Full length article

Underground hydrogen storage (UHS) in natural storage sites: A perspective of subsurface characterization and monitoring

Xiaodong Luo [a,*], Svenn Tveit [a], Raoof Gholami [b], Pål Østebø Andersen [b]

[a] *Norwegian Research Centre (NORCE), Nygårdsgaten 112, Bergen, 5008, Norway*
[b] *Department of Energy Resources, University of Stavanger, Kjell Arholms gate 41, Stavanger, 4021, Norway*

ARTICLE INFO

ABSTRACT

With the long-standing efforts of green transition in our society, underground hydrogen storage (UHS) has emerged as a viable solution to buffering seasonal fluctuations of renewable energy supplies and demands. Like operations in hydrocarbon production and geological CO_2 storage, a successful UHS project requires a good understanding of subsurface formations, while having different operational objectives and practical challenges. Similar to the situations in hydrocarbon production and geological CO_2 storage, in UHS problems, the information of subsurface formations at the field level cannot be obtained through direct measurements due to the resulting high costs. As such, there is a need for subsurface characterization and monitoring at the field scale, which uses a certain history matching algorithm to calibrate a numerical subsurface model based on available field data. Whereas subsurface characterization and monitoring have been widely used in hydrocarbon production activities for a better understanding of hydrocarbon reservoirs, to the best of our knowledge, at present it appears to be a relatively less touched area in UHS problems. This work aims to narrow this noticed gap, and investigates the use of an ensemble-based workflow for subsurface characterization and monitoring in a 3D UHS case study. Numerical results in this case study indicate that the ensemble-based workflow works reasonably well, while also identifying some particular challenges that would be relevant to real-world problems.

1. Introduction

The intergovernmental ambition of taming climate change, in combination with energy crises, has significantly accelerated the development of clean and renewable energies in the past decades. Some types of these renewable energies, such as wind and solar, are abundant in nature, yet a noticeable shortcoming is that their availability may vary seasonally in many geographic locations. For this reason, there is a need for developing energy storage systems to buffer the seasonal fluctuations of renewable energy supplies and demands, apart from the considerations of energy security and strategic reserves [1,2]. In this regard, a well-received notion is to store renewable energy through an electricity-to-hydrogen (H_2) conversion process. For instance, one may use the excess electricity generated from solar energy in the summer to produce H_2, which is then stored somewhere for a period, and converted back to electricity when there is a higher energy demand in the winter [3].

Potential H_2 storage options include underground H_2 storage (UHS), surface H_2 storage and material-based H_2 storage [4]. Surface H_2 storage (e.g., in gas tanks) typically targets end-uses and aims to meet hourly to daily demands. Material-based H_2 storage deposits H_2 through a physical or chemical sorption mechanism, and can conduct a long term storage of H_2 in a safe way. However, the unit storage cost could be high due to the demanding operation conditions to achieve the physical or chemical sorption of H_2, yet the potential storage capacity is not clear so far [4]. In contrast, currently UHS appears to be the only viable option that has the potential to provide long-term storage of H_2 at large scales to balance seasonable supplies and demands [2,4], while experiences in similar underground natural gas or CO_2 storage projects can be extended to facilitate the developments and operations of UHS projects for improved safety and reduced costs.

UHS can be conducted in both engineered and naturally existing storage sites [1,2,4–7]. Engineered sites include salt caverns, hard rock caverns and abandoned mines, and naturally existing sites consist of depleted hydrocarbon (gas or oil) fields and aquifers, among others. Table 1 provides a relative comparison of several characteristic aspects of some UHS options, from which one can see that among these different options, salt cavern possesses the highest technology readiness

---

Table 1
Relative comparison of some characteristic aspects of UHS options, following [2,5].

	Salt caverns	Hard rock caverns or abandoned mines	Depleted hydrocarbon fields	Aquifers
Suitability for UHS	High	High	Site-specific	Site-specific
Technology readiness level (TRL)	High	Low	Low	Low
Geographic availability	Somewhat limited	Somewhat limited	Somewhat limited	Wide
Capital expenditures (CapEx) for project development	Low	High	Low	Low
Operating expenses (OpEx)	Medium	Medium	Low	Low
Cyclic frequency	High	High	Low	Low
Storage capacity	Medium	Medium	High	High

level (TRL) right now, with a few commercial UHS projects running worldwide [2]. Apart from the typically lower storage capacities than those of natural storage sites, an important issue related to the use of salt caverns is that salt-bearing formations may not be available in many geographic locations [1], which makes the use of salt caverns alone inadequate to meet the enormous energy demands of our society. For this reason, other UHS alternatives, namely, hard rock caverns, abandoned mines, depleted hydrocarbon fields and aquifers, should be adopted to complement the use of salt caverns. Engineered hard rock caverns and abandoned mines share similar problems (e.g., capacity and geographical availability) to salt caverns, but are with lower TRLs and practically less used than salt caverns at present [2]. Regarding natural storage sites, the current TRLs of both depleted hydrocarbon fields and aquifers are also lower than those of salt caverns, although there are a few pilot studies reported in the literature (cf. 2 and the references therein). Despite this TRL gap, depleted hydrocarbon fields and aquifers tend to have substantially higher storage capacities than salt caverns, hard rock caverns and abandoned mines [1], and can achieve relatively low unit storage costs due to the economies of scale. With this said, both engineered and natural UHS options have their own benefits and disadvantages. Natural storage sites enjoy the advantages of having, for instance, (1) wider geographic availability; (2) higher storage capacity; and (3) lower project capital expenditures and operating expenses, in comparison to engineered UHS sites. Meanwhile, within the category of natural UHS sites, aquifers share similar characteristics to those of depleted hydrocarbon fields. The technical advantages of aquifers over depleted hydrocarbon fields include, for instance, the exemption from treating legacy wells to avoid H_2 leakage, and the flexibility for storage site selections that allows more economic connections to H_2 production sites, transportation infrastructures and/or end-use facilities; whereas the benefits of depleted hydrocarbon fields relative to aquifers consist of, for instance, better subsurface understanding and existing infrastructure (e.g., pipelines and storage facilities) due to previous exploration, monitoring, and production activities. For more detailed elaboration of the pros and cons of different UHS options, readers are referred to, for instance, [1,2,4,5,7].

The focus of the current work is on UHS in natural storage sites. For this purpose, an important requirement is to obtain a good understanding of the subsurface formations, which is vital for understanding the trapping mechanism of H_2 in the subsurface, imaging potential migration of H_2, detecting potential H_2 leakage in long-term operations, managing the deliverability of H_2 during injection and withdrawal cycles, and so on. To meet this requirement, a common practice is to conduct *subsurface characterization and monitoring*, in which one first builds a numerical model of the subsurface, and then calibrates the numerical model based on available field data (e.g., production and/or geophysical survey data). Typically, the numerical model is represented by a set of parameters (e.g., permeability and porosity), which are referred to as model variables hereafter. Within this context, subsurface characterization and monitoring correspond to a parameter estimation problem.

In connection to the aforementioned focus, in the current work, we investigate the use of an ensemble-based workflow for subsurface characterization and monitoring, within the context of UHS in a subsurface model from [8]. Practically speaking, ensemble-based workflows [9–12] enjoy certain attractive benefits, such as the nature of being derivative-free (i.e., no need to compute the derivatives of an objective function with respect to model variables), the ability to deal with a large number of parameters, and the capacity to provide a means of uncertainty quantification. As such, they can be adopted to address one of the important issues, namely, heterogeneities and uncertainties at the field scale, in UHS problems [2]. In addition, although not investigated in the current work, it is expected that ensemble-based workflows themselves are independent of storage options, and can be applied to UHS problems in different natural storage sites, although site-specific challenges could be encountered therein.

In terms of novelty, within the context of UHS in natural storage sites, much attention has been paid to understanding – at the pore or core scale – PVT properties of H_2 in the subsurface, geochemical and microbial interactions of H_2 to the subsurface environment, and hydrodynamics of H_2 flow in porous media [2,7,13], among others. In the meantime, at the field scale, the focus appears to be more on numerical modeling and simulation, multiphase flow behavior (e.g., wettability, contact angle, relative permeability, capillary pressure and hysteresis) studying, storage site screening and ranking, economics analysis, and so on [2,8,14–19]. To the best of our knowledge, however, in the literature there are few efforts dedicated to subsurface characterization and monitoring, which would be crucial to the developments and operations of UHS projects in natural storage sites, based on similar experience from underground natural gas and CO_2 storage problems [20–22]. The current work can be viewed as one of the first attempts that aim to narrow this noticed gap.

As a side remark, here we also briefly mention certain similarities and differences among hydrocarbon production, geological CO_2 storage and UHS. In terms of similarities, all three subjects involve subsurface operations, for which subsurface characterization and monitoring would play a critical role. On the other hand, they bear different objectives: Hydrocarbon production focuses on non-sustainable natural resource recovery; geological CO_2 storage aims to achieve long-term safe storage of CO_2 in the subsurface for greenhouse gas reduction; and UHS targets H_2 storage and retrieval as a means of renewable energy management. Accordingly, these different objectives lead to distinct procedures and constraints in subsurface operations. For instance, using cushion gas is usually not a prerequisite for hydrocarbon production, but appears to be a standard requirement for both geological CO_2 storage and UHS (since using cushion gas has the benefits of providing pressure supports, helping reduce fluid migration, and so on) [2]. Meanwhile, although both CO_2 storage and UHS involve gas deposits in subsurface conditions, the expected storage time scales may be substantially different (e.g., centurial for CO_2 storage versus annual for UHS). Moreover, while it may be considered beneficial that CO_2 reduction takes place in the subsurface through various trapping mechanisms, it is typically undesirable that H_2 is trapped in similar
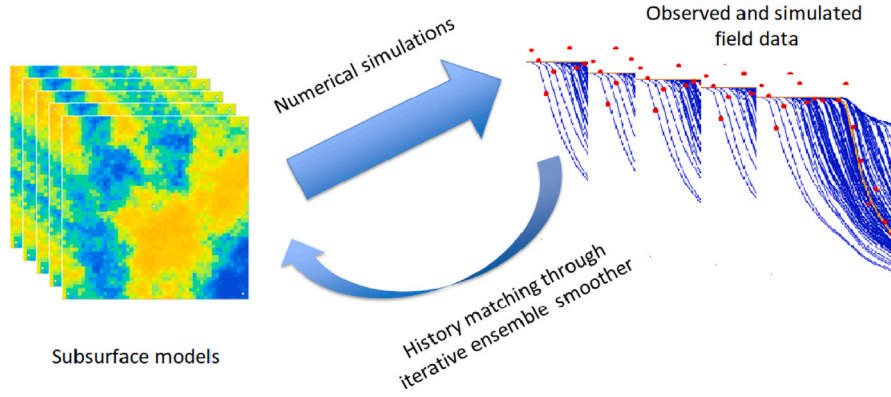
Fig. 1. Subsurface characterization and monitoring workflow based on the iterative ensemble smoother (IES) in [12].

ways. As a result, having a good understanding of the impacts of various trapping mechanisms on UHS is an economically important topic within this context. With this said, it is evident that despite the recognized similarities, these three subjects (hydrocarbon production, geological CO$_2$ storage and UHS) have their own operational objectives and particular practical challenges.

The rest of the paper is organized as follows: We start with an introduction to an ensemble-based history matching workflow for subsurface characterization and monitoring, and then proceed to apply the presented workflow to a 3D case study of UHS in the subsurface model from [8]. We illustrate how the ensemble-based workflow takes into account heterogeneities and uncertainties at the field scale in this case study. In addition, we investigate the performance of the presented workflow in this case study, which initially takes place in a scenario that only uses production data from a single well in history matching, and is extended to another scenario that assimilates both production and 4D seismic data for improved performance. We discuss the implications of the numerical results in the case study, identify some open challenges, and conclude the whole work with potential future research directions.

2. Ensemble-based history matching workflow for subsurface characterization and monitoring

Fig. 1 provides a sketch of the ensemble-based workflow used in the current work. Within this context, subsurface characterization and monitoring consist of a forward modeling process and an inverse modeling process, respectively. The forward modeling process aims to generate simulated field data based on an ensemble of subsurface models and an associated numerical simulator, whereas the inverse modeling process examines the differences between the simulated and observed field data, and adopts a certain algorithm (an iterative ensemble smoother in this case) to adjust the ensemble of subsurface models in such a way that the aforementioned data discrepancies can be reduced to a satisfactory level. Typically, each subsurface model in the ensemble is parameterized by a set of parameters (such as petro-physical parameters like permeability and porosity). In this case, subsurface characterization and monitoring boil down to a parameter estimation problem, which is often referred to as a history matching or data assimilation problem in the community of reservoir engineering [9,23].

In the sequel, the focus of this section will be on introducing one of the iterative ensemble smoothers (IES) [10–12] as the history matching algorithm. For a better comprehension of how this IES algorithm should be applied to large-scale problems, in addition to presenting its basic form, we also include a practical implementation of this IES algorithm for enhanced numerical stability, and discuss the use of two auxiliary techniques, namely, localization and dimensionality reduction, to further improve the performance of this IES algorithm, when it is applied to handle big subsurface models and big field data sets with a relatively small ensemble size.

