger correlation with the magnitude and the shape of the land surface displacement field, in this case the reservoir radius, R, the reservoir depth, c, and the porous

medium compressibility, c_M. On the other hand, the ES could not significantly improve the prior statistics assigned to the other parameters, such as the Poisson ratio ν, the reservoir thickness, h, and the change in fluid pressure, Δp, which are typically identified with less uncertainty and thus have a smaller impact on the statistical variability of the land displacement.

In the case of a loading/unloading sequence, such as the one typically experienced in gas reservoirs/aquifers systems that are developed and subsequently abandoned or in reservoirs used for GSR, the porous medium exhibits a hysteretic behavior with the compressibility in loading conditions decreased in unloading conditions by a factor s. In this situation, the assimilation of surface displacement data at the end of both the loading and the unloading phases proved effective to identify the corresponding compressibility values with reasonable accuracy.

A major strength of the ES algorithm lies in its ability to assimilate simultaneously measurement data and model states at different times, so that system states at a given time are conditioned by measurements collected at other times. With this structure, the ES update algorithm may become quite a powerful tool for the estimation of system parameters. On the other hand, one has to be aware that ES equations are based on the classic KF, which was derived for linear models. The use of ensemble methods, such the ES, has been proposed as a way to extend the KF to non-linear models, and therefore may not provide an optimal solution for the assimilation of data into these models. The overall satisfactory results obtained in the case of Geertsma's model are conducible to the linearity between the land displacements and the product $(1-\nu)\cdot c_M \cdot h \cdot \Delta p$, and the "weak" non-linearity with respect to R and c.

Future work will involve the application of this methodology to a number of gas reservoirs in the Po plain, Italy, currently used for GSR, where time-laps vertical and

horizontal displacements of the land surface have been collected since 2003 using InSAR techniques [24]. The overall effort will be to assimilate these data into a three-dimensional, transversely isotropic, finite-element numerical model, in order to obtain a representative geomechanical characterization of the reservoir formations at the field scale.

Acknowledgements

The Authors wish to thank Eni E&P, Italy, for providing financial support to the majority of this work. The quality of this article benefited from the comments and suggestions provided by two anonymous reviewers, to whom the Authors are very grateful.

References

[1] Rintoul, W. (1981), The case of the disappearing land, in Drilling Ahead: Tapping California"s Richest Oil Fields, pp. 116–137, Valley Publ., Santa Cruz, CA.

[2] Colazas, X. C., and L. J. Olson (1983), Subsidence monitoring methods and beach mark elevation response to water injection, Wilmington oil field, Long Beach, California, in Proceedings of 1982 Forum on *Subsidence due to Fluid Withdrawals*, edited by E. C. Donaldson and H. van Domselaar, pp. 121–132, U.S. Gov. Print. Off., Washington, D. C.

[3] Colazas, X. C., and R. W. Strehle (1995), Subsidence in the Wilmington oil field, Long Beach, California, USA, in *Subsidence Due to Fluid Withdrawal*, edited by G. V. Chilingarian et al., pp. 285–335, Elsevier Science, Amsterdam, The Netherlands.

[4] Finol, A. S., and Z. A. Sancevic (1995), Subsidence in Venezuela, in *Subsidence Due to Fluid Withdrawal*, edited by G. V. Chilingarian et al., pp. 337–372, Elsevier Science, Amsterdam, The Netherlands.

[5] Zaman, M. M., A. Abdulraheem and J. C. Roegiers (1995), Reservoir compaction and surface subsidence in the North Sea Ekifisk field, in *Subsidence Due to Fluid*

Withdrawal, edited by G. V. Chilingarian et al., pp. 373–423, Elsevier Science, Amsterdam, The Netherlands.

[6] Gambolati, G., G. Ricceri, W. Bertoni, G. Brighenti, and E. Vuillermin (1991), Mathematical simulation of the subsidence of Ravenna, *Water Resour. Res*., 27(11), 2899–2918.

[7] Gambolati, G., P. Teatini and W. Bertoni (1998), Numerical prediction of land subsidence over Dosso degli Angeli gas field, Ravenna, Italy, in *Land Subsidence* – Current Research and Case Studies, edited by J. Borchers, pp. 229–238. Star Publishing, Belmont, CA.

[8] Teatini, P., G. Gambolati, L. Tomasi and M. Putti (1998), Simulation of land subsidence due to gas production at Ravenna coastline, in CENAS – Coastline Evolution of the Northern Adriatic Sea due to Sea Level Rise and Natural and Anthropogenic Land Subsidence, edited by G. Gambolati, pp. 135–152. Kluwer Academic Publisher, Norwell, MA.

[9] Teatini, P., D. Baù and G. Gambolati (2000), Water-gas dynamics and coastal land subsidence over Chioggia Mare field, Northern Adriatic Sea, *Hydrogeo. J.*, 8(5), 462–479.

[10] Palozzo, W., G. Cassiani, G. Brighenti and C. Zoccatelli (2000), Three-dimensional simulation of subsidence due to the gas production in the Barbara field and comparison with field data, in Proceedings of the Sixth International Symposium on *Land Subsidence*, edited by L. Carbognin et al., pp. 397–408. La Garangola, Padova, Italy.

[11] Poland, J.F., and Ireland, R.L. (1988), Land subsidence in the Santa Clara Valley, California, as of 1982: U.S. Geological Survey, Professional Paper 497-F, 61 p.

[12] Gabrysch, R.K., and Neighbors, R.J. (2005), Measuring a century of subsidence in the Houston-Galveston region, Texas, USA, in Seventh International Symposium on *Land Subsidence*, Shanghai, P.R. China, October 23–28, 2005, Proceedings: p. 379–387.

