On spatially correlated observations in importance sampling methods for subsidence estimation

Kim, Samantha S.R.; Vossepoel, Femke C.

DOI
10.1007/s10596-023-10264-9

Publication date
2024

Document Version
Final published version

Published in
Computational Geosciences

Citation (APA)
Kim, S. S. R., & Vossepoel, F. C. (2024). On spatially correlated observations in importance sampling methods for subsidence estimation. *Computational Geosciences*, *28*(1), 91-106. https://doi.org/10.1007/s10596-023-10264-9

Computational Geosciences (2024) 28:91–106
https://doi.org/10.1007/s10596-023-10264-9

ORIGINAL PAPER

On spatially correlated observations in importance sampling methods for subsidence estimation

Samantha S. R. Kim[1] · **Femke C. Vossepoel**[1]

Received: 13 October 2022 / Accepted: 16 November 2023 / Published online: 6 December 2023

Abstract
The particle filter is a data assimilation method based on importance sampling for state and parameter estimation. We apply a particle filter in two different quasi-static experiments with models of subsidence caused by a compacting reservoir. The first model considers uncorrelated model state variables and observations, with observed subsidence resulting from a single source of strain. In the second model, subsidence is a summation of subsidence contributions from multiple sources which causes spatial dependencies and correlations in the observed subsidence field. Assimilating these correlated subsidence fields may trigger weight collapse. With synthetic tests, we show in a model of subsidence with 50 independent state variables and spatially correlated subsidence a minimum of $\mathbf{10^{13}}$ particles are required to have information in the posterior distribution identical to that in a model with 50 independent and spatially uncorrelated observations. Spatial correlations cause an information loss which can be quantified with mutual information. We illustrate how a stronger spatial correlation results in lower information content in the posterior and we empirically derive the required ensemble size for the importance sampling to remain effective. We furthermore illustrate how this loss of information is reflected in the log likelihood, and how this depends on the number of model state variables. Based on these empirical results, we propose criteria to evaluate the required ensemble size in data assimilation of spatially correlated observation fields.

Keywords Particle method · Ensemble size · Information theory · Weight collapse · Subsidence · Reservoir

1 Introduction

The Geosciences, specifically the fields of meteorology, oceanography, physical geography, and geophysics, involves the study of complex and nonlinear processes over a range of scales using numerical simulators of varying complexity. Data assimilation is a technique that combines observations with models to estimate model parameters and variables. Parameter values remain constant in time, whereas variable values evolve in time. We use the term state vector for the quantities to be estimated, so the state vector can contain both parameters and state variables. The complexity of a data assimilation system rapidly increases with the number of model state variables and observations and has implications for how we predict the physical processes of the system.

One of the methods used in data assimilation is the particle filter [1] or the particle method [2] for static problems. The particle method that we use is a static application of the particle filter and is in essence an importance sampling method. In the following the particle method refers to importance sampling in static data assimilation problems. Importance sampling, filtering and ensemble-based methods have been applied for subsidence estimation [3, 4]. Other authors [5–7] have used an ensemble smoother (ES) and an iterative ensemble smoother to estimate subsurface geomechanical state variables. An important question in these applications is whether the ensemble spread is sufficiently large to ensure the applicability of the method given the system complexity.

In many applications, the ensemble size is chosen based on trial and error. The particle method, as most other importance sampling methods, suffers from weight collapse, and its performance exponentially degrades as the dimension of the state and observation spaces increases [8–10]. We can prevent this weight collapse by increasing the ensemble size. A more

✉ Samantha S. R. Kim
S.S.R.kim@tudelft.nl

Femke C. Vossepoel
F.C.Vossepoel@tudelft.nl

[1] Department of Geoscience and Engineering, Delft University of Technology, Stevinweg 1, Delft 2628CN, The Netherlands

systematic approach to evaluate the necessary ensemble size has been proposed by [8] and [9] and tested in a practical example through the work of [11]. However, analytic derivations of the ensemble size in these publications are often based on abstract problems, and the results are not always easily translated to geoscience problems [1]. Snyder et al. [8] highlight for example the problem of non-Gaussian prior and observations, and the nontrivial dependencies between state variables and observations. In this study, we will extend the results of [8] to empirical cases with spatially correlated fields of observations, focusing specifically on the feasibility of the particle method and its performance. We illustrate this with an example of subsidence caused by reservoir compaction due to a pressure variation, where observations of subsidence allow us to estimate geomechanical properties of the reservoir [3, 12]. Spatial correlations exist in nearly all geophysical fields and appear in a subsidence field when a single source of subsidence causes a displacement at the surface over a certain region, e.g., a subsidence bowl with a width of several kilometers. In general, a deeper source of subsidence, in the subsurface, creates a subsidence field at the surface with a larger correlation length, and subsidence from a shallower source has a shorter correlation length. Moreover, a spatially correlated subsidence field does not necessarily imply that the measurement errors are correlated since a measurement technique can provide independent measurements and hence uncorrelated measurement errors [13]. In a previous study on the particle filter, [8] derive a theoretical relationship between the maximum weight, $w_{\max}$ of the particles and the dimension of the problem for an example of i.i.d. (independent identically distributed) samples and under the assumption of a standard normal density for the prior and the likelihood. In their study, the authors generalize the results at an asymptotic limit with a linear transformation from model state space to the observation space represented by an I_d (identity) observation operator. Strategies to assimilate spatially correlated observations for a large number of observations are developed by [14]. However, to our knowledge, there is no theoretical approach to estimate the criterion for weight collapse and the required ensemble size in a spatially correlated field, that is when a variable at one location depends on variables at other locations. The objective of the present study is to evaluate the necessary ensemble size that ensures the applicability of the particle method and that avoids weight collapse when 1) the system dimension increases and 2) when the observed field is spatially correlated. For this, we use two conceptual models of subsidence: 1) a one-component model and 2) a multi-component model of subsidence. In the one-component model of subsidence, we use a one-to-one transformation from the model state variables of the reservoir pressure variation to subsidence which gives i.i.d. subsidence values to represent a first-order approximation of a compacting reservoir without spatial correlation. The multi-component model, which includes spatial correlation, is based on the nucleus of strain approach of [15], which has been used in literature to estimate geomechanical reservoir properties [3, 16]. The resulting subsidence shows spatial correlation, as the displacement field is a superposition of subsidence created by a pressure variation in different reservoir compartments. Consequently, the subsidence values are linear combinations of the pressure variation, and therefore not i.i.d.. To derive the ensemble size in problems with spatial correlation, we propose an empirical quantification of the information in both the prior knowledge and the posterior estimate using the formalism of mutual information in information theory.

The paper is organized as follows. In Section 2 we give an overview of importance sampling and of the previous results of [8] for weight collapse in high-dimension. In Section 3 we present the subsidence models and in Section 4, using information theory with the metric of mutual information, we extend the results of [8] to the example of spatially correlated subsidence. Results in Section 5 illustrate how the ensemble size must increase with spatially correlated observations to avoid degradation of the efficiency of the importance sampling. Sections 6 and 7 conclude our study.

2 Importance sampling method

In this section, we first give an overview of importance sampling which is at the base of the particle method, and we introduce the problem of weight collapse.

2.1 Background on importance sampling

The correct description of a physical system can be uncertain when physical processes are not directly observable, for example, processes happening in the subsurface that can only be indirectly observed at the surface. Assuming that we know the probability of a physical process we can approximate its exact probability density function (pdf), i.e., the target pdf, by a discrete distribution. Importance sampling strategies give a sample of N_e particles of model state variables, x, with $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m\}$ sampled from a probability density $p(\mathbf{x}) \sim \sum_{i=1}^{m} w_i \delta(\mathbf{x} - \mathbf{x_i})$, which approximates the exact target density. w_k^i is a scalar weighting each particle and δ the Dirac function. Importance sampling strategies are used in data assimilation to estimate the probability of a model variable or a parameter $\mathbf{x}$ given the observations $\mathbf{y}$ [17–19]. We consider N_e particles obtained with a Monte Carlo sampling of a prior state vector $\mathbf{x}$. Each state in $\{\mathbf{x}_i, i = 1, .., N_e\}$, gives a model representation to be compared with an observation $\{y_i, i = 1, .., N_y\}$, N_y being the number of observations. For the model state vector $\mathbf{x}$, we evaluate the model $\mathcal{M}(\mathbf{x})$

and map it to observations $\mathbf{y}$ via the observation operator $\mathcal{H}$:

$$\mathbf{y} = \mathcal{H}\left[\mathcal{M}(\mathbf{x})\right] + \epsilon, \tag{1}$$

where ϵ represents the measurement error. To estimate the probability density function of the state variables x, the particle method uses a Bayesian approach which provides a probabilistic approach. Its principle is to use the approximated probability density function of the state variables, the prior $p(\mathbf{x})$ which is based on prior knowledge, and to use system observations to infer the uncertainty in $\mathbf{x}$ given $\mathbf{y}$, i.e., estimate the posterior probability density function $p(\mathbf{x}|\mathbf{y})$ with Bayes' formula:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x})}{\int p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x})d\mathbf{x}}, \tag{2}$$

with $p(\mathbf{y}|\mathbf{x})$ the likelihood of the observations, $p(\mathbf{x})$ the prior on the model state variables and the evidence in the denominator, being a normalization factor. Using importance sampling, we use discrete samples of $\mathbf{x}$ to generate the prior $p(\mathbf{x})$. We compare the model predictions $\mathcal{M}(\mathbf{x})$ with the observations $\mathbf{y}$ through the likelihood $p(\mathbf{y}|\mathbf{x})$. The likelihood is chosen as a probability density function which represents the distribution of the uncertainty in data. The posterior distribution $p(\mathbf{x}|\mathbf{y})$ is calculated by weighting each particle with the likelihoods.

$$w_i = \frac{p(\mathbf{y}|\mathbf{x_i})}{\sum_{j=1}^{N_e} p(\mathbf{y}|\mathbf{x_i})} \tag{3}$$

In Eq. 3, the posterior weights are normalized with $\sum_{j=1}^{N_e} w_i = 1$. The expected value of the posterior distribution is an estimator of the state variable values. We later use $\hat{\mathbf{x}}$ to refer to the "analysis", i.e., the posterior quantity in the state space. The quality of the posterior depends on the ensemble size, N_e as well as the dimension of the state- and observation spaces [20]. Moreover, the uncertainty in the analysis depends on the uncertainty in $p(\mathbf{x})$ and in the observations. In Section 4, we introduce the concept of entropy to quantify the quality of the posterior distribution. The entropy, $H(x)$, is minimal if a variable x_i can be estimated with zero uncertainty from $p(\mathbf{x})$, with the probability $p(x_i) = 1$. Conversely, the entropy is maximal and equal to $H(x) = log(N_e)$, if there is equal probability between N_e variables, with $p(x_i) = 1/N_e$.