2.1. The basic iterative ensemble smoother algorithm

As indicated in Fig. 1, the IES algorithm proposed in [12] is adopted as the inversion algorithm. Like other ensemble-based data assimilation algorithms, this IES updates an ensemble of subsurface models based on the differences between the observed and the simulated field data. The main benefits of using an IES algorithm include: (1) the IES algorithm is non-intrusive and derivative-free, meaning that for model updates, there is no need to explicitly evaluate the gradients (nor Hessians) of the objective function with respect to model variables, as can be seen in Eqs (3) – (5) below; (2) the IES algorithm implicitly utilizes the information of both gradients and Hessians for model updates, and converges faster than alternative algorithms that only use the information of gradients [24,25]; (3) the IES algorithm is able to simultaneously update a large number of model variables, as is demonstrated in real field case studies [26]; (4) by generating an ensemble of calibrated subsurface models, instead of a single one, the IES provides a means of uncertainty quantification, which could be further exploited in subsequent procedures like risk analysis and robust production optimization. Putting together these noticed benefits, the IES algorithms become an attractive choice in practice, as they are relatively easy to implement and have the capacity to handle large-scale subsurface characterization and monitoring problems.

In the sequel, we start with an introduction to the basic form of the IES algorithm, and then proceed to present certain modifications of the basic form in order to enhance the performance of the IES in practical applications.

Let $\mathbf{m} \in \mathbb{R}^m$ represent a generic subsurface model, which is an m-dimensional vector containing m parameters (e.g., permeability and porosity) to be updated by the IES algorithm. Also, let $\mathbf{g} : \mathbb{R}^m \to \mathbb{R}^p$ be the forward simulator that generates a p-dimensional vector of simulated field data $\mathbf{d}^{\text{sim}} \equiv \mathbf{g}(\mathbf{m}) \in \mathbb{R}^p$, for a given subsurface model $\mathbf{m}$. In correspondence to $\mathbf{d}^{\text{sim}}$, there exists a p-dimensional vector of observed field data $\mathbf{d}^{obs} \in \mathbb{R}^p$ (called observations hereafter). It is assumed that the observations $\mathbf{d}^{obs}$ are contaminated by certain observation errors that follow a multivariate Gaussian distribution with zero mean and covariance $\mathbf{C}_d \in \mathbb{R}^{p \times p}$.

The IES algorithm in [12] aims to minimize the average of an ensemble of cost functions at each iteration step. Specifically, suppose that $\mathbf{M}^i \equiv \{\mathbf{m}_j^i : \mathbf{m}_j^i \in \mathbb{R}^m\}_{j=1}^{N_e}$ represents an ensemble of N_e subsurface models obtained at the ith iteration step. The IES algorithm updates $\mathbf{M}^i$ to a new ensemble $\mathbf{M}^{i+1} \equiv \{\mathbf{m}_j^{i+1} : \mathbf{m}_j^{i+1} \in \mathbb{R}^m\}_{j=1}^{N_e}$ by solving the following minimum-average-cost (MAC) problem at each iteration step:

$$\min_{\{\mathbf{m}_j^{i+1}\}_{j=1}^{N_e}} \frac{1}{N_e} \sum_{j=1}^{N_e} L_j^{i+1}; \tag{1}$$

$$L_j^{i+1} \equiv \left(\mathbf{d}_j^{obs} - \mathbf{g}\left(\mathbf{m}_j^{i+1}\right)\right)^T \mathbf{C}_d^{-1} \left(\mathbf{d}_j^{obs} - \mathbf{g}\left(\mathbf{m}_j^{i+1}\right)\right)$$
$$+ \gamma^i \left(\mathbf{m}_j^{i+1} - \mathbf{m}_j^i\right)^T \left(\mathbf{C}_m^i\right)^{-1} \left(\mathbf{m}_j^{i+1} - \mathbf{m}_j^i\right), \tag{2}$$

where $\mathbf{d}_j^{obs}$ ($j = 1, 2, \cdots, N_e$) are random samples drawn from the multivariate Gaussian distribution $N(\mathbf{d}^{obs}, \mathbf{C}_d)$; $\mathbf{C}_m^i \equiv \mathbf{S}_m^i \left(\mathbf{S}_m^i\right)^T$ is the sample covariance matrix with respect to the ensemble $\mathbf{M}^i$, with the square root matrix $\mathbf{S}_m^i$ being defined in Eq. (4) below. In Eq. (2), the cost function L_j^{i+1} consists of two terms: the first one (data mismatch term) describes the discrepancies between the observed and the simulated observations, whereas the second one (regularization term) is additionally introduced to mitigate the ill-posedness of history matching as an inverse problem and avoid the issue of overfitting [27]. The relative weights between the data mismatch and the regularization terms depend on the regularization parameter γ^i, whose value varies over the iteration step and is determined using the following rule [12]: $\gamma^i \equiv \alpha^i \times \text{trace}\left(\mathbf{S}_g^i(\mathbf{S}_g^i)^T\right)/\text{trace}\left(\mathbf{C}_d\right)$, with the operator "trace" computing the trace of a matrix, and $\mathbf{S}_g^i$ being a square root matrix defined in Eq. (5). The initial value $\alpha^0 = 1$, and $\alpha^{i+1} = 0.9 \times \alpha^i$ if the average data mismatch (see the definition in Eq. (13) later) is reduced, otherwise $\alpha^{i+1} = 2 \times \alpha^i$.

As shown in [12], the MAC problem (Eqs. (1) – (2)) is approximately solved by the following model update formula:

$$\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \mathbf{S}_m^i(\mathbf{S}_g^i)^T \left(\mathbf{S}_g^i(\mathbf{S}_g^i)^T + \gamma^i \mathbf{C}_d\right)^{-1} \left(\mathbf{d}_j^{obs} - \mathbf{g}\left(\mathbf{m}_j^i\right)\right),$$
$$\text{for } j = 1, 2, \cdots, N_e; \tag{3}$$

$$\mathbf{S}_m^i = \frac{1}{\sqrt{N_e - 1}} \left[\mathbf{m}_1^i - \overline{\mathbf{m}}^i, \cdots, \mathbf{m}_{N_e}^i - \overline{\mathbf{m}}^i\right]; \quad \overline{\mathbf{m}}^i = \frac{1}{N_e}\sum_{j=1}^{N_e} \mathbf{m}_j^i; \tag{4}$$

$$\mathbf{S}_g^i = \frac{1}{\sqrt{N_e - 1}} \left[\mathbf{g}(\mathbf{m}_1^i) - \mathbf{g}(\overline{\mathbf{m}}^i), \cdots, \mathbf{g}(\mathbf{m}_{N_e}^i) - \mathbf{g}(\overline{\mathbf{m}}^i)\right]. \tag{5}$$

From Eqs. (3)–(5), it is clear that the model update formula does not explicitly involve any quantities of derivatives, although it can be shown that the information of gradient and Hessian of the cost function with respect to the model variables is implicitly utilized in the course of deriving the model update formulate [24,25], which is a practically desirable feature, and makes the IES algorithm often reach a relatively good solution after only several (e.g., four to six) iteration steps.

2.2. Practical implementation of the IES algorithm

In a practical implementation, some modifications of the update formulae, Eqs. (3)–(5), are introduced to improve the numerical stability of the algorithm. First of all, to mitigate the potential issue of distinct orders of magnitudes resulting from different types of field data, it is necessary to re-scale relevant quantities in the observation space, leading to the following normalized terms:

$$\tilde{\mathbf{S}}_g^i \equiv \mathbf{C}_d^{-1/2} \mathbf{S}_g^i; \qquad \tilde{\mathbf{d}}_j^{obs} \equiv \mathbf{C}_d^{-1/2} \mathbf{d}_j^{obs}; \qquad \tilde{\mathbf{g}}\left(\mathbf{m}_j^i\right) \equiv \mathbf{C}_d^{-1/2} \mathbf{g}\left(\mathbf{m}_j^i\right), \tag{6}$$

where $\mathbf{C}_d^{-1/2}$ is a square root matrix of the matrix $\mathbf{C}_d^{-1}$ (the inverse of the covariance matrix $\mathbf{C}_d$). Accordingly, Eq. (3) can be re-written as

$$\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \mathbf{S}_m^i(\tilde{\mathbf{S}}_g^i)^T \left(\tilde{\mathbf{S}}_g^i(\tilde{\mathbf{S}}_g^i)^T + \gamma^i \mathbf{I}_p\right)^{-1} \left(\tilde{\mathbf{d}}_j^{obs} - \tilde{\mathbf{g}}\left(\mathbf{m}_j^i\right)\right), \tag{7}$$

where $\mathbf{I}_p$ is the p-dimensional identity matrix.

The second modification is mainly about further improving the numerical efficiency and stability of the update formula in Eq. (7), where the inverse of the matrix $\tilde{\mathbf{S}}_g^i(\tilde{\mathbf{S}}_g^i)^T + \gamma^i \mathbf{I}_p$ needs to be computed. Instead of a straightforward computation of this inverse matrix, one can exploit the fact that in practice, the ensemble size N_e (typically around 100) is much smaller than the data size p. Under this setting, $\tilde{\mathbf{S}}_g^i(\tilde{\mathbf{S}}_g^i)^T$ is a singular matrix with a rank no more than $N_e - 1$. As a result, one can first apply a singular value decomposition (SVD) to the square root matrix $\tilde{\mathbf{S}}_g^i$, such that

$$\tilde{\mathbf{S}}_g^i = \mathbf{U}^i \, \boldsymbol{\Sigma}^i \left(\mathbf{V}^i\right)^T, \tag{8}$$

where $\mathbf{U}^i$ ($p \times p$) and $\mathbf{V}^i$ ($N_e \times N_e$) are unitary matrices containing the left- and right-singular vectors, respectively, and $\boldsymbol{\Sigma}^i$ ($p \times N_e$) is a rectangular

diagonal matrix with its non-zero singular values denoted by $\sigma_1, \sigma_2, \cdots$ (in a descending order). One then proceeds to approximate $\tilde{\mathbf{S}}_g^i$ by truncating insignificant singular values and their associated singular vectors, so that

$$\tilde{\mathbf{S}}_g^i \approx \mathbf{U}_r^i \boldsymbol{\Sigma}_r^i \left(\mathbf{V}_r^i\right)^T, \tag{9}$$

where now $\boldsymbol{\Sigma}_r^i$ ($r \times r$) is a diagonal matrix whose diagonal elements correspond to $\sigma_1, \sigma_2, \cdots, \sigma_r$ (after discarding singular values $\sigma_{r+1}, \sigma_{r+2}, \cdots$), $\mathbf{U}_r^i$ ($p \times r$) and $\mathbf{V}_r^i$ ($N_e \times r$) are matrices resulting from the removals of the left- and right-singular vectors associated with the singular values $\sigma_{r+1}, \sigma_{r+2}, \cdots$ from the original unitary matrices $\mathbf{U}^i$ and $\mathbf{V}^i$, respectively. In the current work, the rank r ($r \leq N_e - 1 \ll p$) is determined using the following criterion: $\left(\sum_{\ell=1}^r \sigma_\ell^2\right)/\left(\sum_{\ell=1}^{N_e-1} \sigma_\ell^2\right) \leq 0.95$ and $\left(\sum_{\ell=1}^{r+1} \sigma_\ell^2\right)/\left(\sum_{\ell=1}^{N_e-1} \sigma_\ell^2\right) > 0.95$.

Inserting Eq. (9) into Eq. (7), one obtains

$$\mathbf{m}_j^{i+1} \approx \mathbf{m}_j^i + \mathbf{S}_m^i \mathbf{V}_r^i \boldsymbol{\Sigma}_r^i \left((\boldsymbol{\Sigma}_r^i)^2 + \gamma^i \mathbf{I}_r\right)^{-1} \left(\mathbf{U}_r^i\right)^T \left(\tilde{\mathbf{d}}_j^{obs} - \tilde{\mathbf{g}}\left(\mathbf{m}_j^i\right)\right), \tag{10}$$

where $(\boldsymbol{\Sigma}_r^i)^2 \equiv \boldsymbol{\Sigma}_r^i(\boldsymbol{\Sigma}_r^i)^T$ and $\mathbf{I}_r$ is the r-dimensional identity matrix. Following this modification, the scalar coefficient γ^i is determined using the following rule: $\gamma^i \equiv \alpha^i \times \text{trace}\left((\boldsymbol{\Sigma}_r^i)^2\right)/\text{trace}\left(\mathbf{I}_r\right) = \alpha^i \sum_{\ell=1}^r \sigma_\ell^2/r$, with the values of α^i being determined in the same way as described in the text below Eq. (2).

In comparison to the update formula in Eq. (7), the modified one, Eq. (10), only requires to compute the inverse of the low-dimensional diagonal matrix $(\boldsymbol{\Sigma}_r^i)^2 + \gamma^i \mathbf{I}_r$, which is more efficient than calculating the inverse of the (much higher dimensional) matrix $\tilde{\mathbf{S}}_g^i(\tilde{\mathbf{S}}_g^i)^T + \gamma^i \mathbf{I}_p$. On the other hand, through the truncated singular value decomposition (TSVD), it can be shown that for a given value γ^i, the condition number[1] of the matrix $\left((\boldsymbol{\Sigma}_r^i)^2 + \gamma^i \mathbf{I}_r\right)^{-1}$ is smaller than that of the matrix $\left(\tilde{\mathbf{S}}_g^i(\tilde{\mathbf{S}}_g^i)^T + \gamma^i \mathbf{I}_p\right)^{-1}$. An implication of this reduced condition number is that the influence of the observation noise in $\tilde{\mathbf{d}}_j^{obs}$ is less amplified, and that the ill-posedness of the inverse problem is further mitigated [27]. Practically speaking, the update formula in Eq. (10) thus tends to be numerically more stable than that in Eq. (7), as was typically observed in the applications of the IES algorithm.