[13] Teatini, P., M. Ferronato, G. Gambolati and M. Gonella (2006), Groundwater pumping and land subsidence in the Emilia-Romagna coastland, Italy: Modeling the past occurrence and the future trend, *Water Resour. Res*, 42, W01406, doi:10.1029/2005WR004242.

[14] Todd, M. R., and G. V. Grand (1993), Enhanced oil recovery using carbon dioxide. *Energy Conversion and Manage.* 34, 1157–1164.

[15] Stancliffe, R. P. W., and M. W. A. van der Kooij (2001), The use of satellite-based radar interferometry to monitor production activity at the Cold Lake heavy oil field, Alberta, Canada, *AAPG Bull.*, 58(5), 781-793.

[16] Gaspar Ravagnani, A. T. F. S., E. L. Ligero and S. B. Suslick (2009), CO_2 sequestration through enhanced oil recovery in a mature oil field. *J. Pet. Sci. Eng.* 65, 129–139.

[17] Ferguson, R. C., C. Nichols, T. Van Leeuwen and V. A. Kuuskraa (2009), Storing CO_2 with enhanced oil recovery. *Energy Procedia 1*, 1989–1996.

[18] Kosloff, D., R. F. Scott and J. Scranton (1980), Finite element simulation of Wilmington oil field subsidence: Part II. Nonlinear modelling, *Tectonophysics*, 70, 159–183.

[19] Holloway, S. (2005), Underground sequestration of carbon dioxide - a viable greenhouse gas mitigation option, *Energy*, 30, 2318–2333.

[20] Preston, C., M. Monea, W. Jazrawi, K. Brown, S. Whitakker, D. White, D. Law, R. Chalaturnyk, and B. Rostron (2005). IEA GHG Weyburn CO_2 monitoring and storage project. *Fuel Processing Technology* 86, 1547–1568.

[21] Ferronato, M., G. Gambolati, C. Janna and P. Teatini (2010), Geomechanical issues of anthropogenic CO_2 sequestration in exploited gas fields. *Energy Conversion and Manage.*, 51, 1918–1928.

[22] Buschback, T. C., and D. C. Bond (1967), Underground storage of natural gas in Illinois - 1967, Illinois State Geological Survey Illinois Petroleum 86, Urbana (IL).

[23] Nagelhout, A. C. G., Roest, J. P. A. (1997), Investigating fault slip in a model of an underground gas storage facility. *Int. J. of Rock Mech. Min. Sci.* 34, 3–4.

[24] Teatini, P., N. Castelletto, M. Ferronato, G. Gambolati, C. Janna, E. Cairo, D. Marzorati, D. Colombo, A. Ferretti, A. Bagliani and F. Bottazzi (2011), Geomechanical response to seasonal gas storage in depleted reservoirs: a case study in the Po River basin, Italy. *J. Geophys. Res.*, 116, F02002, doi:10.1029/2010JF001793.

[25] Kalman, R.E. (1960), A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82, 35-45.

[26] Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99(C5), 10, 143-10,162.

[27] Evensen, G. (2007), Data assimilation. The Ensemble Kalman Filter. Springer-Verlag, Berlin-Heidelberg.

[28] Evensen, G., and P.J. van Leeuwen (2000), An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.*, 128, 1852-1867.

[29] Houtekamer, P.L., and H.L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.*, 126, 796-811.

[30] Reichle, R.H., D.B. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Weather Rev.*, 130, 103-114.

[31] Dunne, S. and D. Entekhabi (2005), An ensemble-based reanalysis approach to land data assimilation. *Water Resour. Res.*, 41, W02013, doi:10.1029/2004WR003449.

[32] Liu, Y., and H.V. Gupta (2007), Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resour. Res.*, 43, W07401, doi:10.1029/2006WR005756.

[33] Camporese, M., C. Paniconi, M. Putti, and P. Salandin (2009), Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow, *Water Resour. Res.*, 45, W10421, doi: 10.1029/2008WR007031.

[34] Hantush, M.M., and M.A. Marino (1997), Estimation of spatially variable aquifer hydraulic properties using Kalman filtering. *J. Hydraul. Engrg. ASCE*, 123(11), 1027-1035.

[35] Moradkhani, H., S. Sorooshian, H.V. Gupta, and P.R. Houser (2005), Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.*, 28, 135-147.

[36] Aksoy, A., F. Zhang, and J.W. Nielsen-Gammon (2006), Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model, *Mon. Weather Rev.*, 134, 2951-2970.

[37] Hendricks Franssen, H. J., and W. Kinzelbach (2008), Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resour. Res.*, 44, W09408, doi: 10.1029/2007WR006505.

[38] Bailey, R.T., and D.A. Baù (2010), Ensemble Smoother assimilation of hydraulic head and return flow data to estimate hydraulic conductivity, *Water Resour. Res.*, 46, W12543, doi: 10.1029/2010WR009147.

[39] Bailey, R.T., and D.A. Baù (2011), Estimating spatially-variable first-order rate constants in groundwater reactive transport systems, *J. Cont. Hydrol.*, 122, 104-121.

[40] Hendricks Franssen, H. J., and W. Kinzelbach (2009), Ensemble Kalman filtering versus sequential self-calibration for inverse modeling of dynamic groundwater flow systems. *J. Hydrol.*, 365, 261-274.

[41] Aanonsen, S.I., G. Naevdal, D.S. Oliver, A.C. Reynolds and B. Valles (2009), The ensemble Kalman filter in reservoir engineering – a review. *SPE J.* 14(3), 393-412. SPE-117274-PA. doi: 10.2118/117274-PA.

[42] Chang, H., Y. Chen and D. Zhang (2010), Data assimilation of Coupled Fluid Flow and Geomechanics Using the Ensemble Kalman Filter. *SPE J.* 15(2), 382-394. SPE-118963.