2.2 Weight collapse in importance sampling

We define the efficacy of importance sampling to estimate the state vector as the ability to get a posterior distribution representing the prior and the likelihood as formulated in Bayes' theorem (Eq. 2). As in any sampling method, the posterior mean becomes a better estimator of the state vector as the number of samples and hence the ensemble size increases. If a given ensemble size is insufficiently large to sample the prior, we can observe a collapse of the weights in the posterior (Eq. 3). Weight collapse occurs when one single particle has almost all the weight in the estimation of the posterior mean. In this case, the maximum weight, $w_{\max}$ of all the particles converges to one. Collapse occurs sooner when the dimensions, N_x, and N_y increase unless the ensemble size increases exponentially as well [9, 10, 21].

Previous results of [8] and [9] give an indication of the applicability of importance sampling, more specifically the particle filter, for a given ensemble size at the asymptotic limit. This asymptotic limit gives the asymptotic condition for collapse for large ensemble size and large number of observations.

In the following, we briefly review the asymptotic theory. For details, we refer the reader to [8] and [9]. The theory of the asymptotic limit is applicable for a case with an i.i.d. prior state vector $\mathbf{x}$ and observation error ϵ, and with an Identity (I_d) observation operator $\mathcal{H}$. If we assume the prior state vector, $\mathbf{x}$, and model $\mathcal{M}(\mathbf{x})$, in the one-component model to be i.i.d, we can apply the asymptotic limit to evaluate the required ensemble size. To compare and understand how weight collapse occurs in the multi-component model with spatial correlation, we address the following question:
Can we identify the required ensemble size, and maximum weight $w_{\max}$, for which importance sampling is practically applicable?

To derive a theoretical relationship between the required ensemble size, N_e, and the dimension N_y, we start with the assumption of standard normal density for the prior and the likelihood.

To derive this relationship, [8] approximate the likelihood of one particle, $p(\mathbf{y}|\mathbf{x}_i)$, as the product of the likelihoods over the observation vector, for an I_d observation operator and with i.i.d. prior state variables and observations error ϵ.

$$p(\mathbf{y}|\mathbf{x}_i) = \prod_{j=1}^{N_y} f\left[y_j - (\mathcal{H}\cdot\mathcal{M}(\mathbf{x}_i))_j\right]. \tag{4}$$

In Eq. 4, for a given observation j, $(\mathcal{H}\cdot\mathcal{M}(\mathbf{x}_i))_j$ is the model prediction from the particle $\mathbf{x}_i$ seen at the observation point j and f is a standard normal density function $f \sim N(0,1)$. The expression of the likelihood in Eq. 4 can be simplified and expressed as a Gaussian distribution, $N(\mu,\tau)$ by defining a re-scaled mean μ and variance τ. To do this, we define a log likelihood V_{ij} using $\Psi() = \log f()$ as follows,

$$V_{ij} = -\Psi\left[y_j - (\mathcal{H}\cdot\mathcal{M}(\mathbf{x}_i))_j\right], \tag{5}$$

s where, using the expression of the log likelihood V_{ij} in Eq. 5, the expression of the likelihood in Eq. 4 can be rewrit-

ten as a re-scaled likelihood of $p(\mathbf{y}|\mathbf{x}_i)$

$$p(\mathbf{y}|\mathbf{x}_i) = \exp(-\mu - \tau S_i). \tag{6}$$

Here, μ is a re-scaled mean and the variance is τ. These are defined as a function of the log likelihood V_{ij}

$$\mu = \sum_{j=1}^{N_y} E(V_{ij}) \qquad \tau^2 = \text{var}\left(\sum_{j=1}^{N_y} V_{ij}\right). \tag{7}$$

An important condition for the expression of the likelihood in Eq. 6 to be a valid approximation of the Eq. 4 is that the scale factor S_i can be approximated by a standard normal density $S_i \sim N(0, 1)$. From Eqs. 6 and 7, we express S_i as a function of the log likelihood, μ and τ and later verify its Gaussianity with the following expression. S_i is now

$$S_i = \left(\sum_{j=1}^{N_y} V_{ij} - \mu\right)\tau^{-1}. \tag{8}$$

Interestingly, [8] emphasize that in the example of a one-component model, the expression of S_i does not directly depend on the dimension, N_x, of the model state vector to be estimated.

Using the expression of the re-scaled likelihood (Eq. 6) and with the approximation $S_i \sim N(0, 1)$, we can now connect the maximum weight, $w_{\max}$, to the dimension N_y, and to the ensemble size N_e, in the case of standard normal distribution for the likelihood and the prior. With this, we find

$$E[1/w_{\max}] - 1 \approx \sqrt{\frac{4}{5}}\sqrt{\frac{\log N_e}{N_y}}. \tag{9}$$

In this expression, the maximum weight, $w_{\max}$, is related to S_i through Eq. 3. [8] use the convergence properties of a $w_{\max} \to 1$ and $S_i \sim N(0, 1)$ for large N_e and N_y to derive the asymptotic expression (Eq. 9). Equation 9 is valid for a Gaussian prior and likelihood under the assumption of S_i converging to a Gaussian distribution.

An important result of [8] is that from Eq. 4 and from the expression of the log likelihood in Eq. 5, the likelihood only depends on the dimension N_y, meaning that the observation dimension, rather than the state dimension, control the weight collapse. The approximation in Eqs. 6 and 7 are valid only if the prior distribution of state variables is close to a Gaussian distribution.

3 Subsidence models

Subsidence can occur as a consequence of reservoir compaction, which is a decrease in reservoir thickness due to a pressure variation, as a result of, for example, gas production. When reservoirs are compartmentalized by faults or in the case of different rock compositions, the reservoir may compact more strongly in certain areas. We discretize the reservoir into 5.5km×5.5km×240m cuboid-shaped elements, with a constant reservoir thickness of 240m (Table 1). In the following, we will use the term *compartment* to refer to reservoir elements. These could be geological reservoir compartments, or volumes within the reservoir that have the same reservoir property. As a simplification, we first simulate the compaction of each compartment, without the created subsidence affecting the surface above other compartments. This is illustrated in Fig. 1a with the "one-component model". In contrast to the one-component model, Fig. 1b shows spatially correlated subsidence created from the cumulative effect of the compaction in all compartments. To create this spatially correlated subsidence, we use the "multi-component model", in the sense that the resulting subsidence is a linear combination of all pressure variations. A commonly used model for subsidence as a result of pressure variation is the nucleus of strain approach of [15, 22, 23]. The nucleus of strain represents a compaction of a reservoir compartment as a point source of pressure variation, ΔP. The analytical solution of the vertical displacement at the surface ($z = 0$) created by a single nucleus of strain is

$$u_z(r, 0) = \frac{-C_m(1-\nu)V\Delta P}{\pi}\frac{D}{\left(r^2 + D^2\right)^{3/2}}, \tag{10}$$

where r is the horizontal distance between an observation point at the surface and the vertical of a nucleus of strain at a depth D and with the compaction coefficient of the reservoir, C_m, the Poisson ratio ν and the volume V of the reservoir, the pressure variation ΔP. In the data assimilation experiment with this subsidence model the pressure variation ΔP for each compartment, $\mathbf{x} = \{\Delta P_i, i = 1, .., N_x\}$ are the unknown variables that are being estimated. These variables form the state vector.

3.1 One-component model of subsidence

The one-component model gives a first approximation of the subsidence caused by the pressure variation (and associated reservoir compaction) and does not take into account the spatial correlation in the subsidence field. To build this model, we start with a 1D geometrical approximation of subsidence (Fig. 1a), and we create adjacent and independent compartments a the reservoir depth with Eq. 10. This gives a discretization of 1D columns of subsurface with the reservoir layer and a 5.5km × 5.5km resolution for subsidence at the Earth's surface (Fig. 1). Because we don't include spatial correlation in the one-component model, an observer at one point only sees the compaction in the reservoir compartment

directly below (Fig. 1), resulting in the horizontal distance r of the nucleus of strain to the observation point, r $= 0$, such as that we have a vertical displacement $u_z(r = 0, 0)$.

In this case, the model $\mathcal{M}$ represents a mapping from a pressure variation of the reservoir to a vertical displacement of the surface with a one-to-one relationship between pressure variation state variables and subsidence. The model state variables are independent hence the the subsidence values associated with each compartment are also independent, resulting in an uncorrelated subsidence field. The number of columns gives the model resolution, which in our case defines the dimension of the state and the observation space. This model is similar to the example in [8] with an I_d observation operator, $\mathcal{H} = I_d$.