2.3. Correlation-based adaptive localization in the IES

In practice, running the forward simulator $\mathbf{g}$ is often computationally expensive in large scale problems, which limits the ensemble size N_e that one can afford. For this reason, the typical ensemble size used in the IES algorithm is in the order of $\mathcal{O}(10^2)$, which is often much smaller than the model size m and the data size p. One consequence of using a relatively small ensemble size is that when the ensemble is used to compute certain matrices (e.g., $\mathbf{S}_m^i(\tilde{\mathbf{S}}_g^i)^T$ and $\tilde{\mathbf{S}}_g^i(\tilde{\mathbf{S}}_g^i)^T$ in Eq. (7)), issues like sampling errors and rank-deficiencies would arise [28], often leading to deteriorated performance of subsurface characterization and monitoring. To address these noticed issues, an auxiliary technique, called *localization* [29–33], is often introduced to improve the performance of the IES algorithm.

There are two major schools of localization techniques used together with IES algorithms. One school of these methods mainly conducts localization based on the distances between the physical locations of model variables and observations (see, for example, 29,30), whereas the other instead utilizes for localization the sample correlations between an ensemble of model variables and the corresponding ensemble of simulated field data (see, for example, 31–33). In comparison to distance-based localization, correlation-based localization does not require physical locations associated with either model variables or observations, and can overcome or mitigate the long-standing challenges

---

[1] For a symmetric matrix, the condition number is defined as the ratio of the largest eigenvalue to the smallest one.

Table 2

Information of the numerical subsurface model used in the UHS case study. The statistics of mean and STD reported below are calculated based on respective values distributed on active gridblocks.

Model information	
Storage site depth (mean $\pm$ STD)	1166.5 m $\pm$ 18.3m
Model size	$61 \times 61 \times 12$ (44 652 gridblocks distributed on 12 vertical layers, all active)
Gridblock size	$\Delta X = \Delta Y = 50.0$m for all gridblocks; $\Delta Z \approx 2.2$m on average (uneven among gridblocks)
Phases	Water and gas
Initial conditions (mean $\pm$ STD)	Temperature (°C): $60.0 \pm 3.8 \times 10^{-6}$; Pressure (bar): 81.5 ± 0.3; Water saturation (%): 21.9 ± 12.2
Number of wells	1 (labeled as 'WELL3'. Vertical well with the perforation zone from Layer 2 to Layer 12)
Simulator	ECLIPSE© 300
Simulation time	01 Jan 2018 - 09 Feb 2022 (1500 days in total)
Operating schedule	– N_2 injection for the first 360 days with 4 sub-cycles; each sub-cycle includes 60 days of N_2 injection, and then 30 days of well shut-in; Operating target: bottom hole pressure (BHP)
	After the first 360 days, four H_2 injection and four H_2 withdrawal cycles alternate:
	– In each injection cycle, H_2 is injected for 90 days, with a subsequent well shut-in for 90 days; Operating target: surface flow rate (RATE) at 6.0×10^5 sm^3/day
	– In each withdrawal cycle, H_2 is produced for 90 days, with a following well shut-in for 15 days; Operating target: gas rate (GRAT) at 6.0×10^5 sm^3/day

when dealing with non-local and time-lapse (or 4D) observations. Additional advantages of correlation-based localization include: (1) It is adaptive and can automatically take into account the effects of various factors, such as the change of fluid dynamics in the subsurface, the use of different ensemble sizes in the IES algorithm, the difference in the combinations of distinct types of model variables and field data, and so on [32,34]; (2) The implementation of correlation-based localization is straightforward and can be conveniently transferred among different case studies.

With these practical benefits said, a correlation-based automatic and adaptive localization (AutoAdaLoc) scheme developed in [35] is applied to the IES algorithm in the case study later, so that the model update formula becomes

$$\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \left(\mathbf{T}^i \circ \mathbf{K}^i \right) \left(\tilde{\mathbf{d}}_j^{obs} - \tilde{\mathbf{g}} \left(\mathbf{m}_j^i \right) \right), \tag{11}$$

where $\mathbf{K}^i$ is a matrix in the form of, e.g., $\mathbf{K}^i = \mathbf{S}_m^i \mathbf{V}_r^i \Sigma_r^i \left((\Sigma_r^i)^2 + \gamma^i \mathbf{I}_r \right)^{-1} \left(\mathbf{U}_r^i \right)^T$ for the update formula in Eq. (10), ∘ stands for Schur product, and $\mathbf{T}^i$ is a tapering matrix whose elements are computed following a certain rule. To avoid distraction, we leave into Appendix A the technical details of how to construct the tapering matrix $\mathbf{T}^i$ in the AutoAdaLoc scheme.

2.4. Dealing with big models and big data in the IES

The issue of big models and big data is relevant to the current work, as in the case study later, 4D seismic surveys are conducted to improve the quality of estimated subsurface models, which leads to a relatively big seismic data set for subsurface characterization and monitoring. Given the high dimensionality of the subsurface model and the field data, applying the update formula of the IES algorithm would require substantial computational resources. This point can be seen in Eq. (11), where the update formula involves the Schur product $\mathbf{T}^i \circ \mathbf{K}^i$. Since both matrices $\mathbf{T}^i$ and $\mathbf{K}^i$ are in the dimension of model size by data size, with big models and big data, it would then require a huge amount of computer memory to store $\mathbf{T}^i$ and $\mathbf{K}^i$ in a straightforward implementation of Eq. (11).

In the literature, there are mainly two categories of strategies to deal with the above-noticed problem. In one category, the main idea is to divide a subsurface model into a number of sub-group model variables, and each sub-group of model variables is updated separately [30,36,37]. In this case, it would be necessary to either calculate parts of the matrices $\mathbf{T}^i$ and $\mathbf{K}^i$ (e.g., each row) on-the-fly [30], or re-compute the matrices $\mathbf{T}^i$ and $\mathbf{K}^i$ for each sub-group of model variables [36,37]. In another category, the essential notion is to conduct dimentionality reduction by finding a (much) lower-dimensional representation of the subsurface model, the field data, or both [38–40].

The lower-dimensional representations are then used in the update formula Eq. (11), so that the dimensionality of the corresponding matrices $\mathbf{T}^i$ and $\mathbf{K}^i$ (accordingly, the required computer memory) can be significantly reduced.

In this work, we follow the previous work [41] and adopt a 3D wavelet-based sparse representation procedure to find a lower-dimensional representation of the 4D seismic data. In comparison to similar sparse representation methods in other work (e.g., 38,40), the advantages of the wavelet-based sparse representation procedure include its computational efficiency, and the ability to provide a reasonably good estimation of the observation noise level in the seismic data (as a by-product of the sparse representation procedure). For brevity, the technical details of the wavelet-based sparse representation procedure are omitted in the current work. Readers are referred to the work of [41] for more elaborations.

3. Application of the ensemble-based workflow to an underground hydrogen storage (UHS) problem

3.1. Case study settings

Table 2 presents a brief summary of the information regarding the numerical subsurface model, which is taken from the UHS study of [8], yet with some minor modifications to facilitate its integration into the ensemble-based workflow. This study considers a UHS scenario in an unsaturated aquifer (gas reservoir) that contains the water and gas phases. The initial gas in place is comprised of nitrogen (N_2), CO_2, methane (C_1), ethane (C_2), propane (C_3) and normal butane (NC_4), which can be used as part of the cushion gas for UHS. In general, depending on geological and technical considerations, the typical depth of a natural storage site (aquifer or depleted hydrocarbon field) for UHS may roughly vary from 300 m to 2700 m [5,42]. The depth of the storage site in the current case study falls within this typical range, with the mean depth being around 1166.5 m and the associated standard deviation (STD) of depth being around 18.3 m.

In terms of the number of gridblocks, the size of the numerical model is $61 \times 61 \times 12$, whereas the size of a gridblock is 50.0 m $\times$ 50.0 m $\times$ 2.2 m on average. There is only a single well (labeled as "WELL3") near the center of the field, which is used for both H_2 injection and withdrawal. In this work, ECLIPSE© 300 is adopted as the numerical simulator (for this reason, ECLIPSE keywords will be adopted to describe several quantities of interest in the text below). The relative permeability and capillary pressure curves used in simulations are plotted in Fig. 2, with the associated critical gas saturation being 0.05 (i.e., 5%). The simulation time is 1500 days in total. During the first 360 days, only N_2 is injected as the cushion gas. After that,
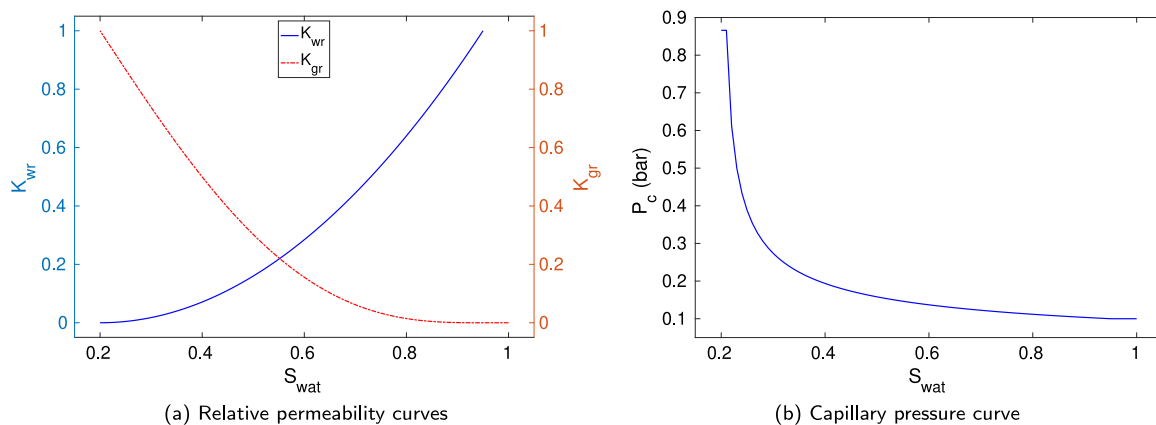
Fig. 2. (a) Relative permeability of water (K_{wr}) and gas (K_{gr}) and (b) capillary pressure (P_c, in the unit of bar) as functions of water saturation (S_{wat}).

Table 3
Summary of history matching settings in the UHS case study.

History matching settings	
Production data	Well bottom hole pressure (WBHP), field gas injection rate (FGIR), field gas production rate (FGPR), field molar injection (FCMIR) and production (FCMPR) rates of water, N_2 and H_2. Total number of production data = 468
4D seismic data	Amplitude versus angle (AVA) data from near- and mid-offsets at three survey times. Total number of seismic data (before dimensionality reduction) $\approx 4.5 \times 10^6$
Dimensionality reduction	3D wavelet-based sparse representation as in [41]. Total number of seismic data (after dimensionality reduction) = 5010
History matching algorithm	IES with correlation-based adaptive localization [35]
Initial ensemble	Containing 100 subsurface models that are generated through a sequential Gaussian simulation algorithm [43]
Parameters to be estimated	Each subsurface model containing permeability along the X direction in the natural logarithmic scale (log PERMX) and porosity (PORO) on all active gridblocks, as the parameters to be estimated (permeability along Y- and Z-directions are assumed to be proportional to permeability along the X direction). Total number of parameters (model size m) = 89 304

alternating H_2 injection and withdrawal cycles start, with more details elaborated in Table 2. The critical temperature and pressure of H_2 are 33.2K and 13.0 bar, respectively. With the settings in Table 2, H_2 is thus injected and contained inside the subsurface in a supercritical state. During the H_2 injection cycle, the injected gas is pure H_2. Other H_2 mixtures (e.g., in the form of town gas) can also be considered, although not investigated in the current work.

Table 3 reports the settings of history matching in this case study. During history matching, the main objective is to estimate an ensemble of subsurface models through the IES algorithm, in such a way that the average data mismatch (cf. Eq. (13)) of the estimated ensemble is reasonably reduced, in comparison to that of the initial ensemble of subsurface models. As indicated in Table 3, the uncertain parameters in a subsurface model include permeability along the x direction in the natural logarithmic scale (log PERMX) and porosity (PORO) on all active gridblocks. Therefore, the total number of parameters to be estimated by the IES algorithm (hence the model size m) is $2 \times 44652 = 89304$.