[43] Wilschut, F., E. Peters, K. Visser, P.A. Fokker, and P.M.E. van Hooff (2011), Joint History Matching of Well Data and Surface Subsidence Observations Using the Ensemble Kalman Filter: A Field Study, SPE Reservoir Simulation Symposium, SPE-141690-PP, DOI: 10.2118/141690-MS.

[44] Iglesias M.A. and D. McLaughlin (2012). Data inversion in coupled subsurface flow and geomechanics models, *Inverse Probl.*, 28 (11), doi:10.1088/0266-5611/28/11/115009.

[45] Emerick A.E. and A.C. Reynolds (2012). Ensemble smoother with multiple data assimilation, *Comput. Geosci*., http://dx.doi.org/10.1016/j.cageo.2012.03.011.

[46] McLaughlin, D. (2002), An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. *Adv. Water Resour.*, 25, 1275-1286.

[47] Geertsma, J. (1973), A basic theory of subsidence due to reservoir compaction: the homogeneous case. *Verhandelingen Kon. Ned. Geol. Mijnbouwk. Gen.* 28, 43-62.

[48] Biot, M. A. (1941), A general theory of three-dimensional consolidation, *J. Appl. Phys.*, 12(2), 155-164.

[49] Pottgens, J. E., and F. J. Browner (1991), Land subsidence due to gas extraction in the northern part of the Netherlands. In *Land Subsidence,* Proceedings of the Houston Symposium, Pub. No. 200, 99–108.

[50] Van Hasselt, J. P. (1992). Reservoir compaction and surface subsidence resulting from oil and gas produtcion. *Geologie en Mijnbouw* 71, 107–118.

[51] Holt, R. M., M. Brignoli, E. Fyaer, C.J. Kenter and T.E.Unander (1994), Core damage effects on compaction behaviour. In Proceedings EUROCK'94, *Rock mechanics for petroleum engineering*, Rotterdam, The Netherlands, Balkema, 55–62.

[52] NAM (1995). Subsidence caused by natural gas production, Technical Report, Nam, Assen.

[53] Cassiani, G. and C. Zocatelli (2000), Towards a reconciliation between laboratory and in-situ measurements of soil and rock compressibility. In *Land Subsidence*, Proceedings of the 6th International Symposium on Land Subsidence, Padova, Vol. 2, 3–15.

[54] Baù, D., Ferronato, M., Gambolati, G. & Teatini, P. (2002), Basin-scale compressibility of the northern Adriatic by the radioactive marker technique, *Geotechnique*, 52, No. 8, 605–616.

[55] Ferronato, M., G. Gambolati, P. Teatini and D. Baù (2004), Radioactive marker measurements in heterogeneous reservoirs: a numerical study. *Int. J. Geomech.* 4, 79–92.

[56] Ferronato, M., G. Gambolati, P. Teatini and C. Janna (2007), Casing influence in reservoir compaction measurement by radioactive markers in the Northern Adriatic, Italy, *Int. J. Geomech.*, 7, 444–447.

[57] Keppenne, C. (2000), Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Weather Rev.*, 128, 1971-1981.

[58] M. Ferronato, G. Gambolati, P. Teatini, and D. Baù (2006), Stochastic Poromechanical Modeling of Anthropogenic Land Subsidence, *Int. J. Solids Struct.*, 43,3324-3336.

[59] Timoshenko, S.P. (1983), History of strength of materials. New York: Dover Publications; 1983.

[60] Gercek, H. (2007), Review - Poisson's ratio values for rocks. *Int. J. Rock Mech. Min. Sci.* 44, 89–101.

[61] Abramowitz, M, and I.A. Stegun, Eds. (1972), Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, New York: Dover Publications, ISBN 978-0-486-61272-0.

[62] Hahn, G. J. and S.S. Shapiro (1967), Statistical models in engineering, John Wiley & Sons, New York.

[63] Baù, D., G. Gambolati and P. Teatini (2000), Waterdrive Dynamics and Enhanced Land Subsidence over Productive Gas Fields: Application to Dosso degli Angeli Reservoir, Ravenna, Italy, in *Theory, Modeling and Field Investigation in Hydrology: A Special Volume in Honor of Shlomo P. Neuman's 60th Birthday*, Geological Society of America, D. Zhang and L.C. Winter Eds., GSA Special Paper 348, 129-149.

Table 1. Statistics prescribed to characterize variability of parameters affecting surface displacements. The far-right column shows the true value of each parameter used in the data assimilation tests.

Parameter (unit)	Symbol	PDF Type	Statistical Parameters		Reference (true) Value
Radius (m)	R	*normal*	$\mu_R = 1{,}000$	$\sigma_R = 100$	975.09
Thickness (m)	h	*normal*	$\mu_h = 40$	$\sigma_h = 10$	37.51
Depth (m)	c	*normal*	$\mu_c = 1{,}200$	$\sigma_c = 100$	1,291.10
Pressure change (Pa)	Δp	*normal*	$\mu_{\Delta p} = -1\times10^6$	$\sigma_{\Delta p} = 0.5\times10^5$	-1.009×10^6
Compressibility (Pa^{-1})	c_M	*log-normal*	$\mu_{\log(c_M)} = -9$	$\sigma_{\log(c_M)} = 1$	2.63×10^{-10}
Poisson Ratio (/)	ν	*uniform*	$\nu_{min} = 0.0$	$\nu_{max} = 0.5$	0.36
Compressibility Ratio (/)	s	*uniform*	$s_{min} = 1$	$s_{max} = 10$	4.26

Table 2. Sensitivity of the performance parameters, AE and AEP, to the number of measurements, and to the standard deviation of measurement errors.