3.2 Multi-component model of subsidence

The one-component model of subsidence computes a local displacement caused by a single nucleus and the resulting subsidence field does not include the response from the entire compacting reservoir. As an alternative, the nucleus of strain approach is able to model an arbitrarily shaped reservoir [22–24] by linearly adding the effect of each nucleus k to the total displacement field

$$\mathbf{u}_z(r, 0) = \sum_{k=1}^{N_x} \mathbf{u}_{z,k}. \tag{11}$$

The geometry of the multi-component model is similar to the one-component model of subsidence with a nucleus of strain in the center of each reservoir compartment. We compute the subsidence (Eq. 10) by calculating the influence of the nucleus of strain over the volume, V, of the reservoir compartment [25]. The difference with the one-component model is that the surface displacement u_z at one location j in space arises from all sources of strain in all reservoir compartments (Fig. 1b). Therefore, $\mathcal{M}$ transforms pressure variation to subsidence and creates a spatially correlated subsidence field at the surface. The spatial correlation in the multi-component model thus implies that the model $\mathcal{M}(\mathbf{x})_j$, computed for an observation point j from the state vector $\mathbf{x}$ can be written as

$$(\mathcal{M}(\mathbf{x}))_j = \sum_{k=1}^{N_x} m_{jk} x_k, \tag{12}$$

where m_{jk} is the jk^{th} element of the model matrix. Recall that the multi-component model integrates the responses of all compartments, and thus is not a one-to-one transformation from the pressure variation to the subsidence (i.e., it is a non-injective transformation), and, in this case, an observation j is linearly dependent to all $k = 1, ...N_x$ components of the state vector. As described in Eq. 1, the observation operator remains $\mathcal{H} = I_d$ as in the one-component model, assuming that the measurement method is the same. In the following, we use these two models of subsidence to investi-

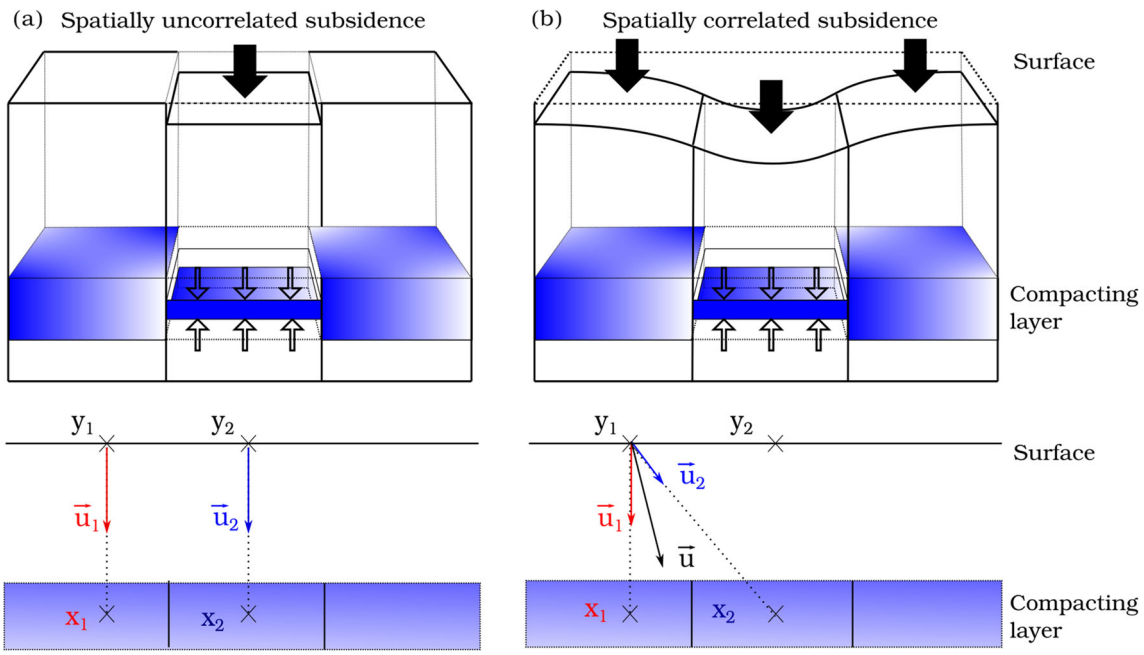

Fig. 1 Models of subsidence. (a) One-component model of subsidence and (b) the multi-component model of subsidence with spatial correlation. 3D models are built with an ensemble of 1D columns. While the one-component model simulates subsidence only for the surface directly above the compartment, the multi-component model computes the deformation field as an integrated response from all compartments

Table 1 Pressure variation, compartment size and thickness of the multi-component model

	ΔP	dx	dy	dh
Nucleus of strain	-0.35MPa	0.05°	0.05°	240m

gate the effect of spatial correlation in weight collapse. The pressure variation, the compartment size, and the thickness are modeled after the Groningen gas reservoir, in the Netherlands (Table 1). The mechanical properties are also taken from this reservoir (Table 2).

In the following we estimate the pressure variation in each of the compartments using observations of subsidence, defining the state vector with $\{\mathbf{x}_k,\ k = 1, .., N_x\}$, N_x being the number of compartments and thus the state space dimension. We set the number of observations N_y, equal to the number of state variables to test the particle method with the two models of subsidence at the asymptotic limit (Section 2.2) and the agreement with the theoretical derivation of weight collapse from [8].

3.3 Synthetic experiments for subsidence models

With synthetic data assimilation experiments, we can assess the efficacy of the particle method to estimate an unknown quantity. The state vector $\mathbf{x}$ of dimension N_x represents the unknown pressure variation, ΔP, that we want to estimate for each grid cell of the reservoir. A pressure variation ΔP in the reservoir creates reservoir compaction and subsidence. We simulate this by applying the forward model operator, $\mathcal{M}$, and with data assimilation we update the distribution of the state variables ΔP given subsidence observations $\mathbf{y}$. Both state and observation vectors $\mathbf{x}$ and $\mathbf{y}$ have values spatially distributed over a regular grid and have the same dimension $N_x = N_y$. The pressure variation and therefore the subsidence can vary in time, usually involving a dynamical forward model $\mathcal{M}$. In the study, we test the method over one-time step considering a quasi-static case. To do this we define a "true" value of state variables, $\mathbf{x}^{truth}$ and sample values from this truth to generate synthetic observations for assimilation $\tilde{\mathbf{y}}, \{\tilde{y}_i, i = 1, .., N_y\}$ with Eq. 1. We use the notation $\tilde{\mathbf{y}}$ to define synthetic observations and avoid confusion with the vector observations, $\mathbf{y}$, in Eq. 1. The observation operator $\mathcal{H}$ maps the true state vector $\mathbf{x}^{truth}$ to the synthetic observations $\tilde{\mathbf{y}}$ and Gaussian noise, ϵ, is linearly added to simulate imperfect observations:

$$\tilde{\mathbf{y}} = \mathcal{H}\left[\mathcal{M}(\mathbf{x}^{truth})\right] + \epsilon. \tag{13}$$

We use the observation operator $\mathcal{H}$, applied to the forward modeling operator $\mathcal{M}$, to compute the model prediction of subsidence from the pressure variation. Note that in the example with synthetic experiments, we use $\mathcal{H} = I_d$, however $\mathcal{H}$ could differ given the origin of the empirical data, e.g., leveling or InSAR techniques. We compare the outcome of synthetic experiments with the "true" values of state variables and subsidence estimates as an indication of the efficacy of the importance sampling. We expect a different required ensemble size, N_e, larger for the multi-component model than for the one-component model. To understand this, let's assume that the one-component model requires N_e Monte Carlo samples to solve N_y independent equations of 1 variable each. One particle is then a vector of N_y values and we solve a system of N_y independent equations, which is relatively straightforward. Now if the N_y equations each contain linear combinations of all N_y variables, as it is in the multi-component model, the equations are no longer independent and it becomes more difficult to solve this problem with N_y samples. Therefore, we expect the ensemble size N_e, to increase when the subsidence in one location depends on pressure variations in other locations, as is the case in the multi-component model.

Table 2 Reservoir properties

Parameter	Symbol	Value	Unit
Depth	D	2800	m
Thickness	h	240	m
Poisson ratio	ν	0.32	-
Compaction coefficient	C_m	1×10^{-10}	Pa^{-1}

4 Entropy and mutual information

We investigate weight collapse in data assimilation problems with spatially correlated observations using the information theory of Shannon [26] which is commonly applied in probability theory in the field of dynamical systems [27]. The measure of entropy quantifies the uncertainty, and thus the information about an unknown quantity. Yustres et al. [28–31] demonstrate the use of information theory and of mutual information in Bayesian estimation problems. Nearing et al. [32] further use entropy and mutual information to evaluate the efficiency of a data assimilation method. For a particle method, sources of uncertainty are model error, imperfect data, and the approximation in the importance sampling algorithm which all propagate into the posterior distribution. The importance sampling algorithm assumes that we can adequately sample the prior. However, if the sampling is inadequate, for example, because of non-independent or insufficient samples, weight collapse can occur. In this case, the expected mean of the posterior is no longer a good representation of the true posterior (Eq. 2). This implies a loss of

information in the posterior estimate. To evaluate the propagation of information from prior and likelihood to posterior, we use the metric of entropy and mutual information as defined in [33] and [34]. Mutual information describes how much information two arguments have in common, similar to a correlation between two random variables normally distributed. For example, it can give the common information between the random variable x contained in the distribution $p(\mathbf{x})$, representing the model state variables and the data for assimilation $\mathbf{y}$, with distribution $p(\mathbf{y})$ (Fig. 2). Using mutual information in data assimilation problems allows us to evaluate how the information before assimilation is conserved in the posterior. For a state variable x sampled from a discrete distribution $p(\mathbf{x})$, the entropy $H(x)$ of $p(\mathbf{x})$, can be interpreted as "can we know x given $p(\mathbf{x})$" and is expressed by

$$H(x) = -\sum_i p(x_i) \log\left(p(x_i)\right). \tag{14}$$

Let us introduce random variables x, y, and z. In assimilation problems, we have samples of the model predictions, $\mathbf{x}$ and we have the observation vector $\mathbf{y}$ for the assimilation (Eq. 1). In the case of a synthetic experiment the analysis can be compared with the truth. However, in realistic data assimilation problems, we don't know the truth, and therefore, to test the performance of the data assimilation, we compare the analysis with independent observations, which we refer to as validation data. Let us assume that $\mathbf{z}$ is a vector of validation data of dimension N_y. By construction, observation vectors $\mathbf{y}$ and $\mathbf{z}$ are sampled from the same density, $N(0, 1)$, and the same observation model (Eq. 1). We assume them to be independent. $H(x)$ is the entropy in the model and it measures the uncertainty on the model predictions $\mathbf{x}$. Similarly, for the entropy in a data set, we can compute $H(y)$, from observations $\{y_i,\ i = 1,\ ...,\ N_y\}$ with Eq. 14.