To initialize the ensemble-based workflow, an initial ensemble of 100 subsurface models is randomly generated through a sequential Gaussian simulation (SGSIM) algorithm [43], so that at the field scale, heterogeneities and uncertainties of model variables are taken into account. The initial ensemble is then updated by the IES algorithm, by conditioning subsurface models on available field data (cf. Eq. (11)). To mitigate the adversary effects of the relatively small ensemble size, the IES algorithm is equipped with a correlation-based adaptive localization scheme, which was introduced in the previous sub- Section 2.3. As an iterative algorithm, the IES needs to have certain stopping criteria for runtime control. In this work, the following two criteria are adopted: (1) the maximum number of iteration steps reaches 10; (2) the change of average data mismatch in two consecutive iteration steps is less than 0.01%. The IES will stop if either of these two criteria is satisfied. For the case study herein, only the first criterion is triggered in our

numerical investigations. As such, the final ensembles of subsurface models are those generated at the 10-th iteration steps.

In this work, there are two categories of field data used in history matching. One constitutes 9 types of production data, including well bottom hole pressure (WBHP), field gas injection rate (FGIR), field gas production rate (FGPR), field molar injection (FCMIR) and production (FCMPR) rates of water, N_2 and H_2, collected from the well WELL3 at 52 report time steps. Accordingly, the total number of production data is $9 \times 52 = 468$. In this synthetic case study, the production data are generated by inputting a reference subsurface model into ECLIPSE© 300 to run a forward simulation, and then collecting respective production data at specified time steps. In addition, to mimic the situation in practice, a sample of zero-mean Gaussian white noise is then added to each production-data point, where the STD of the noise is 5% of the magnitude of the data point. As will be shown later, when only using these production data in the IES algorithm (referred to as the "Prod" scenario for convenience of discussion later), the performance of history matching is not entirely satisfactory, in the sense that the IES algorithm is not able to make substantial updates of the initial ensemble of subsurface models, largely due to the limited information contents contained in the production data from the single well in the field.

To tackle the issue of limited information in production data, another category of field data, 4D seismic, is also adopted in history matching together with production data, which is referred to as the "ProdSeis" scenario for distinction. In this case, the seismic attribute is the amplitude versus angle (AVA) data from near- (10°) and mid-offsets (20°) at three survey times, on day 30, 735, and 1455, respectively. Appendix B explains the procedure to conduct forward simulations to generate AVA data, whereas the observations of 4D AVA data used in history matching are generated by plugging into the forward AVA simulator relevant static and dynamical properties (e.g., porosity, pressure and saturation fields etc.) of the reference subsurface model at three survey times, similar to the way used to obtain the observations of production data.
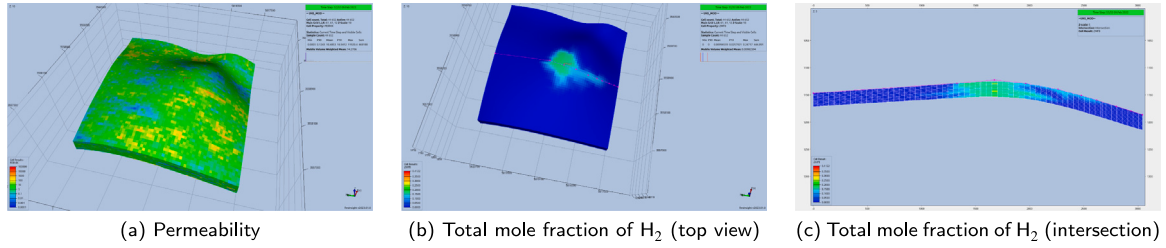
Fig. 3. Fields of (a) permeability along the X-direction and (b–c) total mole fraction (ZMF) of H_2 of the reference model at the end of the simulation period (day 1500). For better visualization, Panels (b) and (c) present a top view and a 2D intersection of the ZMF field, respectively. Figures are produced by ResInsight©.

For each seismic survey, the dimension of either the near- or mid-offset AVA data set is in the dimension of $61 \times 61 \times 201$, for the reason explained in Appendix B. Therefore, the total number of the 4D seismic data set is $3 \times 2 \times 61 \times 61 \times 201 \approx 4.5 \times 10^6$, which, together with the relatively big model size ($m = 89304$), makes it difficult for the IES algorithm to directly use the 4D seismic data for subsurface characterization and monitoring. To solve this problem, the 3D wavelet-based dimensionality reduction procedure in [41] is adopted, so that the original 4D seismic data set can be sparsely represented by a set of 5010 leading wavelet coefficients in the wavelet domain. After this dimensionality reduction procedure, it is these 5010 leading wavelet coefficients that will be used as the effective observations in history matching. With this data processing procedure, the dimensionality of the effective observations becomes manageable by the IES algorithm. As a side remark, zero-mean Gaussian white noise is also introduced to the original 4D AVA data, with the noise STD being 30% of the magnitudes of the data. This noise STD is assumed unknown, whereas a wavelet-based algorithm can be employed to estimate the noise STD of the leading wavelet coefficients, as a by-product of the wavelet-based dimensionality reduction procedure [41].

In the current work, two metrics, namely, average data mismatch (DM) and average root mean squared error (RMSE), are used to monitor or measure the performance of the IES algorithm during history matching. Intuitively, DM (RMSE) measures the discrepancies between simulated field data (estimated subsurface model) and real observations (reference subsurface model) in the observation (model) space. As mentioned in Section 2, the information of the average DM at each iteration step will be utilized to determine how the iteration should be carried out. Meanwhile, other than generating the observations of production and seismic data, the reference model is not used elsewhere for subsurface characterization and monitoring. As such, the information of the reference subsurface model (hence that of the average RMSE) is not directly exploited by the IES algorithm during history matching. Instead, in this case, the average RMSE is mainly used to cross-validate the performance of the IES algorithm, after history matching is finished. Since RMSE calculates the distances between estimated and reference subsurface models, and is a metric for performance cross-validation, it serves as a more suitable metric than DM to evaluate the qualities of estimated subsurface models.

Specifically, the DM $\zeta\left(\mathbf{m};\mathbf{d}^{obs}\right)$ of a generic $\mathbf{m}$ with respect to the observation $\mathbf{d}^{obs}$ is defined as

$$\zeta\left(\mathbf{m};\mathbf{d}^{obs}\right) \equiv \left(\mathbf{d}^{obs} - \mathbf{g}(\mathbf{m})\right)^T \mathbf{C}_d^{-1}\left(\mathbf{d}^{obs} - \mathbf{g}(\mathbf{m})\right). \tag{12}$$

Accordingly, the average DM $\hat{\zeta}^i\left(\mathbf{M}^i;\mathbf{d}^{obs}\right)$ of an ensemble of subsurface model $\mathbf{M}^i = \{\mathbf{m}_j^i\}_{j=1}^{N_e}$ at the ith iteration step is

$$\hat{\zeta}^i\left(\mathbf{M}^i;\mathbf{d}^{obs}\right) \equiv \frac{1}{N_e} \sum_{j=1}^{N_e} \zeta\left(\mathbf{m}_j^i;\mathbf{d}^{obs}\right). \tag{13}$$

On the other hand, the RMSE $\varepsilon\left(\mathbf{m};\mathbf{m}^{ref}\right)$ of a generic model $\mathbf{m}$ with respect to the reference model $\mathbf{m}^{ref}$ is written as

$$\varepsilon\left(\mathbf{m};\mathbf{m}^{ref}\right) \equiv \|\mathbf{m} - \mathbf{m}^{ref}\|_2 / \sqrt{m}, \tag{14}$$

where the operator $\|\cdot\|_2$ returns the euclidean norm of its operand •, and m is the dimension of the model $\mathbf{m}$. Consequently, the average RMSE $\hat{\varepsilon}^i\left(\mathbf{M}^i;\mathbf{m}^{ref}\right)$ of the ensemble $\mathbf{M}^i = \{\mathbf{m}_j^i\}_{j=1}^{N_e}$ at the ith iteration step becomes

$$\hat{\varepsilon}^i\left(\mathbf{M}^i;\mathbf{m}^{ref}\right) \equiv \frac{1}{N_e} \sum_{j=1}^{N_e} \varepsilon\left(\mathbf{m}_j^i;\mathbf{m}^{ref}\right). \tag{15}$$

3.2. Numerical results

In the sequel, we compare numerical results in the Prod and ProdSeis scenarios from a few different angles, including history matching performance in terms of DM and RMSE, predicted profiles of certain types of production data, H_2 storage performance, and visual inspections on some estimated subsurface models and the corresponding predicted spatial distributions of H_2.

3.2.1. Data mismatch (DM) and root mean squared error (RMSE)

Fig. 3 shows the fields of (a) permeability along the X-direction, and (b–c) total mole fraction (ZMF) of H_2 on 09 Feb. 2022 (day 1500), with respect to the reference model. As can be noticed in Figs. 3(b) and 3(c), a part of the H_2 plume enters into a strip of subsurface zone and stays relatively far away from the wellbore, which is often referred to as the fingering phenomenon in the literature [4]. This phenomenon is likely due to the relatively high permeability values distributed on the same strip of the subsurface model, as can be seen in Fig. 3(a). For the purpose of UHS, the migration of the H_2 plume into a region relatively far away from the wellbore is not desired, as this may reduce the deliverability of H_2 during the withdrawal cycles. In a practical situation, however, the true subsurface conditions (hence the corresponding spatial distribution of the H_2 plume) are typically unknown at the field scale. As such, from the perspective of subsurface characterization and monitoring, an important question is that when the reference model is not directly exposed to the history matching workflow, can the IES algorithm help improve subsurface understanding and make a better prediction of the spatial distribution of the H_2 plume? As will be shown through the numerical results below, the answer to this question appears to be dependent on which kinds of field data are available for history matching.

The above point is reflected by the history matching results in Table 4, where the values of DM and RMSE (in terms of mean/average $\pm$ STD) of the initial ensemble and the final ones obtained in both Prod and ProdSeis scenarios are presented. The additional messages from Table 4 include: (a) In comparison to the average DM of production data of the initial ensemble, the IES algorithm can substantially reduce the average DM of production data in the final ensembles of both Prod and ProdSeis scenarios. Meanwhile, the efficacy of the IES algorithm in abating the average DM of seismic data appears less substantial; (b) In terms of average RMSE (especially that of porosity), using production data alone in the IES algorithm can only marginally improve the qualities of subsurface models, relative to those in the initial ensemble, whereas more model improvements can be obtained by conditioning subsurface models on both production and seismic

Table 4
Data mismatch (DM) and root mean squared error (RMSE), in terms of ensemble mean $\pm$ standard deviation (STD), with respect to the initial ensemble of subsurface models, and the final ones obtained in the Prod and ProdSeis scenarios. Note that seismic data are not used in the Prod scenario, thus DM of seismic data is not available (N/A) therein.

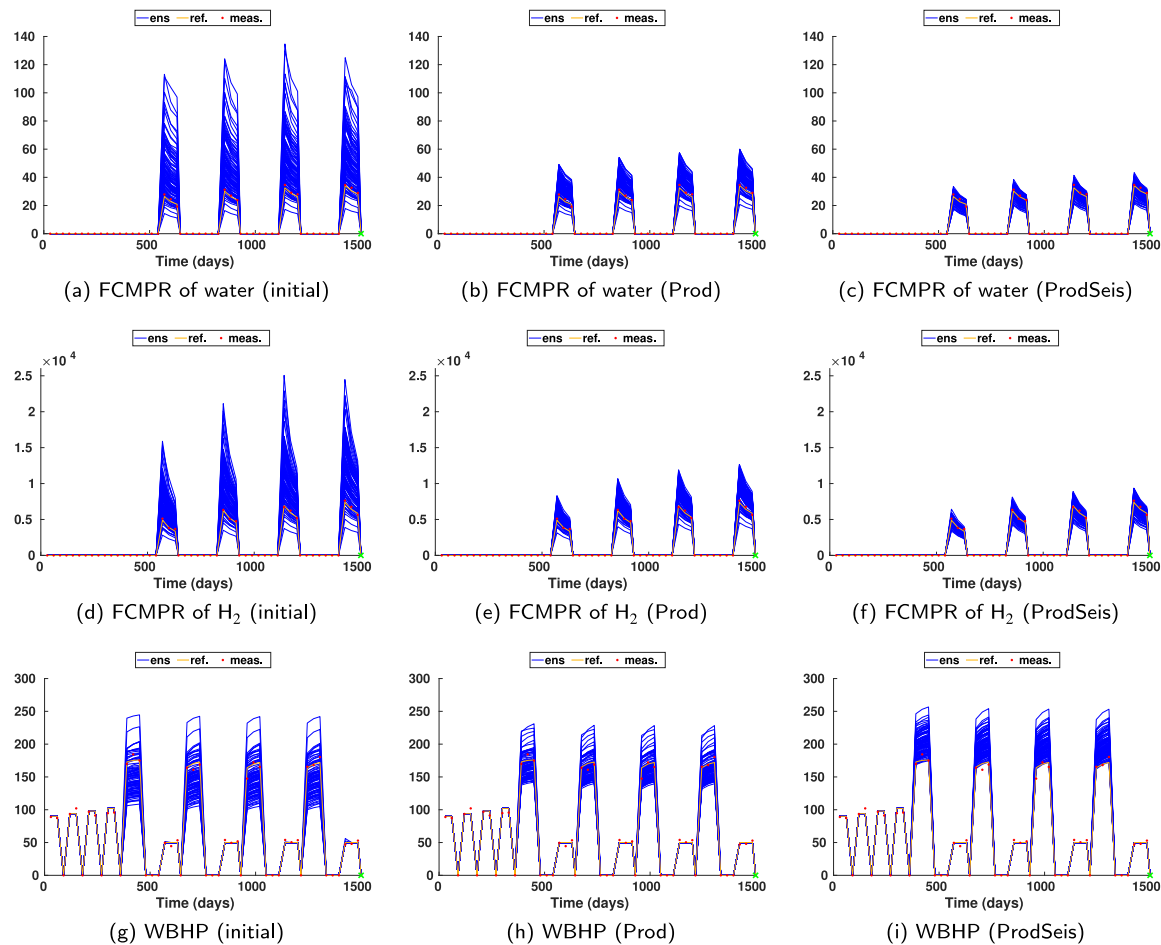|	DM of production data	DM of seismic data	RMSE of log PERMX	RMSE of PORO
Initial ensemble	$(0.61 \pm 1.38) \times 10^6$	$(3.74 \pm 0.64) \times 10^5$	2.8115 ± 0.0731	0.0931 ± 0.0023
Final ensemble (Prod)	$(7.13 \pm 7.30) \times 10^4$	N/A	2.7794 ± 0.0783	0.0930 ± 0.0023
Final ensemble (ProdSeis)	$(3.11 \pm 1.87) \times 10^4$	$(3.49 \pm 0.35) \times 10^5$	2.5300 ± 0.0644	0.0888 ± 0.0023