				Surface Displacement		**Reservoir Parameter**					
	Analysis		**Scenario**	$\log(\boldsymbol{u}_r)$ *AE* *AEP*	$\log(\boldsymbol{u}_z)$ *AE* *AEP*	R *AE* *AEP*	c *AE* *AEP*	h *AE* *AEP*	Δp *AE* *AEP*	$\log(c_M)$ *AE* *AEP*	$\Psi^{-1}(\nu)$ *AE* *AEP*
Forecast			0	1.016 0.833	1.011 0.836	83.9 81.2	107.3 76.0	8.00 7.78	4.04×10^4 3.95×10^4	0.944 0.820	0.945 0.792
Update	σ	0.50	1	0.132 0.127	0.130 0.128	81.6 80.3	104.7 75.2	7.85 7.75	4.04×10^4 3.95×10^4	0.182 0.174	0.910 0.785
		0.05	2	0.017 0.017	0.018 0.017	70.1 61.5	63.7 50.3	7.84 7.74	4.04×10^4 3.95×10^4	0.149 0.125	0.917 0.785
		0.01	3	0.0043 0.0042	0.0042 0.0041	29.1 26.7	22.7 21.0	7.86 7.73	4.04×10^4 3.95×10^4	0.133 0.116	0.927 0.783
		0.0001	4	0.0002 0.0001	0.0002 0.0001	3.4 3.2	1.6 1.3	7.82 7.70	4.04×10^4 3.94×10^4	0.126 0.114	0.881 0.780
	m	2	5	0.036 0.035	0.044 0.041	80.0 77.4	100.1 73.1	7.85 7.75	4.04×10^4 3.95×10^4	0.158 0.132	0.908 0.783
		10	2	0.017 0.017	0.018 0.017	70.1 61.5	63.7 50.3	7.84 7.74	4.04×10^4 3.95×10^4	0.149 0.125	0.917 0.785
		16	6	0.014 0.014	0.014 0.014	66.3 57.7	56.8 45.2	7.85 7.75	4.04×10^4 3.95×10^4	0.146 0.123	0.916 0.783

Table 3. Coefficients of the Kalman gain matrix for Scenario 3. Rows correspond to the six transformed reservoir parameters $\boldsymbol{R}$, $\boldsymbol{h}$, $\boldsymbol{c}$, $\boldsymbol{\Delta p}$, $\mathbf{log}(\boldsymbol{c_M})$, and $\boldsymbol{\Psi^{-1}(\nu)}$, whereas the two columns correspond to the log-transformed surface displacements, $\mathbf{log}(\boldsymbol{u_r})$ and $\mathbf{log}(\boldsymbol{u_z})$, at the locations $\boldsymbol{r} =$ 1,100 m and $\boldsymbol{r} = 0$ m.

		Observation	
		log[u_r(1100,0)]	log[u_z(0,0)]
Parameter		column: 3	column: 6
R	row: 603	+0.1404	-0.3177
h	row: 604	+0.0043	+0.0195
c	row: 605	+0.0050	-0.2579
Δp	row: 606	+0.0017	-0.0060
log c_M	row: 607	+0.0862	+0.1078
$\Psi^{-1}(\nu)$	row: 608	-0.0070	-0.0050

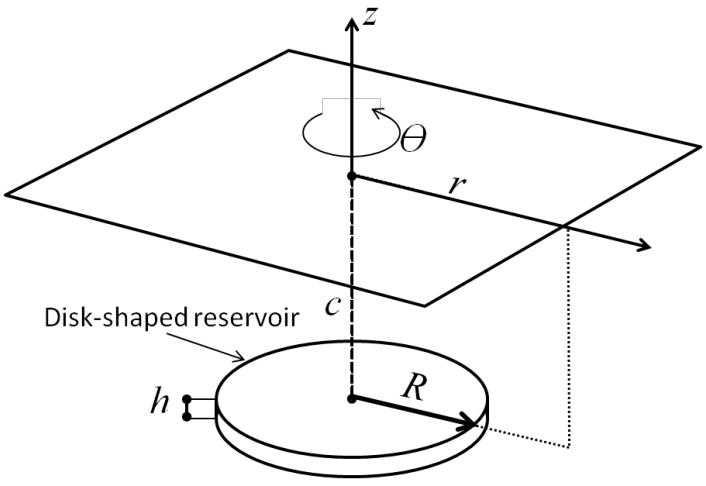

Figure 1. Geometry for Geertsma's analytical solution to the land subsidence problem. The reservoir is disk-shaped and embedded in a semi-infinite elastic homogeneous and isotropic medium. Reservoir compaction occurs due to a uniform change in gas pressure within the cylindrical volume.

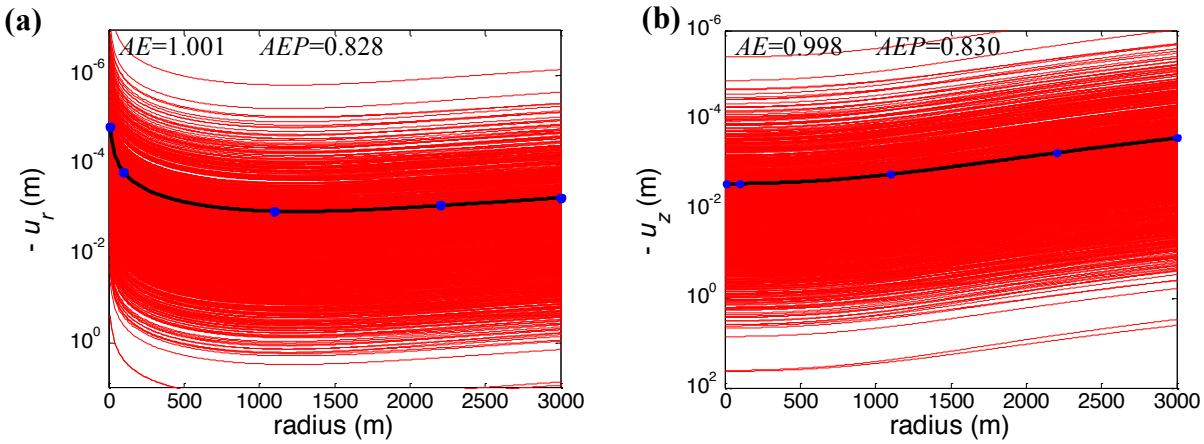

Figure 2. Semi-log plots of the forecast ensembles for (a) horizontal and vertical surface displacement as a function of the radial distance from the center of the reservoir (Table 2, Scenario 0). In each subpanel, the solid black line represents the displacement for the "true" reference system, from which measurement data, indicated by blues dots, are collected.