We can now compute the mutual information $I(z; x)$ between $p(\mathbf{z})$ and $p(\mathbf{x})$

$$I(z; x) = \sum_z \sum_x p(\mathbf{x}, \mathbf{z}) \log\left(\frac{p(\mathbf{z}, \mathbf{x})}{p(\mathbf{z})p(\mathbf{x})}\right). \tag{15}$$

In Eq. 15, the prior distribution $p(\mathbf{x})$, of the model predictions $\mathcal{M}(\mathbf{x})$, is compared with the validation data $\mathbf{z}$, using the joint probabilities $p(\mathbf{z}, \mathbf{x})$.

As mentioned in Section 3.2, the state and the observation spaces have the same dimension, $N_x = N_y$, thus the vector of the model prediction has the dimension of N_x. A more practical computation of the mutual information [35], written as a function of the entropy and the joint entropy $H(z, x)$ is

$$I(z; x) = H(x) + H(z) - H(z, x). \tag{16}$$

Also Eq. 16 can be applied to estimate the mutual information $I(z; y)$. $I(z; x)$ and $I(z; y)$ now give an indication

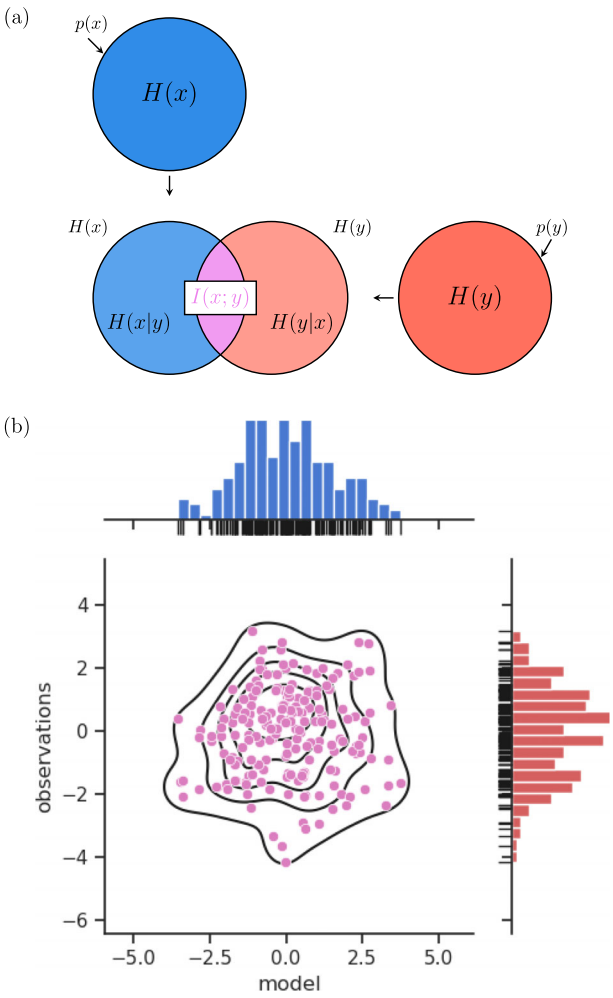

Fig. 2 Venn diagram to illustrate the metric of entropy before assimilation. (a) The colored area in the circle represents the entropy of a distribution (e.g., entropy $H(x)$ of the prior distribution $p(\mathbf{x})$). In this drawing, the intersection of the entropy of prior and data visualizes the mutual information I(x; y). The remaining area of entropy (e.g., $H(x|y)$) represents the reduced uncertainty on the variable x given the knowledge brought by the variable y. (b) Joint probability gives an intuitive approach to mutual information as a measure of similarity between probability distributions. In the example with discrete Gaussian distributions for the prior and the data, if the joint probability shows a strong correlation then the mutual information increases as the joint probability increases (Eq. 15)

of the information content in the model variable x and the assimilated data y, which can be used to evaluate a certain data-assimilation setup. By quantifying the propagation of information content before and after assimilation we can relate weight collapse to the importance sampling performance and therefore evaluate the required ensemble size given the model complexity.

To evaluate how the information content before assimilation propagates to the posterior we define the differential information I_{diff} and the data assimilation efficiency $\mathcal{E}_{DA}$ [33]. The differential information is the difference between

the mutual information in the posterior and in the prior and is expressed as $I_{diff} = I(z;\hat{x}) - I(z;x)$. This quantity can be negative if the particle method *corrupts* the prior information available before assimilation (e.g., in the model or in the observations). To describe how the information from both the prior and observations is conserved in the posterior distribution, we define the data assimilation efficiency $\mathcal{E}_{DA}$. $\mathcal{E}_{DA}$ is the ratio of the posterior mutual information $I(z;\hat{x})$ and the total information in model and observations $I(z;x,y)$,

$$\mathcal{E}_{DA} = \frac{I(z;\hat{x})}{I(z;x,y)} \tag{17}$$

The posterior mutual information represents the mutual information between the distribution of the validation data $p(\mathbf{z})$ and the analysis estimate, evaluated from the posterior weighted ensemble $p(\hat{\mathbf{x}})$. The total information in model and observation $I(z;x,y)$ represents the information available before assimilation and is expressed as a function of the mutual information $I(z;\hat{x})$ and of the mutual conditional information $I(z;y|x)$:

$$I(z;x,y) = I(z;x) + I(z;y|x). \tag{18}$$

$I(z;y|x)$ is the information that z and y have in common conditioned on the prior $p(\mathbf{x})$ and can be calculated as follows in order to derive $\mathcal{E}_{DA}$:

$$I(z;y|x) = \sum_x \sum_y \sum_z p(\mathbf{x},\mathbf{y},\mathbf{z}) \log\left(\frac{p(\mathbf{y}|\mathbf{z},\mathbf{x})}{p(\mathbf{y}|\mathbf{x})}\right). \tag{19}$$

The Posterior quantity of mutual information computed in the same manner as that of the prior information in Eq. 16. The entropy of the posterior distribution $p(\mathbf{x}|\mathbf{y})$ is proportional to the weights $w_i \log(w_i)$, and can be derived in the discrete case with the approximation of

$$p(\mathbf{x}|\mathbf{y}) = \sum_{i=1}^{N_e} w_i \delta(x - x_i)$$

and the expression of the updated weights in Eq. 3. A maximum entropy represents a maximum uncertainty on the analysis $\hat{x}$, and likewise, a decreasing entropy suggests a decreasing uncertainty. Quantities of mutual information are empirically estimated with histogram estimation methods. Algorithm 1 gives a pseudo algorithm to evaluate the prior information before assimilation $I(z;x)$ (Eq. 16) and the posterior information $I(z;\hat{x})$. Other quantities of mutual information are computed in the same manner.

Algorithm 1 Derivation of the mutual information in synthetic experiments.

1) Prior information $I(z;x)$
→ Sample $N_x = 50$ state variables of pressure variation x_i
→ Compute subsidence at $N_y = 50$ observation points $x_i = \mathcal{M}(x_i)$
→ Generate a synthetic observation vector of dimension $N_y = 50$ with $\tilde{y}_i = \mathcal{H}\left[\mathcal{M}(x_i^{truth})\right] + \epsilon$ and the validation data $\tilde{z}_i = \mathcal{H}\left[\mathcal{M}(x_i^{truth})\right] + \epsilon$.
for $n = 1 : nbin$ **do**
  Create histograms of x and $\tilde{z}$, *nbin* being the number of bins
  Compute the probability of $p(x)$, $p(\tilde{z})$
  Evaluate the entropy of $H(x)$, $H(\tilde{z})$
end for
→ Compute the mutual information $I(\tilde{z};x)$ with Eq. 16.
1) Posterior information $I(\tilde{z};\hat{x})$
for $n = 1 : nbin$ **do**
  Create the histogram of x given the weight vector from the assimilation and get $p(\hat{x})$ per bin.
end for
→ Compute $I(\tilde{z};\hat{x})$ in the manner of Eq. 16.

5 Subsidence state estimation

5.1 Weight collapse and asymptotic limit for subsidence models

We observe a stronger weight collapse in the case of spatial correlation in the observation field. Figure 3 illustrates the weight collapse in the posterior distribution in the subsidence models of this study. The histograms show the distribution of the maximum weight $w_{\max}$, for both the one-component and the multi-component models of subsidence and for a model dimension of $N_x = [10,\ 30,\ 100]$. The experiment is repeated 1000 times for a consistent estimation. The maximum weight can be used as an indicator of the importance sampling performance. It shows that the method performance can rapidly decrease with spatially correlated observations if the ensemble size is inadequate. To evaluate weight collapse in the case of spatial correlation and an increasing dimension of the state, N_x, and observations spaces, N_y, we test the particle method at the asymptotic limit (Eq. 9). We perform experiments with both the one-component and the multi-component models with a large number of observations $N_y = [600, 800, 1000, 1200, 1400, 1600, 1800, 2000]$. In each experiment, we assimilate synthetic observations of subsidence into ensembles with $N_e = N_y^n$ with $n = [0.75, 0.875, 1.0, 1.25]$. Not surprisingly, the results of the one-component model (Fig. 4a and b) show a good accordance between simulations and the theoretical prediction of $E[1/w_{\max}]$ from Eq. 9. Connecting the maximum weight $w_{\max}$, to the ensemble size N_e, and the number of state variables N_x, through a linear relationship. A linear interpolation of the results suggests that the line of