Fig. 4. Field molar production (FCMPR) rates of water (top row) and H2 (middle row), and well bottom hole pressure (WBHP) (bottom row) obtained at different simulation time, with respect to the initial ensemble (first column), and the final ensembles obtained in the Prod (second column) and ProdSeis (third column) scenarios, respectively. For FCMPR of water and H_2, their units are "kg-M/day" (kilogram-mole per day), and for WBHP, its unit is "bar". In each sub-figure, the observed field data at different report time steps are marked by red dots, the field data generated by the reference model (without observation noise) are represented by an orange curve, whereas the field data corresponding to the initial or final ensemble of subsurface models are depicted by blue curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

data; (c) In the ProdSeis scenario, the IES algorithm tends to result in lower uncertainties (in terms of ensemble standard deviations) in both DM and RMSE values, in comparison to those of the initial ensemble and the Prod scenario.

3.2.2. Production data profiles

For further illustration, Fig. 4 shows the profiles of FCMPR of water and H_2, and WBHP. Quantities presented in each sub-figure include the observed production data at different report time steps (red dots), which are parts of the observations used in history matching; production data from the reference model without any observation noise (orange curve); and a set of simulated production data generated from an ensemble of initial or final subsurface models (blue curves). For FCMPR of water and H_2, it is clear that after history matching, the simulated data match their respective observations better than those of the initial ensemble, regardless of whether seismic data are used in history matching or not. It is also evident that the simulated FCMPR

data of water and H_2 in the ProdSeis scenario fit the observations better than those in the Prod scenario. In the meantime, for WBHP, the simulated data corresponding to the Prod scenario have a better agreement with the observations than those of the initial ensemble. In the ProdSeis scenario, during the cycles of H_2 injection, the simulated data tend to provide a higher estimation of WBHP than those in the Prod scenario, hence a worse performance in terms of data coverage.[2] The comparison of the results in Fig. 4 thus indicates the complexity of history matching with multiple types of field data: Often adopting an additional type of field data in history matching can help improve the data match of some other types of field data, but not necessarily all of them. Note that in this case study, with the wavelet-based sparse representation

---

[2] From the perspective of uncertainty quantification, it would be desirable that the profile of simulated data has a good coverage of the underlying observations.
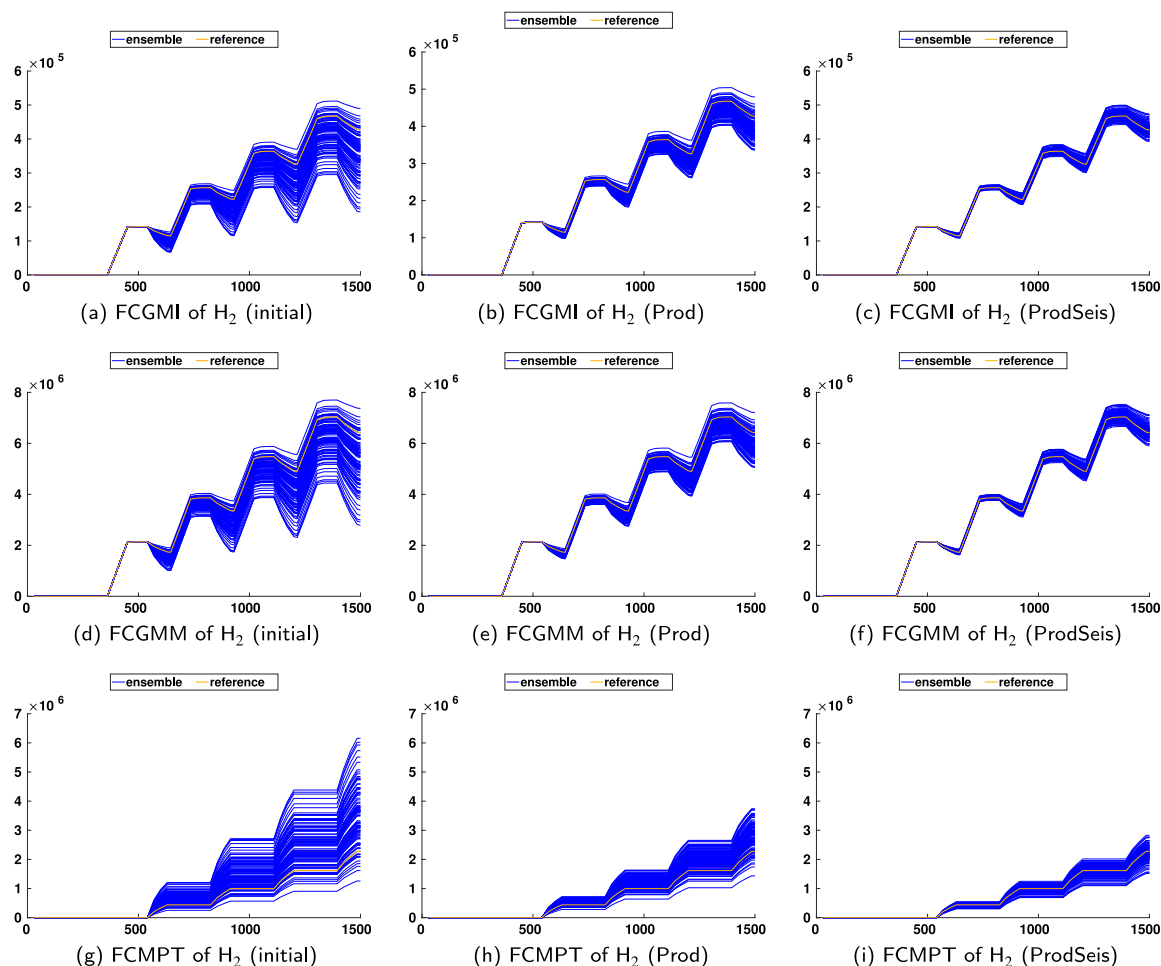
Fig. 5. Field molar amount of H$_2$ trapped in gas (FCGMI) (top row), field molar amount of H$_2$ mobile in gas (FCGMM) (middle row) and field molar production totals of H$_2$ (FCMPT) (bottom row), at different simulation time. The units of FCGMI, FCGMM and FCMPT are "kg-M" (kilogram-mole). Similar to the settings in Fig. 4, in each sub-figure, data of the reference model are plotted along an orange curve, whereas those of the initial or final ensemble are connected by blue curves. Note that the quantities presented here, FCGMI, FCGMM and FCMPT of H$_2$, are not used as observations in history matching. Therefore, unlike the situation in Fig. 4, there are no red dots in each sub-figure here. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

procedure, seismic data are represented by a set of leading wavelet coefficients in the wavelet domain. As such, it becomes complicated to interpret the changes of these leading wavelet coefficients. For this reason, the profiles of seismic data (similar to those of production data in Fig. 4) are not included here, whereas some sample sections of the observed seismic data are shown in Appendix B (cf. Fig. 9).

Additional simulated data, including those of field molar amount of H$_2$ trapped in gas (FCGMI), field molar amount of H$_2$ mobile in gas (FCGMM) and field molar production totals of H$_2$ (FCMPT), are displayed in Fig. 5. Note that the above-mentioned types of data are not used in history matching. Consequently, a comparison on the profiles of these data can be considered as an additional way to cross-validate the performance of history matching. In Fig. 5, one can see that after history matching, all three types of the simulated data from either the Prod or the ProdSeis scenario can provide better predictions of respective reference data with less uncertainties, as reflected by the quantitative results on day 1500 in Table 5 later. In particular, introducing seismic data to history matching can further improve the prediction accuracies in the ProdSeis scenario, in comparison to those in the Prod scenario.

3.2.3. H$_2$ storage performance

The simulation results in Fig. 5 can also be used to assess the ability of an ensemble of subsurface models to predict the performance of a

UHS storage project. For illustration, Table 5 reports the actual and predicted values of molar total H$_2$ injected, total H$_2$ recovered, H$_2$ trapped (immobile), and H$_2$ in place (mobile) at the end of the simulation time window, with respect to the reference model, and those of the initial and final ensembles, respectively. Here, the numbers of total injected and recovered H$_2$ are described by the values associated with the keywords FCMIT and FCMPT in ECLIPSE© 300, respectively. In particular, because the operating target during the H$_2$ injection cycles is surface flow rate (cf. Table 2), the total amount of injected H$_2$ is the same in all subsurface models (including reference model and those of the initial and final ensembles). Meanwhile, H$_2$ trapped (immobile) corresponds to the amount of H$_2$ that are captured and consequently become immobile in the subsurface through various mechanisms, e.g., solubility trapping, mineral trapping and residual trapping, similar to the situation in CO$_2$ storage problems [44]. In this UHS case study, since the mineral trapping mechanism is not considered, and the solubility of H$_2$ in saline water is very low (and thus negligible), the amount of H$_2$ trapped is mainly due to the mechanism of residual trapping, and is depicted by the keyword FCGMI in ECLIPSE© 300. Finally, in the last column of Table 5, H$_2$ in place (mobile) is the fraction of H$_2$ that is in the gas phase and remains mobile in the subsurface. This portion of H$_2$ can serve as the cushion gas for future H$_2$ injections (after day 1500), but may or may not be recovered during subsequent withdrawal cycles. In ECLIPSE© 300, H$_2$ in place is defined by the keyword FCGMM.

Table 5
Assessment of UHS performance by analyzing the allotments of H_2, in terms of molar total H_2 injected, recovered, trapped, and in place (in kilogram-mole, kg-M), on day 1500. The simulation results of the reference model are included as the baselines, whereas those corresponding to the initial or final ensemble of subsurface models are presented in the form of mean $\pm$ STD, below which the percentages of deviations of the ensemble means from respective baselines are presented in parentheses.

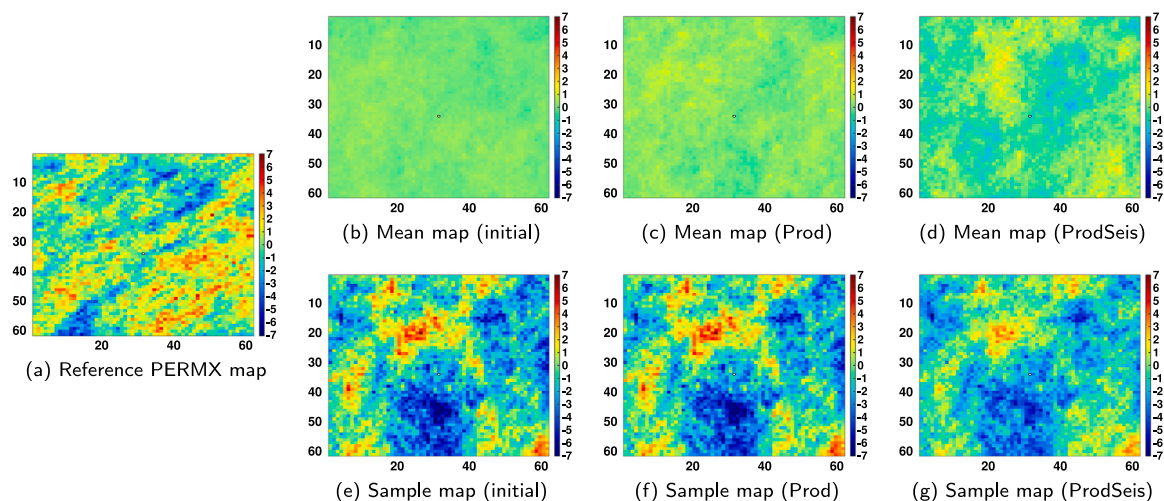|	Total H_2 injected (FCMIT, kg-M)	Total H_2 recovered (FCMPT, kg-M)	H_2 trapped (immobile) (FCGMI, kg-M)	H_2 in place (mobile) (FCGMM, kg-M)
Reference model		2.27×10^6	0.43×10^6	6.42×10^6
Initial ensemble	9.12×10^6	$(3.52 \pm 1.03) \times 10^6$ $(+55.1\%)$	$(0.35 \pm 0.06) \times 10^6$ (-18.6%)	$(5.25 \pm 0.97) \times 10^6$ (-18.2%)
Final ensemble (Prod)		$(2.80 \pm 0.47) \times 10^6$ $(+23.3\%)$	$(0.39 \pm 0.03) \times 10^6$ (-9.3%)	$(5.93 \pm 0.44) \times 10^6$ (-7.6%)
Final ensemble (ProdSeis)		$(2.22 \pm 0.28) \times 10^6$ (-2.2%)	$(0.43 \pm 0.02) \times 10^6$ (0.0%)	$(6.47 \pm 0.23) \times 10^6$ $(+0.8\%)$

Fig. 6. Log PERMX maps on Layer 2 of (a) the reference model, (b) the mean model of the initial ensemble, (c) the mean model of the final ensemble in the Prod scenario, (d) the mean model of the final ensemble in the ProdSeis scenario, (e) a sample model (ensemble member) of the initial ensemble, (f) updated sample model obtained in the Prod scenario, and (g) updated sample model obtained in the ProdSeis scenario. In each map, there is a small circle indicating the location of the well "WELL3".