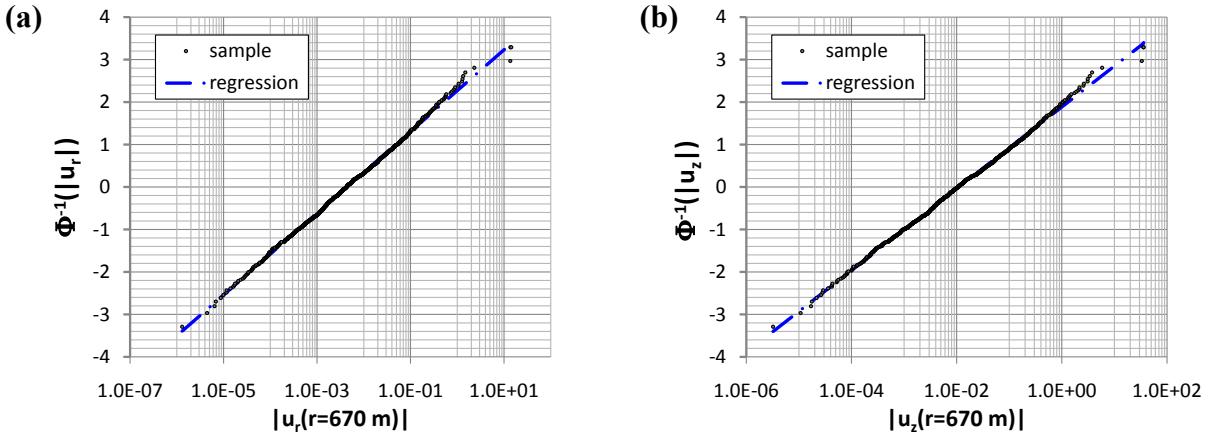

Figure 3. Results of the fit analysis for the sample distribution of (a) the horizontal and (b) the vertical surface displacement at a radial distance of 670 m from the center of the reservoir. The accurate fit between the sampled displacements and the regression line reveals that both u_r and u_z are log-normally distributed.