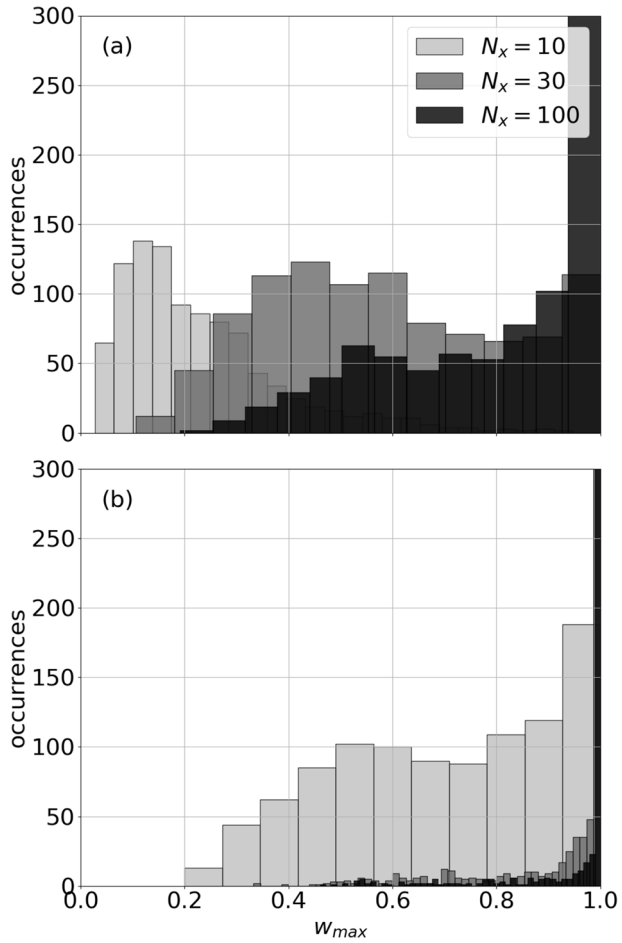

Fig. 3 Maximum weight $w_{\max}$ for the one-component model of subsidence (a) and for the multi-component model of subsidence with spatial correlation (b) from simulations with, N_x=[10, 30, 100] state variables and an ensemble size $N_e = 1000$. Histograms of $w_{\max}$ are computed over 1000 simulations

$E[1/w_{\max}] - 1 = -0.008 + 1.0721\sqrt{\log(N_e)/N_y}$ describes the variability of the maximum weight which exponentially depends on N_e. The particle method in the one-component model of subsidence shows good agreement with the results of [8]. However, it clearly appears that the particle method in the multi-component model does not fit the theoretical prediction of Eq. 9. The main difference between the one-component model and the multi-component model is reflected by the negative log likelihood (Eq. 5) that results from the summation in the model operator $\mathcal{M}$ (Eq. 12). If we assume a Gaussian density for the expression of the negative log likelihood in the one-component model

$$V_{ij} \propto \frac{1}{2}\frac{\left[y_i - (\mathcal{H}[\mathcal{M}(\mathbf{x}_i)])_j\right]^2}{\sigma^2}, \tag{20}$$

this becomes

$$V_{ij} \propto \frac{1}{2}\frac{\left[y_j - \mathcal{M}(x)_{ij}\right]^2}{\sigma^2}. \tag{21}$$

In the case of spatial correlation, the expression of the negative log likelihood differs as we take into account the linear combinations of state variables (Eq. 12). Because of the mapping from the model input to the model output reflected by $\mathcal{M}$, the log likelihood depends on the state space dimension N_x (Eq. 22) in the case of spatial correlation. This can be explained by the fact that in the case of spatial correlation, we consider the differences between the vector observation y_j with the model computed from all model state variables.

$$V_{ij} \propto \frac{1}{2}\frac{\left[y_j - \sum_{k=1}^{N_x} m_{jk}.x_{ik}\right]^2}{\sigma^2}. \tag{22}$$

To assess how the weight collapse in the assimilation with the multi-component model (Eq. 22) differs from the theoretical prediction of weight collapse in Eq. 9, we test the approximation of the observation likelihood in Eq. 6 against the probability distribution of the term S_i in Eq 8. The term S_i is the scale factor that allows us to express the re-scaled likelihood (Eq. 4) as a Gaussian distribution. The main assumption to re-scale this likelihood is that S_i follows a standard normal density, $S_i \sim N(0, 1)$.

The histograms in Fig. 4 show the distributions of S_i in the one-component and the multi-component model. The distribution of S_i in the multi-component model shows a skewness compared to the result for the one-component model, which approaches a standard normal distribution. Distributions of S_i provide evidence of the effect of state variables dependency in Eq. 22 [11].

The histograms in Fig. 4 show the result of S_i for a single simulation. To confirm the deviation to a standard normal density in the distribution of S_i, we perform the K-S (Kolmogorov-Smirnov) tests over 1000 simulations (Table 3).

The analytical cdf (cumulative density function) of the standard normal distribution and the cdf of sampled distributions from the standard normal density, $N(0, 1)$ are first compared to evaluate the spread around the analytical solution (Fig. 4e and f), due to the sampling. For a dimension $N_x = N_y = 1000$ and the ensemble size of $N_e = 1000$, we compute the value of S_i (Eq. 8), and compare to the analytical and sampled cdf. We choose the values of N_x, N_y, and N_e to perform simulations of S_i at the asymptotic limit, in a regime where Eq. 9 is valid.

Results in Fig. 4e and f, show a very good overlap of the cdf for the one-component model for both the sampled distributions and the S_i distributions. K-S tests confirm the main assumption that S_i is approximately normal in the case of the one-component model of subsidence.

The K-S test for the multi-component model exhibits a skewness in the cdf of S_i, corresponding with the deviation from a standard normal density in the histogram

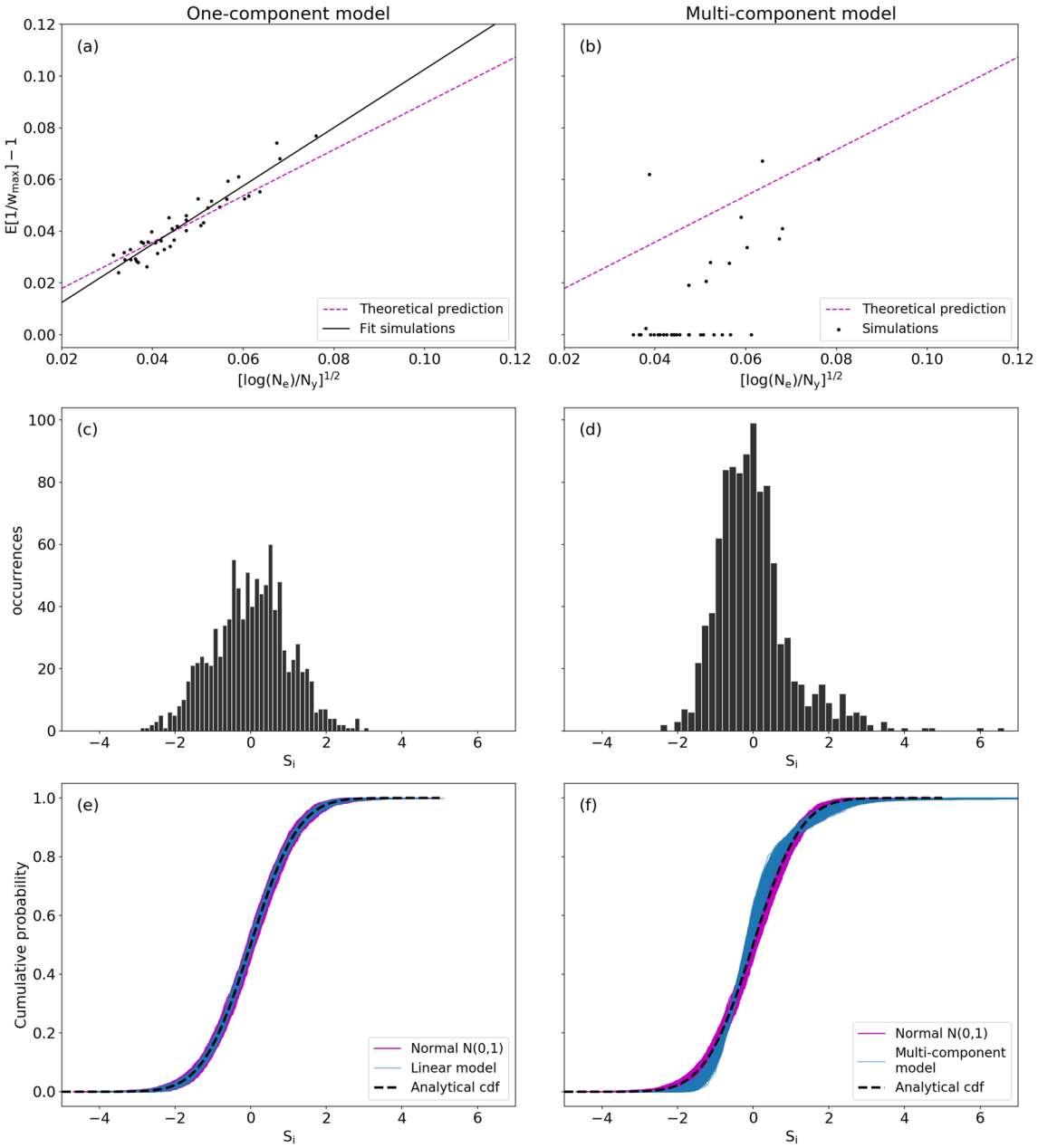

Fig. 4 Results at the asymptotic limit for the one-component model of subsidence and the multi-component model. Verification of the condition for collapse ((a), (b)). The dashed line gives the theoretical prediction of Eq. 9 and the solid line gives the best fit to the simulations given N_x and N_e. The histograms of the distribution of the scale factor S_i from the re-scaled likelihood in Eq. 8 ((c), (d)). In (e) and (f), Kolmogorov-Smirnov (K-S) tests of the distribution of S_i show the cumulative probability distributions over 1000 simulations with 1) a sampled standard normal density $N(0, 1)$, 2) the subsidence model and in dashed line the analytical cumulative probability of a density $N(0, 1)$

Table 3 K-S test results of the S_i distribution in the one-component and the multi-component model. S_i distributions are compared to Gaussian distributions and average over 1000 simulations

	KS statistic	p-value
Gaussian	0.032	0.66
S_i: one-component	0.034	0.60
S_i: multi-component	0.095	0.015

(Fig. 4d). We have computed the K-S test statistic and the p-values between two distributions for the cases: Gaussian/Gaussian, Gaussian/S_i one-component, and Gaussian/S_i multi-component. The result is an average of the comparison of the S_i distribution against 1000 Gaussian distributions: Results show a lower p-value for the multi-component model than for the one-component model, confirming the deviation from the Gaussian distribution. Most of the p-values are

below the threshold of 0.05 usually taken as resemblance criterion. This implies that the Gaussian approximation of S_i is not valid in the multi-component model. Results with the multi-component model thus suggest that S_i can not always be approximated by a standard normal distribution when the observations are spatially correlated and consequently when the log likelihood explicitly depends on the state dimension, N_x in Eq. 22. Implications of non-Gaussian S_i are that 1) the re-scale likelihood in Eq. 6 is not a valid approximation and 2) the relationship between the maximum weight and the required ensemble size is not linear (Figs. 4a and b). With these empirical results, we highlight a limit of the analytical derivation of the required ensemble size in a spatially correlated data assimilation problem.