With the above definitions, it can be seen that in Table 5, the amount of total H_2 injected equals the summed amount of total H_2 recovered, H_2 trapped (immobile) and H_2 in place (mobile). It also indicates that on average, the subsurface models in the initial ensemble over-estimate (more than 50%) the amount of H_2 that can be recovered, while under-estimating (more than 18%) the amount of H_2 remaining in the subsurface. Using production data to calibrate the models in the initial ensemble helps mitigate this issue, leading to the respective percentage of over- or under-estimation more than halved, whereas including both production and seismic data in history matching achieves the best performance, making the respective percentage of over- or under-estimation become no more than 2.2%, apart from reduced uncertainties.

Based on the information in Table 5, one can calculate the average recovery factor of H_2 on day 1500, in terms of the ratio of average total H_2 recovered to total H_2 injected, which are 24.9%, 38.6%, 30.7%, and 24.4%, for the reference model, the models of the initial ensemble, and those of the final ensembles in the Prod and ProdSeis scenarios, respectively. These recovery factors appear to be relatively low, given the settings in the current study. One may further improve the recovery factor by, for instance, adding more wells in the field with a suitable spatial pattern [42], which, however, is a topic beyond the scope of this work. Meanwhile, one can also estimate the minimum average loss factor of H_2 on day 1500, as the ratio of average H_2 trapped (immobile) to total H_2 injected, which are 4.7%, 3.8%, 4.3% and 4.7%, accordingly. Comparing the numbers with respect to the initial and final ensembles, it is evident that among others, the subsurface models obtained in the ProdSeis scenario provide the best prediction of the actual performance of UHS in the reference model, in terms of the average recovery and loss factors of H_2.

3.2.4. Estimated subsurface models and the corresponding H_2 distributions

Figs. 6 and 7 proceed to compare log PERMX and PORO, respectively, in the subsurface models before and after history matching. Since there are 100 models in each ensemble, while each model contains 12 vertical layers, it would be tedious to include the information of all individual models. For this reason, here we choose to plot in Figs. 6 and 7, respectively, the log PERMX and PORO maps on Layer 2 of the mean models of the initial and final ensembles (Panels (b) – (d)). We also show the log PERMX and PORO maps on Layer 2 of an ensemble member (a sample model) in the initial ensemble, and the same ensemble member obtained in the Prod and ProdSeis scenarios, respectively (Panels (e) – (g)). In this way, an inspection on the differences among the log PERMX and PORO maps of both the mean model and the sample model in the initial and final ensembles would help analyze the effects of different field data on history matching. In addition, the corresponding maps of the reference model (Panel (a)) are also included as the baselines for comparison.

The results in Figs. 6 and 7 indicate that when only using production data in history matching, the subsurface models in the final ensemble stay close to those in the initial ensemble. In other words, the IES algorithm cannot make significant changes to the model variables in the initial ensemble (especially for PORO), which is consistent with the results of RMSEs in Table 4. The main reason behind this observation is that in this particular case study, there is only a single well "WELL3" in the field, which does not provide a good spatial coverage of the subsurface. Accordingly, production data from "WELL3" do not contain substantial information for model update. From the perspective of sensitivity analysis, this means that production data are not sensitive to many model variables in the whole field, especially PORO parameters, as was also found in some previous work [39,41].
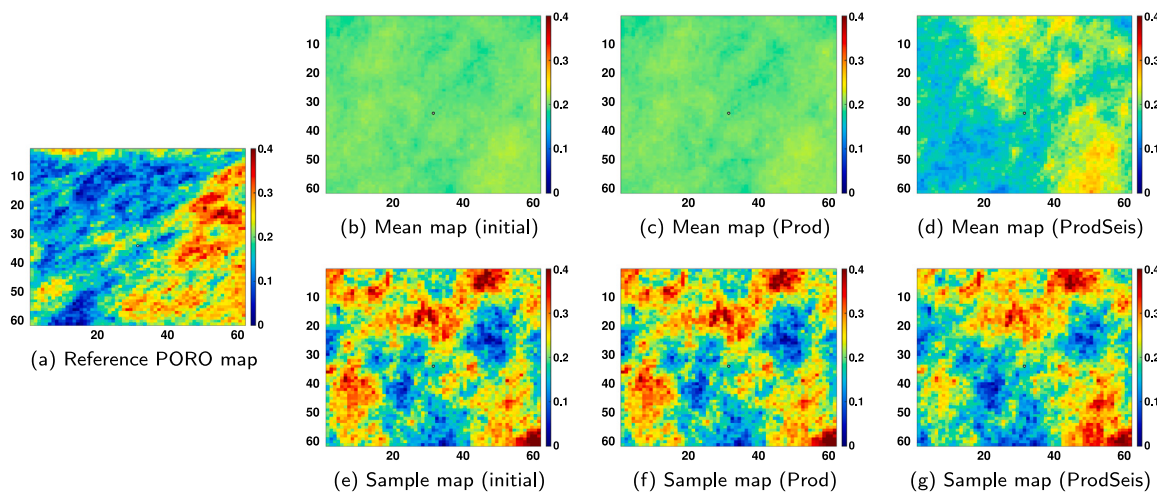
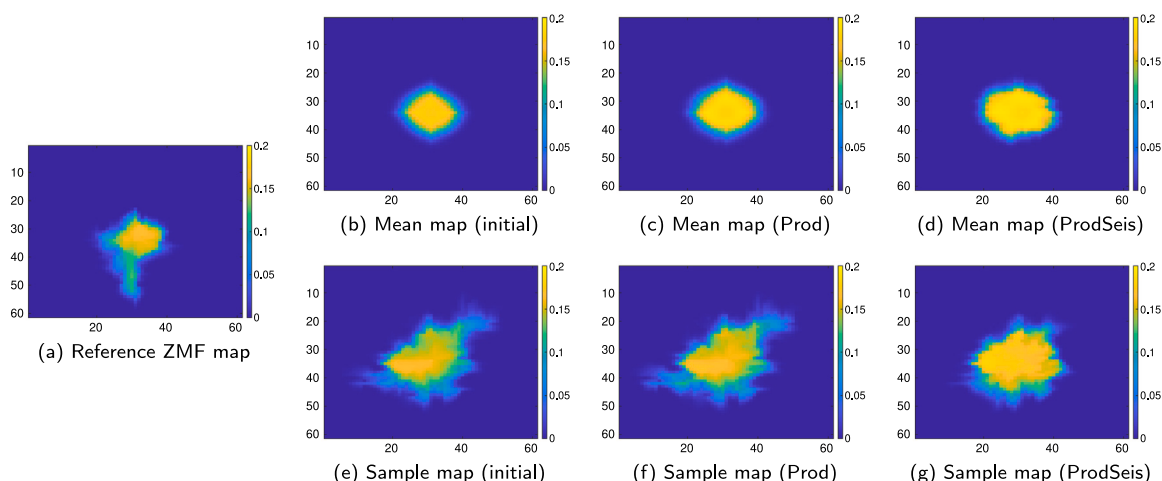**Fig. 7.** As in Fig. 6, but for PORO maps.

Fig. 8. As in Fig. 6, but for ZMF maps of H_2 (on day 1500) at Layer 1 of respective models. Note that here a "mean map" corresponds to the result obtained by running a mean subsurface model of an initial or final ensemble.

In contrast, when both production and seismic data are incorporated into history matching, more substantial differences between the mean or sample models of the initial and final ensembles can be noticed, also in line with the results of RMSEs in Table 4. The main reason behind these more noticeable model changes is that seismic data provide a more extensive spatial coverage than production data, as elaborated in Fig. 9 and the corresponding text in Appendix B. On top of the improved spatial coverage, there is an explicit dependence of seismic data on PORO parameters in the forward seismic simulator, as described in Appendix B. Moreover, there is also an implicit relation of seismic data to PERMX and PORO parameters. This is because the forward seismic simulator uses the information of fluid pressure and saturations in the whole field to simulate seismic data, whereas these pressure and saturation profiles are obtained through the numerical simulator ECLIPSE© 300, which in turn takes a subsurface model (consisting of PERMX and PORO parameters) as the input. As such, when seismic data are included in history matching, the IES algorithm makes more considerable changes to not only PERMX, but also PORO parameters, and results in improved performance of subsurface characterization and monitoring.

Finally, Fig. 8 presents ZMF maps of H_2 (on day 1500) at Layer 1 of subsurface models, with the settings therein similar to those in Figs. 6 and 7. For the same reasons explained in the preceding paragraphs, one can also observe that the ZMF maps in the Prod scenario remain closer to those of the initial ensemble, in comparison to the maps in

the ProdSeis scenario. Including seismic data in history matching helps improve the prediction of the distribution of the H_2 plume. This point is more evident when comparing the sample maps in Panels (e) – (g): In the sample maps of both the initial ensemble and the Prod scenario, there are incorrectly predicted pinch-outs of the H_2 plume along the upper right and lower left directions. In the corresponding map of the ProdSeis scenario, the wrong predictions are corrected to some extent, and a large part of the previously predicted pinch-outs has vanished. Meanwhile, in comparison to the reference ZMF map in Panel (a), none of the maps from either the initial or the final ensembles can provide a correct prediction of the fingering phenomenon along the downward direction, which indicates that even in the presence of seismic data, there is no guarantee that the IES algorithm can achieve a perfect estimation of the underlying truth (reference model in this case). This deficiency can be attributed to a few potential issues. Among others, one is the attenuation of seismic waves when they propagate through the subsurface [45], which leads to reduced sensitivity of seismic data to the formation properties in deeper subsurface. Another is that for the purpose of runtime control, in practice it is not possible to run a history matching algorithm with too many iteration steps. As such, often the history matching algorithm could end up with a local minimum that may not be sufficiently close to the underlying truth. With this said, the results presented here suggest that there is still room to further improve the performance of subsurface characterization and monitoring within the adopted ensemble-based workflow.

4. Discussion and conclusion

This work studies the underground hydrogen storage (UHS) problem in a 3D case study from the perspective of subsurface characterization and monitoring. To this end, we first present an ensemble-based history matching workflow, and apply it to the 3D UHS case study. In this case study, we investigate two scenarios: one of them only uses production data for history matching (labeled as "Prod" scenario); and the other incorporates both production and 4D seismic data (labeled as "ProdSeis" scenario). Through a set of numerical investigations, we show that

- the ensemble-based history matching workflow can handle heterogeneities and uncertainties at the field scale;
- In either the Prod or the ProdSeis scenario, the proposed workflow helps improve the qualities of an initial ensemble of subsurface models, and increase the prediction accuracies of certain quantities of interest (e.g., observed or unobserved production data, recovery and loss factors of H_2) while properly reducing the uncertainties of the predictions;
- Meanwhile, because production data typically have limited spatial coverage, including 4D seismic data helps mitigate this problem and thus improves the history matching performance.

In particular, the last bullet point is supported through the comparison of the numerical results in the Prod and ProdSeis scenarios, which indicates that in the ProdSeis scenario, the history matching algorithm tends to provide better predictions to various inspected quantities, such as different types of production data and model variables, than that in the Prod scenario does. One implication of this comparison result is that incorporating 4D seismic data would be beneficial for the development, operation and management of a UHS project. At the time of this writing, there only exist a few commercial UHS projects in salt caverns [2], for which 4D seismic (or other geophysical monitoring tools) does not appear to be commonly used. Nevertheless, according to the recent survey in [22], among others, 4D seismic is routinely adopted for subsurface characterization and monitoring in several existing geological CO_2 storage projects. Given the similarities between geological CO_2 storage and UHS projects, it is thus expected that 4D seismic would be among the most useful monitoring tools for UHS in natural storage sites. Meanwhile, we also stress that from the practical point of view, the optimal choice of monitoring tools should take into account site-specific risk factors, and would be typically project-dependent [22].