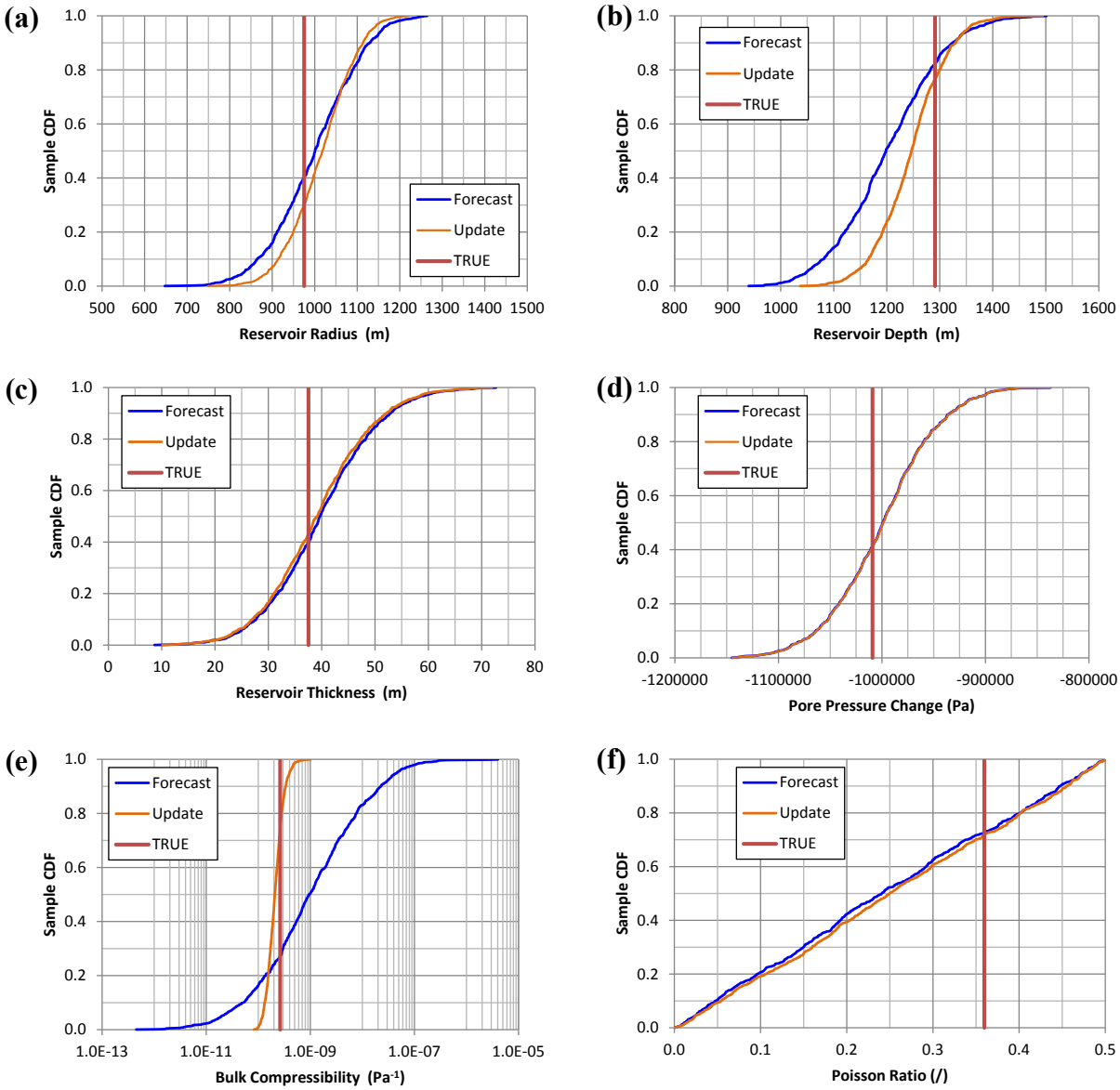

Figure 4. Comparison between the prior (forecast) and posterior (update) sample CDFs for the six reservoir parameters (a) R, (b) c, (c), h (d) Δp, (e) c_M, and (f) ν (Table 2, Scenarios 0 and 2). In each subpanel, the "true" reference value of the corresponding parameter is also shown.

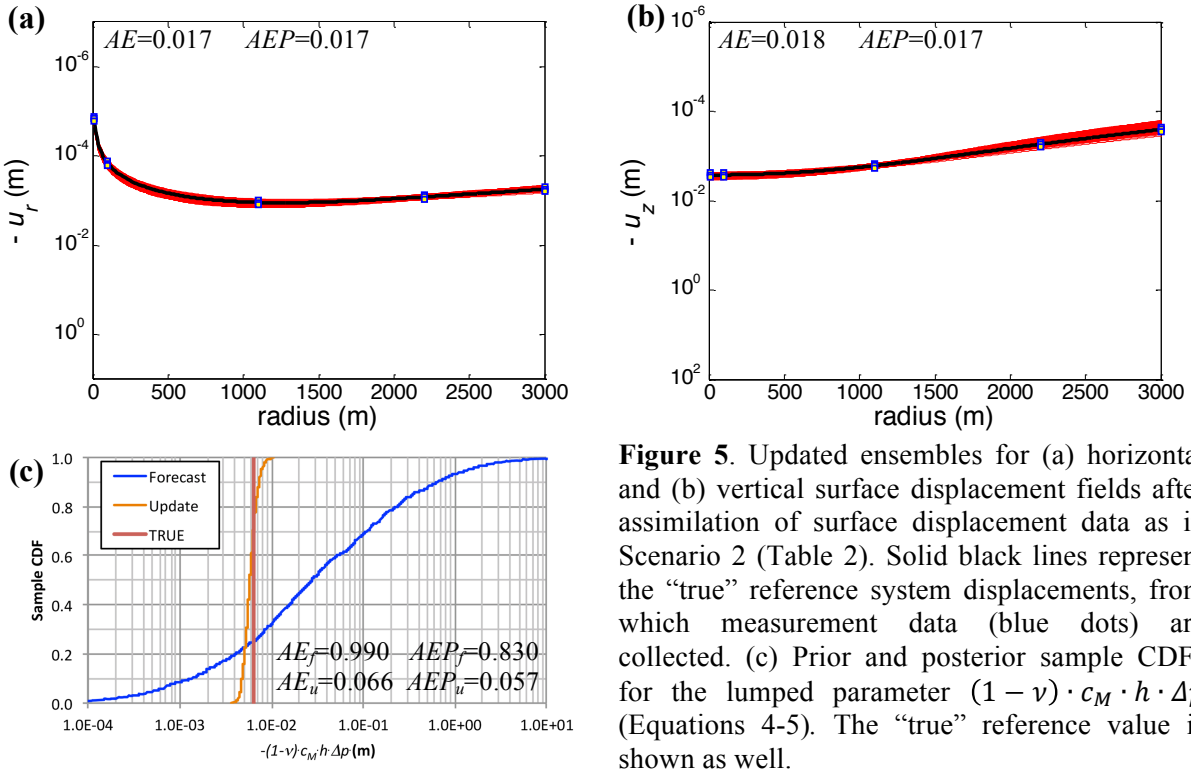

Figure 5. Updated ensembles for (a) horizontal and (b) vertical surface displacement fields after assimilation of surface displacement data as in Scenario 2 (Table 2). Solid black lines represent the "true" reference system displacements, from which measurement data (blue dots) are collected. (c) Prior and posterior sample CDFs for the lumped parameter $(1-\nu)\cdot c_M \cdot h \cdot \Delta p$ (Equations 4-5). The "true" reference value is shown as well.

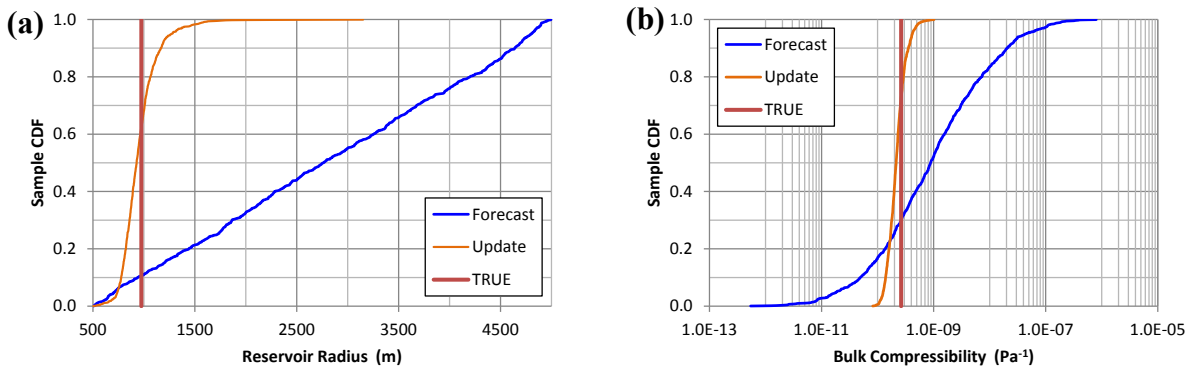

Figure 6. Comparison between the prior (forecast) and posterior (update) sample CDFs for the reservoir parameters (a) R and (b) c_M in the case where the prior uncertainty in the reservoir radius is sampled from a uniform distribution $U(500,5000)$ (m). In each subpanel, the "true" reference value of the corresponding parameter is also shown.