5.2 Entropy and mutual information for subsidence models

With $H(z)$ being the uncertainty about a set of observations, with the metric of mutual information we evaluate the information content in the model, $I(z; x)/H(z)$, and in the data $I(z; y)/H(z)$ that resolves the uncertainty on z. We apply the same methodology for both the one-component and the multi-component model. With synthetic data assimilation experiments, we compute the entropy and the mutual information with a histogram method. The histogram method requires a sufficient sample size to evaluate the probability of the sample into bins. The sample size in this simulation is the number of model predictions, N_x, and the number of observations N_y, respectively for $I(z; x)/H(z)$ and $I(z; y)/H(z)$. The binwidth is chosen such that the histogram covers the range of the values of subsidence (e.g., for both model predictions and synthetic observations). For our experiments with the dimension of $N_x = N_y = 50$, sensitivity tests on the robustness of the histogram method led us to choose a bin resolution of 0.5mm to create histograms of subsidence. Table 4 shows the mutual information for the one-component and the multi-component model before the assimilation. The information content in both the model and in the data is 0.76. The one-component and the multi-component model have a similar information content before assimilation, which is not surprising, as both have been sampled from the same, Gaussian distribution. We chose a relatively small $N_x = N_y$ of 50, which is sufficient to make sure that we avoid ensemble collapse. This ensemble size is used to reproduce the results of entropy and mutual information with the histogram estimation and we keep the same $N_x = N_y$ of 50 in this study to compare results. The importance sampling experiments are performed for an ensemble size $N_e = 1000$. As expected from histograms of $w_{\max}$ in Fig. 3, the weight collapse is stronger in the multi-component model than in the one-component model (Table 4) as the dimension increases. This suggests that the importance sampling itself is the main cause of information loss in the posterior in case of weight collapse. The posterior entropy, $H(\hat{x})$ in Fig. 5, which represents the uncertainty on the vector of state variables $\mathbf{x}$, after the assimilation in the posterior distribution $p(\mathbf{x}|\mathbf{y})$, is computed for the dimensions $N_x = [10,\ 30,\ 100]$. Comparing Fig. 5 with Fig. 3, we observe that the posterior entropy decreases as the maximum weight increases. Comparison of Fig. 5b with Fig. 5a shows that the posterior entropy converges faster to zero in the case of a model with correlation than in the case of the one-component model. Similarly to the

Table 4 Prior information content in model and in data with a bin resolution of subsidence of 0.5mm and 40 bins for $N_x = N_y = 50$

	One-component	Multi-component
I(z; x)/H(z)	0.76	0.76
I(z; y)/H(z)	0.76	0.77
$w_{\max}$ ($N_e = 10^3$)	0.71	0.94

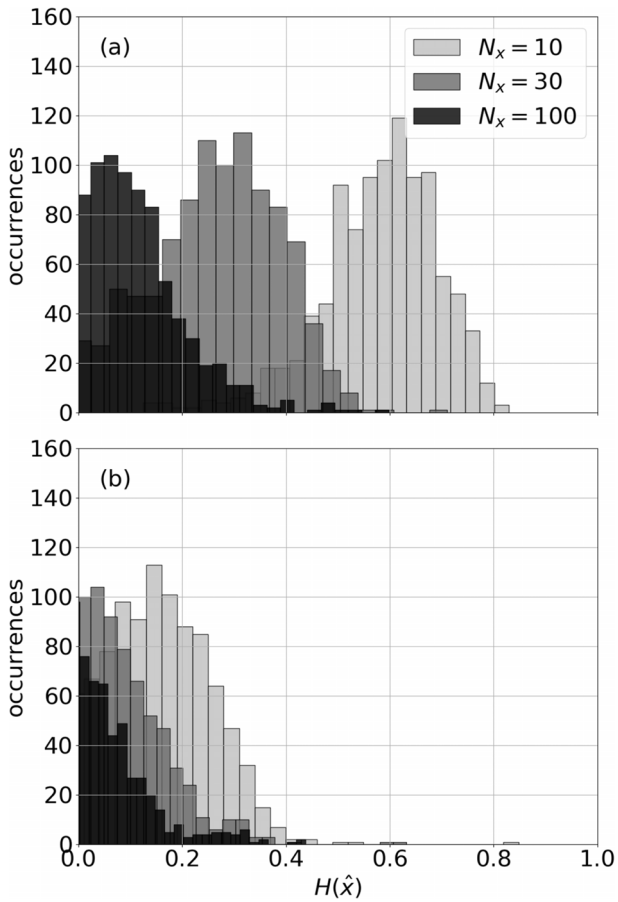

Fig. 5 Posterior entropy for the one-component model and b the multi-component model of subsidence for the dimension N_x=[10, 30, 100] and 1000 ensemble members. Histograms of entropy (Eq. 14) applied to the weighted posterior probability distribution. The entropy is computed and averaged over 1000 simulations

maximum posterior weight in Fig. 3, the entropy decreases as the dimension increases, highlighting that both the maximal weight and the posterior entropy indicate weight collapse.

5.3 Information content and efficiency

In the following, to evaluate the ability of the data assimilation algorithm to conserve information in the posterior, we use the definitions of the data assimilation efficiency $\mathcal{E}_{DA}$ (Eq. 17), and of differential information I_{diff}. According to the results of the information content before assimilation (Table 4), we compute $\mathcal{E}_{DA}$ and I_{diff} with the binwidth of 0.5mm and the dimensions $N_x = N_y = 50$ for an increasing ensemble size N_e. Figure 6 shows results of I_{diff} for the one-component and the multi-component model. Results clearly show a negative I_{diff} for a small ensemble size, in agreement with a stronger weight collapse. In the example of the one-component model, $I_{diff} < 0$ for ensemble size $N_e < 10^3$ and a maximum weight $w_{\max} \sim 0.7$. Differential information increases as a function of the ensemble size N_e, for both the one-component and the multi-component model. As expected, the ensemble size of $N_e = 100$ and $N_e = 1000$ are not large enough to avoid weight collapse in the one-component model and give a negative differential information I_{diff}. In this example, the information in the posterior distribution is less than in the prior. We refer to this as *corrupted* information with negative differential information. I_{diff} becomes positive for larger ensemble sizes, reflecting an increasing information content in the posterior and a respectively decreasing weight collapse.

Results of the differential information with the one-component model show good agreement with the estimation of the required ensemble size of [8] and are used as a performance benchmark for our study. Transposing this approach to the multi-component model, Fig. 6 shows that with an ensemble size of $N_e = 10^7$, the particle method still corrupts the prior information, with $I_{diff} < 0$ and the maximum weight approaching max $w_i = 0.8$ (Fig. 7).

To evaluate the ensemble size which preserves the prior information and then ensures the applicability of importance sampling in the multi-component model, we compare the efficiency $\mathcal{E}_{DA}$ with the maximum weight and the differ-

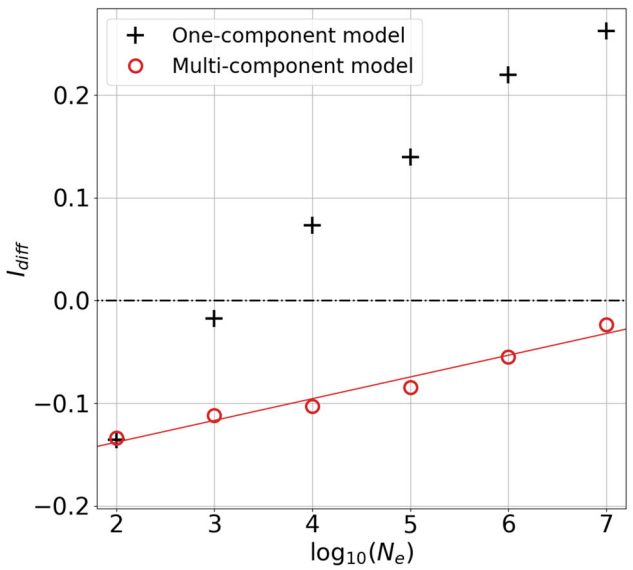

Fig. 6 The differential information I_{diff}, indicates the ability of the posterior to conserve the prior information content. The best-fit line of the differential information with the multi-component model is given by $I_{diff} = 0.018 \log_{10}(N_e) - 0.17$