On the other hand, there is also a noticed issue in the ProdSeis scenario, that is, even after assimilating seismic data, the spatial distribution of the H_2 plume still cannot be correctly predicted (especially for the fingering phenomenon), which appears to be a common challenge when using seismic data for subsurface characterization and monitoring [22]. Hence, this observation leads to an open research question: can we further improve the history matching performance by including more types of field data? While in the literature there are some preliminary investigations in this regard (e.g., 46), there does not seem to be a definite answer to this question, especially in the context of UHS. As a rule of thumb, whenever possible, our recommendation is to include different types of field data, e.g., tracer, distributed acoustic sensing (DAS) or electromagnetic (EM), which may offer new information contents complementary to those in existing production and 4D seismic data. With the inclusion of more types of field data, however, it might be necessary that the IES algorithm be modified accordingly to mitigate the negative impacts of conflicting information contents from different sources [26,46]. Within this context, more efforts on history-matching algorithm developments are thus expected in our future investigations.

CRediT authorship contribution statement

Xiaodong Luo: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Svenn Tveit:** Writing – review & editing, Methodology, Conceptualization. **Raoof Gholami:** Writing – review & editing, Formal analysis, Conceptualization. **Pål Østebø Andersen:** Writing – review & editing, Formal analysis, Conceptualization.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgments

Appendix A. Technical details of correlation-based adaptive localization

For better elaboration of the localization technique, we re-write the model update formula, either Eq. (7) or Eq. (10), in a more compact form:

$$\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \mathbf{K}^i \left(\tilde{\mathbf{d}}_j^{obs} - \tilde{\mathbf{g}} \left(\mathbf{m}_j^i \right) \right), \tag{A.1}$$

where $\mathbf{K}^i$ is a matrix similar to the Kalman gain in a conventional Kalman filter [47], with $\mathbf{K}^i = \mathbf{S}_m^i (\tilde{\mathbf{S}}_g^i)^T \left(\tilde{\mathbf{S}}_g^i (\tilde{\mathbf{S}}_g^i)^T + \gamma^i \mathbf{I}_p \right)^{-1}$ in Eq. (7), and $\mathbf{K}^i = \mathbf{S}_m^i \mathbf{V}_r^i \boldsymbol{\Sigma}_r^i \left((\boldsymbol{\Sigma}_r^i)^2 + \gamma^i \mathbf{I}_r \right)^{-1} (\mathbf{U}_r^i)^T$ in Eq. (10).

To conduct localization in the IES algorithm, one replaces the Kalman-gain-like matrix $\mathbf{K}^i$ by the Schur product between $\mathbf{K}^i$ and a tapering matrix $\mathbf{T}^i$, leading to a further modified update formula:

$$\mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \left(\mathbf{T}^i \circ \mathbf{K}^i \right) \left(\tilde{\mathbf{d}}_j^{obs} - \tilde{\mathbf{g}} \left(\mathbf{m}_j^i \right) \right), \tag{A.2}$$

where ∘ stands for Schur product. As such, the focus of the localization technique is on how to determine the tapering matrix $\mathbf{T}^i$ to further improve the performance of history matching.

In the current work, the correlation-based adaptive localization (AdaLoc) scheme in [35] is adopted for its practical benefits (readers are referred to 35 for more information of these benefits). To see the main idea behind this localization scheme, we re-write Eq. (A.2) in an element-wise form, as follows:

$$m_{j,k}^{i+1} = m_{j,k}^i + \sum_{s=1}^p \left(t_{k,s}^i k_{k,s}^i \right) \Delta \tilde{d}_{j,s}^i, \tag{A.3}$$

where $m_{j,k}^i$ ($m_{j,k}^{i+1}$) corresponds to the kth element of $\mathbf{m}_j^i$ ($\mathbf{m}_j^{i+1}$), $\Delta \tilde{d}_{j,s}^i$ to the sth element of the innovation term $\Delta \tilde{\mathbf{d}}_j^i \equiv \tilde{\mathbf{d}}_j^{obs} - \tilde{\mathbf{g}} \left(\mathbf{m}_j^i \right)$, and $t_{k,s}^i$ and $k_{k,s}^i$ to the elements on the kth row and the sth column of the matrices $\mathbf{T}^i$ and $\mathbf{K}^i$, respectively.

In [35], the tapering coefficient $t^i_{k,s}$ is calculated by the following formula:

$$t^i_{k,s} = f_{GC}\left(\frac{1 - \text{abs}(\rho^0_{k,s})}{1 - \theta_{G_k,s}}\right), \text{ for } k = 1, 2, \cdots, m; s = 1, 2, \cdots, p, \quad \text{(A.4)}$$

where abs is the operator taking absolute values of its inputs,

$$f_{GC}(z) = \begin{cases} -\frac{1}{4}z^5 + \frac{1}{2}z^3 + \frac{5}{8}z^3 - \frac{5}{3}z^2 + 1, & \text{if } 0 \leq z \leq 1; \\ \frac{1}{12}z^5 - \frac{1}{2}z^4 + \frac{5}{8}z^3 + \frac{5}{3}z^2 - 5z + 4 - \frac{2}{3}z^{-1}, & \text{if } 1 < z \leq 2; \\ 0, & \text{if } z > 2, \end{cases}$$

$$\text{(A.5)}$$

and

$$\theta_{G_k,s} = \sqrt{2 \ln(\#\rho^0_{G_k,s})}\, \sigma_{G_k,s}; \quad \text{(A.6)}$$

$$\sigma_{G_k,s} = \frac{\text{median}(\text{abs}(\epsilon^0_{G_k,s}))}{0.6745}. \quad \text{(A.7)}$$

The tapering function f_{GC} in Eq. (A.4) corresponds to the Gaspari-Cohn function [48], which is defined in Eq. (A.5). Meanwhile, $\rho^0_{k,s}$ stands for the sample correlation between the kth model variable and the sth simulated observation data point, obtained from the initial ensembles of model variable $\{m^0_{j,k}\}^{N_e}_{j=1}$ and the corresponding simulated data point $\{[\tilde{g}(\mathbf{m}^0_j)]_s\}^{N_e}_{j=1}$, where $[\mathbf{x}]_s$ represents the sth element of a dummy vector $\mathbf{x}$. Eq. (A.4) implies that the tapering coefficients $t^i_{k,s}$ only depend on the initial ensemble of subsurface models, but not on the ensembles at later iteration steps. For this reason, the tapering coefficients $t^i_{k,s}$ are computed only once and for all model updates later.

In Eq. (A.6), one combines the universal rule [49] and the median absolute deviation (MAD) estimator [50] to calculate a common threshold value $\theta_{G_k,s}$ of a group (indexed by G_k) of model variables, for a given data point (indexed by s). In the implementation, the groups are created according to the types of the model variables. For instance, model variables corresponding to the porosity values on the gridblocks of a numerical subsurface model are put into the same group. Meanwhile, $\#\rho^0_{G_k,s}$ corresponds to the number of model variables (equivalently, the number of sample correlation coefficients) in a group for a given data point indexed by s, and $\epsilon^0_{G_k,s}$ stands for the substitute sampling errors in the group of sample correlation coefficients $\rho^0_{G_k,s}$, where the substitute sampling errors can be conveniently obtained by the random shuffle method proposed by [35].

Appendix B. Simulation of amplitude versus angle (AVA) data

AVA data depict the variation of seismic wave amplitudes over the angles (hence distances) between sources and receivers in seismic acquisitions. Typically, the change of seismic amplitudes is related to the modification of lithology or fluid contents in the subsurface formations. As a result, AVA data can be used to estimate static and dynamical properties (e.g., porosity, pressure and fluid saturations) of subsurface formations through a subsurface characterization and monitoring workflow [39,51,52]. For this purpose, it is necessary to build a forward AVA simulator that involves a few steps (cf. Fig 1 of [41]): (1) Run a numerical subsurface model to obtain the spatial distributions of pore pressure and fluid saturations in the field; (2) Input the pore pressure and fluid saturation fields into a petro-elastic model (PEM) to compute certain relevant seismic parameters, such as P- and S-wave velocities and formation/fluid densities in the subsurface; (3) Plug the calculated seismic parameters into an AVA equation to generate AVA attributes at different angles [53].

In this work, a soft sand model from [53] is used as the PEM. This model assumes that the cement is deposited without grain contacts, and that the initial framework of the uncemented sand rock is a dense

Table 6
Parameters in the petro-elastic model (PEM) and their values used in this work.

Parameter	Notation	Value (unit)
Critical porosity	ϕ_c	36%
Bulk modulus	K_s	35 (GPa)
Shear modulus	μ_s	44 (GPa)
Lithostatic pressure	P_{litho}	25.3 (MPa)
Coordination number	C_p	9
Degree of root	n	3
Bulk modulus of water	K_w	2.7 (GPa)
Bulk modulus of gas	K_g	1.43×10^{-4} (GPa)

random pack of spherical grains with a critical porosity value (denoted by ϕ_c hereafter) around 36%, which is the maximum porosity value of a rock before suspension happens. The dry bulk modulus (K_{HM}) and shear modulus (μ_{HM}) at the critical porosity are then be computed based on the Hertz–Mindlin model [54]:

$$\nu_s = (3K_s - 2\mu_s)/(6K_s + 2\mu_s); \quad \text{(B.1)}$$

$$P_{eff} = P_{litho} - P; \quad \text{(B.2)}$$

$$K_{HM} = \sqrt[n]{\frac{C_p^2(1 - \phi_c)^2 \mu_s^2}{18\pi^2(1 - \nu_s)^2} P_{eff}}; \quad \text{(B.3)}$$

$$\mu_{HM} = \frac{5 - 4\nu_s}{5(2 - \nu_s)} \sqrt[n]{\frac{3C_p^2(1 - \phi_c)^2 \mu_s^2}{2\pi^2(1 - \nu_s)^2} P_{eff}}, \quad \text{(B.4)}$$

with K_s and μ_s standing for the bulk and shear moduli of solid mineral, respectively, and P_{litho} for lithostatic pressure. In Eq. (B.1), ν_s corresponds to Poisson's ratio, whereas in Eq. (B.2), P_{eff} represents the effective pressure, which is the difference between the lithostatic pressure P_{litho} and the pore pressure P (obtained from the numerical simulation of a subsurface model). In Eqs. (B.3) and (B.4), C_p is the coordination number that describes the average number of contacts per sphere, and n is the degree of root. The values of K_s, μ_s, P_{litho}, C_p and n (as well as some other parameters) used in the current work are reported in Table 6.

For a porosity value ϕ smaller than the critical value ϕ_c, the modified lower Hashin–Shtrikman (MLHS) bound [53] is used to compute the effective dry rock bulk (K_d) and shear (μ_d) moduli:

$$Z = (9K_{HM} + 8\mu_{HM})/(K_{HM} + 2\mu_{HM}); \quad \text{(B.5)}$$

$$K_d = \left[\frac{\frac{\phi}{\phi_c}}{K_{HM} + \frac{4}{3}\mu_{HM}} + \frac{\frac{1 - \phi}{\phi_c}}{K_s + \frac{4}{3}\mu_{HM}}\right]^{-1} - \frac{4}{3}\mu_{HM}; \quad \text{(B.6)}$$

$$\mu_d = \left[\frac{\frac{\phi}{\phi_c}}{\mu_{HM} + \frac{\mu_{HM}}{6}Z} + \frac{\frac{1 - \phi}{\phi_c}}{\mu_s + \frac{\mu_{HM}}{6}Z}\right]^{-1} - \frac{\mu_{HM}}{6}Z. \quad \text{(B.7)}$$

Meanwhile, the corresponding saturation effect is described based on the Gassmann model [55], in which the saturated bulk (K_{sat}) and shear (μ_{sat}) moduli are determined by:

$$K_{sat} = K_d + \frac{(1 - \frac{K_d}{K_s})^2}{\frac{\phi}{K_f} + \frac{1 - \phi}{K_s} - \frac{K_d}{K_s^2}}; \quad \text{(B.8)}$$

$$\mu_{sat} = \mu_d, \quad \text{(B.9)}$$

where in Eq. (B.8), K_f is the effective fluid bulk modulus estimated by the Reuss average method [56]. In the current study, since the subsurface model contains only the water and gas phases, K_f is thus given by

$$K_f = \left(\frac{S_w}{K_w} + \frac{S_g}{K_g}\right)^{-1}, \quad \text{(B.10)}$$

(a) Sample inline section (b) Sample cross-line section (c) Sample time section
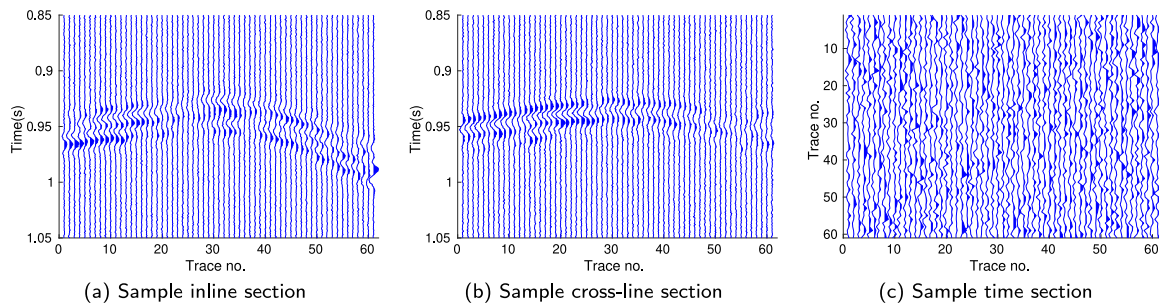
Fig. 9. (a) A section of 3D seismic data along the Y-Z plane with X = 20 (called inline section with the inline number 20), from the mid-offset AVA data of the third survey (day 1455); (b) A section of 3D seismic data along the X-Z plane with Y = 15 (called cross-line section with the cross-line number 15), from the near-offset AVA data of the first survey (day 30); (b) A section of 3D seismic data along the X-Y plane with Z = 10 (at travel time 0.86 s) (called time section with the time number 10), from the near-offset AVA data of the second survey (day 735).

where K_w, K_g correspond to the bulk moduli of water and gas, and S_w and S_g to the saturations of water and gas, respectively. Note that K_w and K_g are considered as parameters of the PEM, whose values are given in Table 6. Meanwhile, S_w and S_g are dynamical properties of the subsurface formations, whose values are obtained by numerical model simulations.