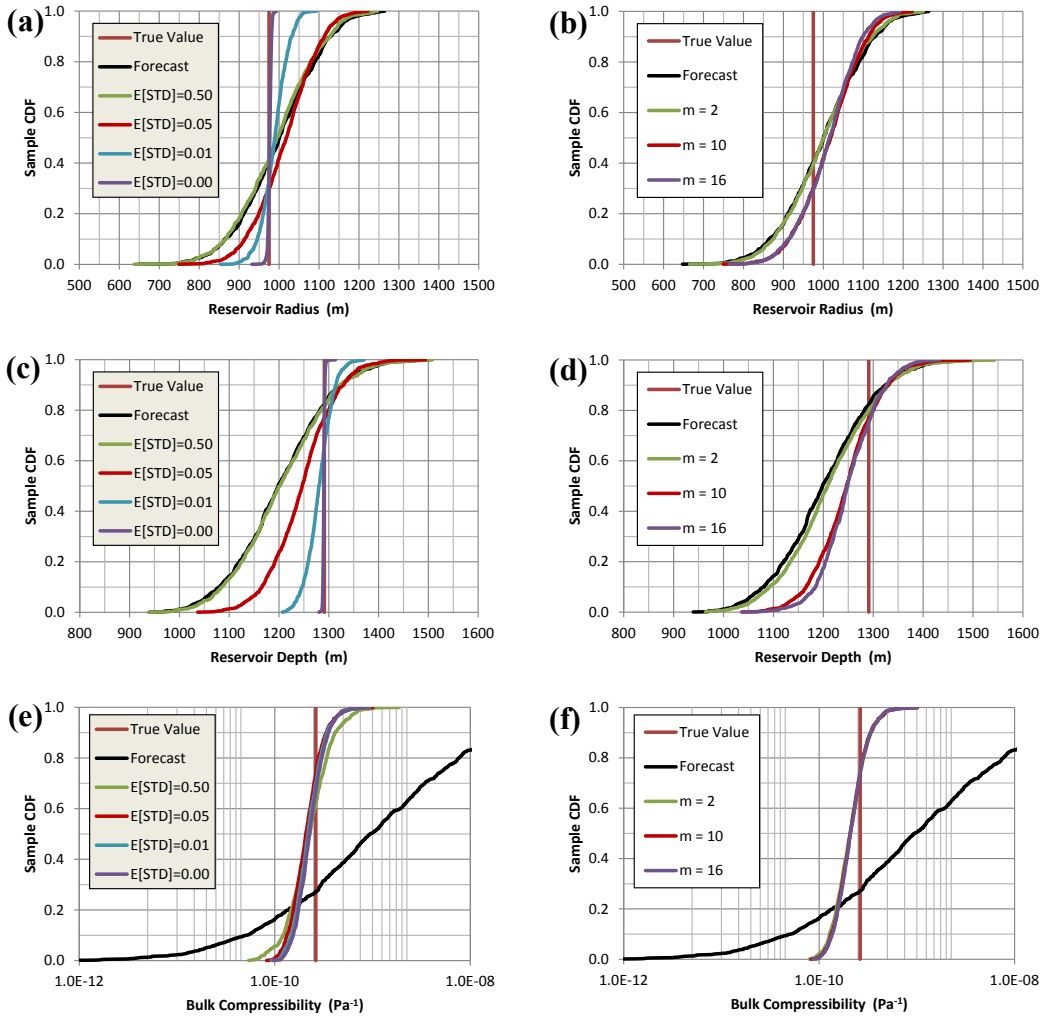

Figure 7. Comparison between the prior (forecast) and posterior (update) sample CDFs for (a-b) the reservoir radius R, (c-d) the reservoir average depth c, and (e-f) the bulk compressibility, c_M, obtained using increasing values of the standard deviation σ of the measurement error (subpanels a-c-e) and increasing values of the number m of the measurements (subpanels b-d-f). In each subpanel, the "true" reference value for the reservoir parameter is also shown.