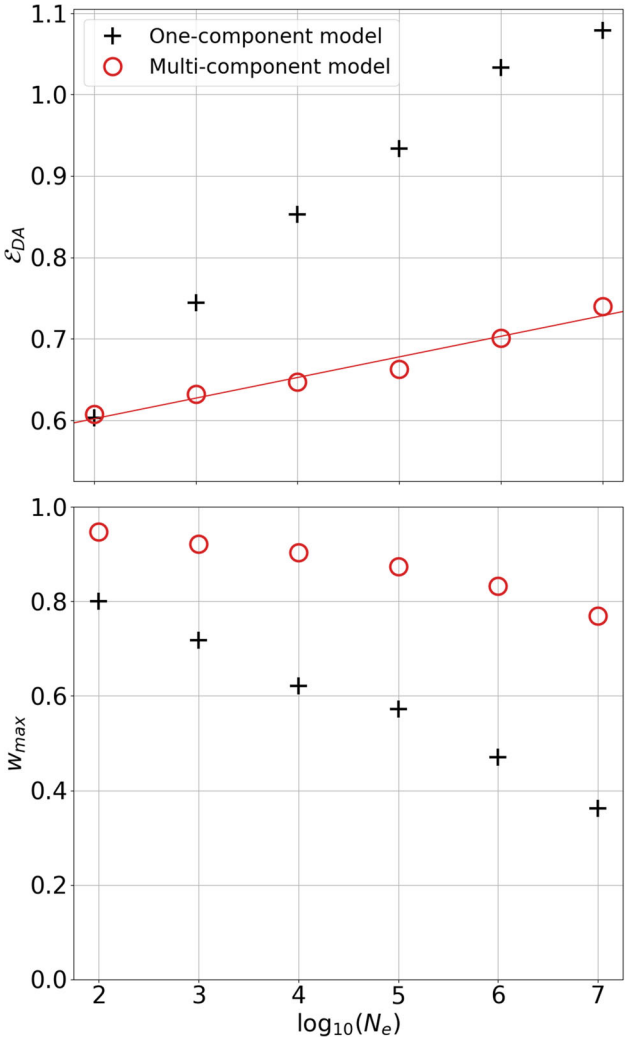

Fig. 7 Efficiency $\mathcal{E}_{DA}$ of the particle method for the one-component model and the multi-component model of subsidence for a dimension $N_x = 50$, and with an increasing ensemble size N_e. The best fit of $\mathcal{E}_{DA}$ with the multi-component model is given by $\mathcal{E}_{DA} = 0.022 \log_{10}(N_e) + 0.56$

ential information I_{diff}. The efficiency $\mathcal{E}_{DA}$ as illustrated in Fig. 7 measures the quality of the posterior given the information content in the prior model and the assimilated observations. Comparison of $\mathcal{E}_{DA}$ with the differential information (Fig. 6), shows that the information in the posterior ($\mathcal{E}_{DA} \sim 0.75$) is at least equal to the prior information (Table 4) for a positive I_{diff}. An efficiency $\mathcal{E}_{DA}$ larger than the prior information implies that the particle method conserves the prior information content. With the example of the one-component model with $N_e = 100$, the particle method does not conserve the prior information with an efficiency of $\mathcal{E}_{DA} = 0.6$ less than the mutual information before assimilation $I(z;x)/H(z) = 0.76$ (Table 4). For an increasing ensemble size $N_e > 10^4$, the particle method now conserves the prior information ($\mathcal{E}_{DA} > 0.75$) and does not corrupt the information in the posterior ($I_{diff} > 0$). This result shows that the required ensemble size should be of $N_e > 10^4$, which is consistent with the previous results of [8]. For equivalent ensemble sizes and equal prior information content, the efficiency in the multi-component model is less than in the one-component model. This result confirms that the importance sampling algorithm causes the loss is information when the observations are spatially correlated. For an increasing ensemble size, the efficiency becomes larger than one (Fig. 7a) despite the normalization. This may come from the uncertainty in the histogram method for the calculation of the mutual information. This could be reduced using a state dimension larger than $N_x = N_y = 50$ (i.e., model prediction or data).

Using the differential information and the efficiency $\mathcal{E}_{DA}$ we can evaluate the minimum required efficiency of a particle method. We consider an acceptable performance in the one-component model for a $\mathcal{E}_{DA}$ at least equal to either the prior information in the model or in the observation. Using this approach, the differential information has a positive value at $N_e = 10^4$ with the efficiency of $\mathcal{E}_{DA} = 0.85$, which is larger than the prior information content in the model and in the data (Table 4). This then suggests $\mathcal{E}_{DA} = 0.85$ as the minimum required efficiency corresponding to $I_{diff} > 0.07$ for an ensemble size of $N_e = 10^4$.

In the multi-component model, linear interpolation in Fig. 7a gives an equivalent performance of $\mathcal{E}_{DA} \sim 0.85$ with an ensemble size larger than $N_e > 10^{13}$. An ensemble size of $N_e > 10^{13}$ is thus required in the example of the multi-component model to have the same performance as that in a one-component model with an ensemble size $N_e = 10^4$. Likewise, the results of I_{diff} in Fig. 6 suggest that we need an ensemble size larger than $N_e > 10^9$ to lift the differential information to a positive value. For I_{diff} to reach ~ 0.07, the required ensemble size should be larger than $N_e > 10^{13}$ (Table 5).

Table 5 Required ensemble size N_e to ensure the particle method applicability in the models of subsidence based on the differential information and the data assimilation efficiency $\mathcal{E}_{DA}$. Experiments are performed with a bin resolution of subsidence of 0.5mm and 40 bins for dimensions $N_x = N_y = 50$

	One-component	Multi-component
$I_{diff} > 0$	$N_e > 10^3$	$N_e > 10^9$
$I_{diff} > 0.07$	$N_e > 10^4$	$N_e > 10^{13}$
$\mathcal{E}_{DA} > 0.85$	$N_e > 10^4$	$N_e > 10^{13}$

6 Discussion

Setting an adequate ensemble size can prevent weight collapse in importance sampling methods in high-dimensional problems. Spatial correlation in the observed field increases the model complexity and requires importance sampling strategies with a large ensemble size. In this study, we show an example with a transformation from model input to model output involving non i.i.d. model predicted subsidence which requires increasing ensemble sizes to ensure the applicability of importance sampling.

As in most data-assimilation systems, the observables sample a spatially correlated field. In data assimilation methods, these non-injective transformations from the state space to the observation space are associated with spatial correlations, and, depending on the strength of the correlations, the tendency for weight collapse varies. Thus, data-assimilation practitioners can expect a possible deviation from the asymptotic results of weight collapse as derived by [8]. Our empirical results can help derive the required ensemble size, showing that information theory, specifically the metric of mutual information can give empirical criteria to ensure a minimum importance sampling efficiency. Results of a so-called multi-component model at the asymptotic limit provide insights to understand the deviation from the results of the importance sampling with the one-component model. In the first part of this study we highlight that in the case of spatial correlation in subsidence, the approximation of the likelihood and the distribution of the scale factor S_i deviates from a standard normal probability distribution. The deviation remains small, however, the log likelihood explicitly depends on the dimension of the state space, N_x. This could explain why the particle method in the multi-component model suffers from a stronger weight collapse than in the one-component model. In the results of [11] with examples of nonlinear models, we observe a similar deviation at the asymptotic limit in our multi-component model, with a reservoir model that has a varying strength of compaction. Evaluating the required ensemble size N_e can be very difficult

given model complexity and this complicates the definition of a generalized methodology to evaluate N_e. We propose criteria to evaluate the required ensemble size in problems of realistic complexity, involving high dimension and spatial correlation.

Information theory gives a means to quantify the information in the model and in the data. It shows how this information is propagated in the posterior distribution [32]. The quality of the data assimilation estimate is often assessed with the variance of the posterior distribution or the effective sample size [36, 37]. However, the posterior variance can be biased because of weight collapse causing a narrow and non-representative posterior distribution. In data assimilation, a common procedure to reduce the observation dimension and increase the posterior variance is by applying localization [31, 38, 39]. A consequence of localization could be that the information of a dataset is optimized by not taking into account the redundant information. In our approach, we evaluate the bias caused by weight collapse by assessing the method performance using mutual information through the quantities of differential information and data assimilation efficiency. The differential information I_{diff} and the data assimilation efficiency $\mathcal{E}_{DA}$ reveal that weight collapse corrupts the information in the posterior distribution, highlighting that the algorithm itself is the main cause of information loss in importance sampling.

To compute the metric of mutual information we used the histogram method and set the sample size of model predictions and the data to $N_x = N_y = 50$ to test the sensitivity of the particle method to the ensemble size N_e. In this study, we choose the number of state variables, of observations of $N_x = N_y = 50$ to have the same spread and binwidth in the histogram, since we want to compare results with a space for an ensemble size between $Ne = 100$ to $N_e = 10^7$. Increases N_x and N_y for the histogram method would result in ensemble collapse or would require an unpractical ensemble size. The one-component model of subsidence provides a means to compare the empirical results of mutual information with the theoretical background on weight collapse in the particle method [8–10, 21]. Results of mutual information show good agreement with previous results of [8]. The same methodology with mutual information could be applied to larger datasets or a time series of data in filtering and ensemble smoother methods. Weight collapse also occurs in those methods [40]. From the result of the efficiency of the data assimilation, $\mathcal{E}_{\mathcal{DA}}$ we derive by linear interpolation the required ensemble size in the multi-component model. It has been shown that the ensemble size must scale exponentially with the dimensions N_y and N_x [8, 9, 21]. As spatial correlation depends on the data-assimilation system, we may also expect a deviation from the linear interpolation proposed in this study. For example, $\mathcal{E}_{\mathcal{DA}}$ in the linear model slightly levels off, and this may suggest that the interpolated values of the required ensemble size in the multi-component model are slightly underestimated. Our approach is also of interest to either assess the method performance (e.g., information loss, ensemble size) or even to optimize the data assimilation before assimilation by evaluating the information content in the prior and in the data. Other criteria to ensure the applicability of importance sampling, are the maximum weight or an effective sample size of the prior ensemble. This could be a threshold to choose the ensemble size and set a minimum level for the required data assimilation efficacy.

However, taken alone it is not clear how we should choose these values. Using a positive differential information I_{diff} and a minimum data assimilation efficiency $\mathcal{E}_{DA}$, we can only obtain a first impression of what is the relevant amount of information that the posterior should contain in a specific problem.