The next step is to compute the saturated density ρ_{sat} of the water-gas mixture, based on the follow formula [53]:

$$\rho_{sat} = (1 - \phi)\rho_m + \phi S_w \rho_w + \phi S_g \rho_g, \tag{B.11}$$

where ρ_m, ρ_w, and ρ_g represent the densities of mineral, water, and gas, respectively. In this work, the values of ρ_m, ρ_w, and ρ_g are determined by the default model in ECLIPSE© 300.

Subsequently, one can proceed to calculate P- and S-wave velocities, denoted by V_P and V_S, respectively, through the following formulae [53]:

$$V_P = \sqrt{\frac{K_{sat} + \frac{4}{3}\mu_{sat}}{\rho_{sat}}}; \tag{B.12}$$

$$V_S = \sqrt{\frac{\mu_{sat}}{\rho_{sat}}}. \tag{B.13}$$

With the velocities V_P and V_S and the density ρ_{sat}, the Zoeppritz equation [53] is then adopted to calculate the reflection coefficients at the interfaces between the layers of a subsurface model. In this case, the reflectivity series is a function of two-way travel time, which is computed based on the P-wave velocity V_P and the vertical thicknesses of the layers in the numerical subsurface model. The computed reflectivity series is then further convolved with a Ricker wavelet (with the dominant frequency of 45 Hz) to obtain the AVA data. In this work, the AVA data are generated with the angles of 10° and 20°, respectively, and are referred to as the near- and mid-offset traces, accordingly.

For illustration, Fig. 9 shows three sample seismic sections from the observations of 4D AVA data used in history matching. As mentioned in Section 3.1, the 4D AVA data consist of vintages at three survey times. At each survey time and for each AVA angle, we sample the values of seismic traces within the travel-time window [0.85, 1.05] (in seconds), with a sampling frequency of 1 millisecond. As such, within the travel-time window, the total number of sample points is 201. Meanwhile, at a given sample time, the number of seismic traces along either the X- or Y-direction is equal to the number of gridblocks of the numerical subsurface model along the same direction. Accordingly, at each survey time, the dimension of the AVA data with either the near- or the mid-offset is $61 \times 61 \times 201$.

References

[1] Ennis-King J, Michael K, Strand J, Sander R, Green C. Underground storage of hydrogen: Mapping out the options for Australia. Future Fuel CRC; 2021.

[2] Hydrogen TCP-Task 42. Underground hydrogen storage: Technology monitor report. Technical report, 2023, URL https://www.ieahydrogen.org/download/17/task-reports/7067/task42_uhs_technologymonitoringreport.pdf. [Access date: 10 January 2024].

[3] Jacobs T. Chevron assumes lead role in largest US hydrogen storage project. J Pet Technol 2023. online 12 September, 2023.

[4] Amirthan T, Perera M. The role of storage systems in hydrogen economy: A review. J Nat Gas Sci Eng 2022;104843.

[5] Cihlar J, Mavins D, Van der Leun K. Picturing the value of underground gas storage to the European hydrogen system. Chicago, IL, USA: Guidehouse; 2021, p. 52.

[6] Minougou JD, Gholami R, Andersen P. Underground hydrogen storage in caverns: Challenges of impure salt structures. Earth-Sci Rev 2023;104599.

[7] Zivar D, Kumar S, Foroozesh J. Underground hydrogen storage: A comprehensive review. Int J Hydrogen Energy 2021;46(45):23436–62.

[8] Hogeweg S, Strobel G, Hagemann B. Benchmark study for the simulation of underground hydrogen storage operations. Comput Geosci 2022;26:1367–78.

[9] Aanonsen S, Nævdal G, Oliver D, Reynolds A, Vallès B. The ensemble Kalman filter in reservoir engineering: A review. SPE J 2009;14:393–412, SPE-117274-PA.

[10] Chen Y, Oliver DS. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math Geosci 2012;44:1–26.

[11] Emerick AA, Reynolds AC. Ensemble smoother with multiple data assimilation. Comput Geosci 2012;55:3–15.

[12] Luo X, Stordal A, Lorentzen R, vdal GN. Iterative ensemble smoother as an approximate solution to a regularized minimum-average-cost problem: Theory and applications. SPE J 2015;20:962–82. http://dx.doi.org/10.2118/176023-PA, SPE-176023-PA.

[13] Gholami R. Hydrogen storage in geological porous media: Solubility, mineral trapping, H$_2$S generation and salt precipitation. J Energy Storage 2023;59:106576.

[14] Chen F, Ma Z, Nasrabadi H, Chen B, Mehana MZS, Van Wijk J. Capacity assessment and cost analysis of geologic storage of hydrogen: A case study in Intermountain-West Region USA. Int J Hydrogen Energy 2023;48(24):9008–22.

[15] Ghaedi M, Andersen PØ, Gholami R. Hydrogen diffusion into caprock: A semi-analytical solution and a hydrogen loss criterion. J Energy Storage 2023;64:107134.

[16] Ghaedi M, Andersen PØ, Gholami R. Maximum column height and optimum storage depth for geological storage of hydrogen. Int J Hydrogen Energy 2023.

[17] Harati S, Gomari SR, Ramegowda M, Pak T. Multi-criteria site selection workflow for geological storage of hydrogen in depleted gas fields: A case for the UK. Int J Hydrogen Energy 2023.

[18] Luboń K, Tarkowski R. Numerical simulation of hydrogen injection and withdrawal to and from a deep aquifer in NW Poland. Int J Hydrogen Energy 2020;45(3):2068–83.

[19] Sáinz-García A, Abarca E, Rubí V, Grandia F. Assessment of feasible strategies for seasonal underground hydrogen storage in a saline aquifer. Int J Hydrogen Energy 2017;42(26):16657–66.

[20] Guises R, Auger E, Bordoloi S, Ofi A, Cranfield C, Freitag H-C. The importance of subsurface characterization and monitoring during development and operation of underground gas storage facilities. In: SPE Middle East oil and gas show and conference. SPE; 2021, D021S010R005.

[21] Krevor S, De Coninck H, Gasda SE, Ghaleigh NS, de Gooyert V, Hajibeygi H, et al. Subsurface carbon dioxide and hydrogen storage for a sustainable energy future. Nat. Rev. Earth Environ. 2023;4(2):102–18.

[22] IEAGHG. Monitoring and modelling of CO$_2$ storage: The potential for improving the cost-benefit ratio of reducing risk, 2020-01. Technical report, 2020, URL https://climit.no/app/uploads/sites/4/2020/05/2020-01-Monitoring-and-Modelling-of-CO2-Storage.pdf. [Access date: 10 January 2024].

[23] Oliver D, Chen Y. Recent progress on reservoir history matching: A review. Comput Geosci 2010;15:185–221.

[24] Luo X. Novel iterative ensemble smoothers derived from a class of generalized cost functions. Comput Geosci 2021;25:1159–89.

[25] Zhang XL, Xiao H, Luo X, He G. Ensemble Kalman method for learning turbulence models from indirect observation data. J Fluid Mech 2022;949:A26.

[26] Lorentzen R, Bhakta T, Grana D, Luo X, Valestrand R, Nævdal G. Simultaneous assimilation of production and seismic data: Application to the Norne field. Comput Geosci 2020;24:907–20. http://dx.doi.org/10.1007/s10596-019-09900-0.

[27] Engl HW, Hanke M, Neubauer A. Regularization of inverse problems. Springer; 2000.

[28] Hamill TM, Whitaker JS, Anderson JL, Snyder C. Comments on "sigma-point Kalman filter data assimilation methods for strongly nonlinear systems". J Atmos Sci 2009;66:3498–500.

[29] Chen Y, Oliver DS. Cross-covariances and localization for EnKF in multiphase flow data assimilation. Comput Geosci 2010;14:579–601.

[30] Emerick A, Reynolds A. Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. Comput Geosci 2011;15:251–69.

[31] Lacerda JM, Emerick AA, Pires AP. Methods to mitigate loss of variance due to sampling errors in ensemble data assimilation with non-local model parameters. J Pet Sci Eng 2019;172:690–706.

[32] Luo X, Bhakta T, Nævdal G. Correlation-based adaptive localization with applications to ensemble-based 4D seismic history matching. SPE J 2018;23:396–427. http://dx.doi.org/10.2118/185936-PA, SPE-185936-PA.

[33] Ranazzi PH, Luo X, Sampaio MA. Improving pseudo-optimal Kalman-gain localization using the random shuffle method. J Pet Sci Eng 2022;215:110589.

[34] Luo X, Lorentzen RJ, Valestrand R, Evensen G. Correlation-based adaptive localization for ensemble-based history matching: Applied to the Norne field case study. SPE Reserv Eval Eng 2019;22:1084–109. http://dx.doi.org/10.2118/191305-PA, SPE-191305-PA.

[35] Luo X, Bhakta T. Automatic and adaptive localization for ensemble-based history matching. J Pet Sci Eng 2020;184:106559.

[36] Chen Y, Oliver DS. Localization and regularization for iterative ensemble smoothers. Comput Geosci 2017;21:13–30.

[37] Soares RV, Luo X, Evensen G, Bhakta T. Handling big models and big data sets in history-matching problems through an adaptive local analysis scheme. SPE J 2021;26(02):973–92.

[38] Jafarpour B, McLaughlin DB. History matching with an ensemble Kalman filter and discrete cosine parameterization. Comput Geosci 2008;12:227–44.

[39] Luo X, Bhakta T, Jakobsen M, vdal GN. An ensemble 4D-seismic history-matching framework with sparse representation based on wavelet multiresolution analysis. SPE J 2017;22:985–1010. http://dx.doi.org/10.2118/180025-PA, SPE-180025-PA.

[40] Soares R, Luo X, Evensen G, Bhakta T. 4D seismic history matching: Assessing the use of a dictionary learning based sparse representation method. J Pet Sci Eng 2020;195:107763.

[41] Luo X, Bhakta T, Jakobsen M, Nævdal G. Efficient big data assimilation through sparse representation: A 3D benchmark case study in petroleum engineering. PLoS One 2018;13:e0198586.

[42] Harati S, Gomari SR, Gasanzade F, Bauer S, Pak T, Orr C. Underground hydrogen storage to balance seasonal variations in energy demand: Impact of well configuration on storage performance in deep saline aquifers. Int J Hydrogen Energy 2023.

[43] Nussbaumer R, Mariethoz G, Gloaguen E, Holliger K. Which path to choose in sequential Gaussian simulation. Math Geosci 2018;50:97–120.

[44] Cao C, Liu H, Hou Z, Mehmood F, Liao J, Feng W. A review of CO2 storage in view of safety and cost-effectiveness. Energies 2020;13(3):600.

[45] Yilmaz Ö. Seismic data analysis: Processing, inversion, and interpretation of seismic data. Society of exploration geophysicists; 2001.

[46] Cruz WC, Luo X, Petvipusit KR. Joint history matching of multiple types of field data in a 3D field-scale case study. Energies 2022;15(17):6372.

[47] Simon D. Optimal state estimation: Kalman, H-Infinity, and nonlinear approaches. Wiley-Interscience; 2006, p. 552.

[48] Gaspari G, Cohn SE. Construction of correlation functions in two and three dimensions. Q J R Meteorol Soc 1999;125:723–57.

[49] Donoho DL, Johnstone JM. Ideal spatial adaptation by wavelet shrinkage. Biometrika 1994;81:425–55.

[50] Donoho DL, Johnstone IM. Adapting to unknown smoothness via wavelet shrinkage. J Amer Statist Assoc 1995;90:1200–24.

[51] Bhakta T, Landrø M. Estimation of pressure-saturation changes for unconsolidated reservoir rocks with high VP/VS ratio. Geophysics 2014;79:M35–54.

[52] Landrø M. Discrimination between pressure and fluid saturation changes from time-lapse seismic data. Geophysics 2001;66:836–44.

[53] Mavko G, Mukerji T, Dvorkin J. The rock physics handbook: Tools for seismic analysis of porous media. Cambridge University Press; 2009.

[54] Mindlin RD. Compliance of elastic bodies in contact. J Appl Mech 1949;16:259–68.

[55] Gassmann F. Über die Elastizität poröser Medien. Vierteljahresschr. Naturforsch Ges 1951;96:1–23.

[56] Reuss A. Berechnung der Fließgrenze von Mischkristallen auf Grund der Plastizitätsbedingung für Einkristalle. ZAMM-J Appl Math Mechnics/Z Angew Math Mech 1929;9:49–58.