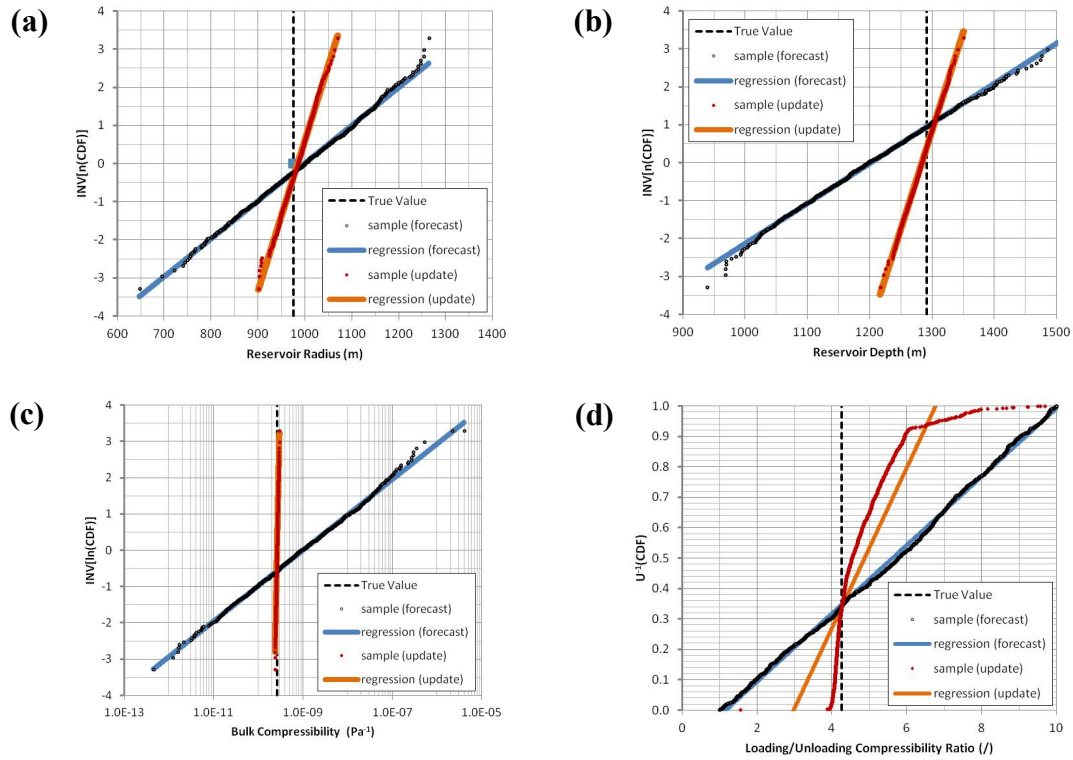

Figure 8. Results of the fit analyses for the sample distributions of (a) reservoir radius R, (b) reservoir depth c, (c) compressibility c_M and (d) the loading/unloading compressibility ratio s. After the ES updating, the statistical distributions of R and c remain normal, the statistical distribution of c_M remains log-normal, whereas the forecast statistics of s is not preserved.

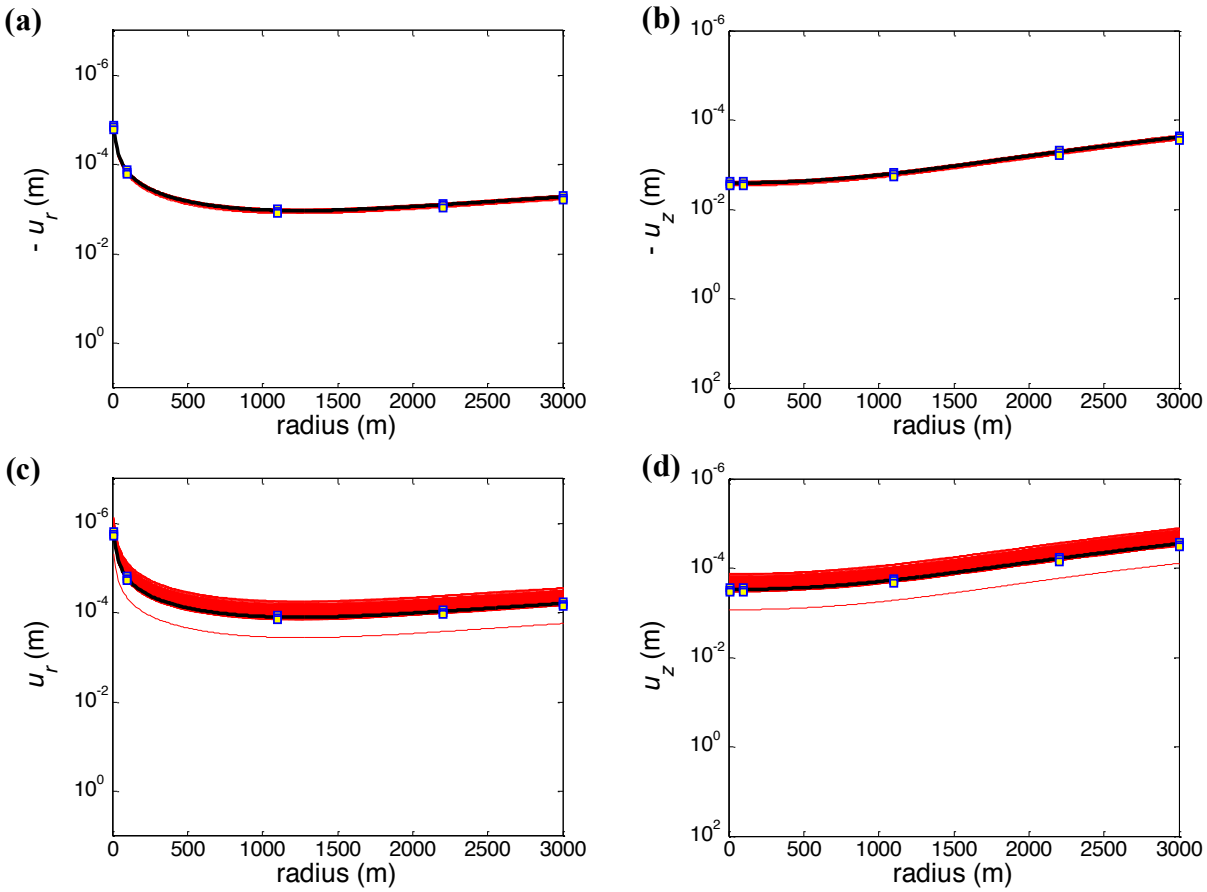

Figure 9. Semi-log plots of the updated ensembles for the horizontal and the vertical land surface surface displacement as a function of the radial distance from the center of the reservoir, obtained at the end of the loading phase (a-b) and after the unloading phase (c-d). In each subpanel, the solid black line represents the reference displacement profile, from which five measurement data, indicated by blue dots, have been collected.