7 Conclusion

With an example of subsidence models, we show that the main cause of performance loss in the particle method comes from the sampling strategy. By choosing a larger ensemble size, we can effectively prevent weight collapse. The required ensemble size in a problem with spatial correlation can be underestimated if evaluated based on non-representative transformation from the state to observation space. In this study, we propose two criteria based on the information theory: the differential information and the data assimilation efficiency using the metric of mutual information. This can be used to empirically derive the required ensemble size in the case of spatial correlation in the observed field. For this, we relate the weight collapse and the performance of importance sampling to the information content before and after assimilation. We find that in the one-component model of subsidence, the particle method requires an ensemble size larger than $N_e > 10^4$ for a dimension $N_x = N_y = 50$ to propagate the information available before assimilation and to obtain positive differential information. Our results show good agreement with an earlier study of [8] and provides an empirical method to measure the applicability of a data assimilation method through the differential information and the relative (i.e., relative to the prior information) criterion of the data assimilation efficiency, to choose the ensemble size. This approach could be further used to track the information content in other data assimilation problems and to optimize the prior information content.

Acknowledgements We are thankful to R.F. Hanssen, R. Govers, and G.F. Nane for the valuable discussions and to the anonymous reviewers for their valuable comments on our manuscript.

Funding This project is funded by NWO project DEEP.NL.2018.052, "Monitoring and Modelling the Groningen Subsurface based on inte-

grated Geodesy and Geophysics: improving the space-time dimension". The authors have no financial or proprietary interests in any material discussed in this article.

References

1. van Leeuwen, P.J., Künsch, H.R., Nerger, L., Potthast, R., Reich, S.: Particle filters for high-dimensional geoscience applications: a review. Q. J. R. Meteorol. Soc. **145**(723), 2335–2365 (2019). https://doi.org/10.1002/qj.3551
2. Vossepoel, F.C., van Leeuwen, P.J.: Parameter estimation using a particle method: inferring mixing coefficients from sea level observations. Mon. Weather. Rev. **135**(3), 1006–1020 (2007). https://doi.org/10.1175/MWR3328.1
3. Fokker, P., Wassing, B., van Leijen, F., Hanssen, R., Nieuwland, D.: Application of an ensemble smoother with multiple data assimilation to the bergermeer gas field, using ps-insar. Geomech. Energy Environ. **5**, 16–28 (2016). https://doi.org/10.1016/j.gete.2015.11.003
4. Zoccarato, C., et al.: Data assimilation of surface displacements to improve geomechanical parameters of gas storage reservoirs. J. Geophys. Res. Solid Earth **121**(3), 1441–1461 (2016). https://doi.org/10.1002/2015JB012090
5. Gazzola, L., et al.: A novel methodological approach for land subsidence prediction through data assimilation techniques. Comput Geosci **25**(5), 1731–1750 (2021). https://doi.org/10.1007/s10596-021-10062-1
6. Evensen, G., Van Leeuwen, P.J.: An ensemble Kalman smoother for nonlinear dynamics. Mon. Weather Rev. **128**(6), 1852–1867 (2000)
7. Evensen, G., Raanes, P.N., Stordal, A.S., Hove, J.: Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. Front. Appl. Math. Stat. **5**, 47 (2019)
8. Snyder, C., Bengtsson, T., Bickel, P., Anderson, J.: Obstacles to high-dimensional particle filtering. Mon. Weather Rev. **136**(12), 4629–4640 (2008). https://doi.org/10.1175/2008MWR2529.1
9. Bengtsson, T., Bickel, P., Li, B., et al.: Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems. Probab. Stat.: Essays in honor of David A. Freedman **2**, 316–334 (2008). https://doi.org/10.1214/193940307000000518
10. Beskos, A., Crisan, D., Jasra, A., et al.: On the stability of sequential Monte Carlo methods in high dimensions. Ann. Appl. Probab. **24**(4), 1396–1445 (2014). https://doi.org/10.1214/13-AAP951
11. Slivinski, L., Snyder, C.: Exploring practical estimates of the ensemble size necessary for particle filters. Mon. Weather Rev. **144**(3), 861–875 (2016). https://doi.org/10.1175/MWR-D-14-00303.1
12. Fokker, P.A., Visser, K., Peters, E., Kunakbayeva, G., Muntendam-Bos, A.: Inversion of surface subsidence data to quantify reservoir compartmentalization: A field study. J. Pet. Sci. Eng. **96**, 10–21 (2012). https://doi.org/10.1016/j.petrol.2012.06.032
13. Hanssen, R.F.: Radar interferometry: data interpretation and error analysis. Kluwer Academic Publishers, Dordrecht (2001)
14. Simonin, D., Waller, J.A., Ballard, S.P., Dance, S.L., Nichols, N.K.: A pragmatic strategy for implementing spatially correlated observation errors in an operational system: an application to doppler radial winds. Q. J. R. Meteorol. Soc. **145**(723), 2772–2790 (2019). https://doi.org/10.1002/qj.3592
15. Geertsma, J.: Land subsidence above compacting oil and gas reservoirs. J. Pet. Technol. **25**(06), 734–744 (1973). https://doi.org/10.2118/3730-PA
16. Du, J., Olson, J.E.: A poroelastic reservoir model for predicting subsidence and mapping subsurface pressure fronts. J Petrol Sci Eng **30**(3–4), 181–197 (2001). https://doi.org/10.1016/S0920-4105(01)00131-0
17. Doucet, A., De Freitas, N., Gordon, N.: An introduction to sequential Monte Carlo methods. Sequential Monte Carlo methods in practice 3–14 (2001)
18. Van Leeuwen, P.J., Cheng, Y., Reich, S., van Leeuwen, P.J.: Nonlinear Data Assimilation for high-dimensional systems: with geophysical applications. Springer (2015)
19. Morzfeld, M., et al.: Iterative importance sampling algorithms for parameter estimation. SIAM J Sci Comput **40**(2), B329–B352 (2018)
20. Evensen, G., Vossepoel, F.C., van Leeuwen, P.J.: Data assimilation fundamentals: a unified formulation of the State and parameter estimation problem (2022). https://doi.org/10.1007/978-3-030-96709-3
21. Beskos, A., Crisan, D., Jasra, A., Kamatani, K., Zhou, Y.: A stable particle filter for a class of high-dimensional state-space models. Adv. Appl. Probab. **49**(1), 24–48 (2017). https://doi.org/10.1017/apr.2016.77
22. Muñoz, L.F.P., Roehl, D.: An analytical solution for displacements due to reservoir compaction under arbitrary pressure changes. Appl. Math. Model. **52**, 145–159 (2017). https://doi.org/10.1016/j.apm.2017.06.023
23. Tempone, P., Fjær, E., Landrø, M.: Improved solution of displacements due to a compacting reservoir over a rigid basement. Appl Math Model **34**(11), 3352–3362 (2010). https://doi.org/10.1016/j.apm.2010.02.025
24. Candela, T., et al.: Depletion-induced seismicity at the Groningen gas field: coulomb rate-and-state models including differential compaction effect. J. Geophys. Res. Solid Earth **124**(7), 7081–7104 (2019). https://doi.org/10.1029/2018JB016670
25. Geertsma, J., Opstal, v., et al.: A numerical technique for predicting subsidence above compacting reservoirs, based on the nucleus of strain concept **30**, 53–78 (1973)
26. Shannon, C.: A Mathematical Theory of Cryptography (1945)
27. Jost, J.: Dynamical systems: examples of complex behaviour (2005). https://doi.org/10.1007/3-540-28889-9
28. Yustres, Á., Asensio, L., Alonso, J., Navarro, V.: A review of Markov Chain Monte Carlo and information theory tools for inverse problems in subsurface flow. Comput. Geosci. **16**(1), 1–20 (2012). https://doi.org/10.1007/s10596-011-9249-z
29. van Leeuwen, P.J.: A variance-minimizing filter for large-scale applications. Mon. Weather Rev. **131**(9), 2071–2084 (2003). https://doi.org/10.1175/1520-0493(2003)131<2071:AVFFLA>2.0.CO;2
30. Fowler, A., van Leeuwen, P.J.: Measures of observation impact in non-gaussian data assimilation. Tellus A Dyn. Meteorol. Oceanogr. **64**(1), 17192 (2012)
31. Fowler, A., Jan Van Leeuwen, P.: Observation impact in data assimilation: the effect of non-Gaussian observation error. Tellus A Dyn. Meteorol. Oceanogr. **65**(1), 20035 (2013)

32. Nearing, G.S., Gupta, H.V., Crow, W.T., Gong, W.: An approach to quantifying the efficiency of a Bayesian filter. Water. Resour. Res. **49**(4), 2164–2173 (2013). https://doi.org/10.1002/wrcr.20177
33. Nearing, G., et al.: The efficiency of data assimilation. Water Resour Res **54**(9), 6374–6392 (2018). https://doi.org/10.1029/2017WR020991
34. Brewer, B.J.: Computing entropies with nested sampling. Entropy **19**(8), 422 (2017). https://doi.org/10.3390/e19080422
35. Paninski, L.: Estimation of entropy and mutual information. Neural Comput. **15**(6), 1191–1253 (2003). https://doi.org/10.1162/089976603321780272
36. Li, T., Sun, S., Sattar, T.P., Corchado, J.M.: Fight sample degeneracy and impoverishment in particle filters: a review of intelligent approaches. Expert Syst Appl **41**(8), 3944–3954 (2014). https://doi.org/10.1016/j.eswa.2013.12.031
37. Martino, L., Elvira, V., Louzada, F.: Effective sample size for importance sampling based on discrepancy measures. Signal Proc. **131**, 386–401 (2017). https://doi.org/10.1016/j.sigpro.2016.08.025
38. Poterjoy, J.: A localized particle filter for high-dimensional nonlinear systems. Mon. Weather Rev. **144**(1), 59–76 (2016)
39. Luo, X., Lorentzen, R.J., Valestrand, R., Evensen, G.: Correlation-based adaptive localization for ensemble-based history matching: applied to the norne field case study. SPE Reserv. Eval. Eng. **22**(03), 1084–1109 (2019)
40. Stordal, A.S., Elsheikh, A.H.: Iterative ensemble smoothers in the annealed importance sampling framework. Adv. Water Resour. **86**, 231–239 (2015). https://doi.org/10.1016/j.advwatres.2015.09.030