Quantitative assessment of monitoring strategies for conformance verification of CO_2 storage projects

Barros, E. G.D.; Leeuwenburgh, O.; Szklarz, S. P.

DOI
10.1016/j.ijggc.2021.103403

Publication date
2021

Document Version
Final published version

Published in
International Journal of Greenhouse Gas Control

International Journal of Greenhouse Gas Control 110 (2021) 103403

Contents lists available at ScienceDirect

International Journal of Greenhouse Gas Control

journal homepage: www.elsevier.com/locate/ijggc

Quantitative assessment of monitoring strategies for conformance verification of CO_2 storage projects

E.G.D. Barros [a,*], O. Leeuwenburgh [a,b], S.P. Szklarz [a]

[a] TNO, Applied Geosciences, Princetonlaan 6, Utrecht, the Netherlands
[b] Delft University of Technology, Geoscience and Engineering, Mekelweg Delft, the Netherlands

ARTICLE INFO

Keywords:
Subsurface CO2 storage
Conformance verification
Risk management
Monitoring design
Uncertainty management
Quantitative workflow

ABSTRACT

We propose a quantitative model-based workflow for conformance verification of CO_2 storage projects. Bayesian inference is applied to update an ensemble of simulation models that capture prior uncertainty based on mismatches with measured data. Conformance assessments are derived by comparison of updated model predictions with storage permit requirements and confidence criteria. Two examples, one conceptual and one based on a real candidate storage site, are provided in which the quantitative workflow is applied to the a priori assessment of candidate monitoring strategies. The examples illustrate the limitations of pressure monitoring in the presence of realistic subsurface uncertainties, and the potential for cost saving by informed design of geophysical monitoring surveys. Approximate methods are discussed that could make the workflow also applicable for (quasi) real-time conformance monitoring.

1. Introduction

The main storage-related challenges for deployment of large-scale Carbon Capture and Storage (CCS) are capacity, confidence and cost. There must be high certainty that the site has the capacity to permanently hold the volumes that are planned to be injected, and that it can be operated safely and economically. These elements are normally addressed in plans that operators submit to the regulating authority before approval to operate a storage site is granted. In Europe, the basis for assessment of these plans is provided by the European Commission (EC) CCS Directive (EC, 2009). It states that a monitoring plan is required to enable comparison between actual and modelled behavior, detection of irregularities, detection of the migration of CO_2, and the assessment of the effectiveness of corrective measures in case of leakages or significant irregularities. It furthermore states that reports need to be submitted, at a frequency determined by the competent authority, that contain all information relevant for assessment of compliance with storage permit conditions, and for 'increasing the knowledge of CO_2 behavior in the storage site'.

The specific requirements for safe storage operations are identified in a storage permit. A general requirement is that injected CO_2 remains within the storage complex permanently. The storage permit will also specify maximum injection rates and pressures, as well as the maximum allowed reservoir pressure. Additional requirements may result from site-specific risks and could therefore differ from one project to another. Key concepts in the monitoring of these risks are captured by the notions of containment and conformance. Containment refers to the basic requirement that CO_2 must remain permanently within the storage complex or within clearly identified boundaries inside that storage complex. According to the EU CCS Directive, conformance (or conformity) refers to consistency of the actual behavior of the injected CO_2 with the modelled behavior. Conformance is a requirement for the transfer of responsibility after site closure, together with the absence of detectable leakage, and a demonstration that the site is evolving towards a situation of long-term stability. A slightly broader definition of conformance, that we will adopt here, also includes compliance with any additional requirements as specified in the storage permit. This definition is in agreement with that used by the government of Alberta, Canada, in its the regulatory framework for CCS, and combines the concepts of concordance (agreement between models and data) and performance (agreement with permit requirements) as discussed by Oldenburg (2018). We will define conformance verification to mean the activity aimed at establishing if a situation of conformance exists at any moment during operation of the site and will remain to exist in the future. Since a storage permit will not be granted if initial modelling suggests that conditions qualifying as non-conformance are likely to

* Corresponding author.
E-mail address: eduardo.barros@tno.nl (E.G.D. Barros).

https://doi.org/10.1016/j.ijggc.2021.103403
Received 22 December 2020; Received in revised form 14 May 2021; Accepted 7 July 2021
Available online 22 July 2021

develop, it may be assumed that initial models prepared by the operator will suggest that the site can be operated in accordance with storage permit requirements.

The actual behavior of the CO_2 after injection in the reservoir will be strongly controlled by conditions (fluid pressure, temperature and composition, stress field) and rock properties (porosity, permeability, connectivity, fault stability) inside the storage reservoir, which are site specific. While these properties may be approximately known at the locations of wells that have been drilled into the reservoir before the start of injection, they will generally be poorly known everywhere else. As a consequence, also the dynamic behavior of the CO_2 will be uncertain. This will also be the case to some extent for storage in depleted gas reservoirs for which information obtained from monitoring data is generally limited to storage volume dimensions and pressure. Some of the uncertainties can be associated with risks of non-conformance situations, such as poor injectivity (preventing injection of the planned volumes), leakage through old wellbores or through the overburden, or migration of CO_2 outside of the licensed storage area (all leading to non-containment), and high seismicity (violating safe operation standards). Some of these risks may be more relevant in some sites than in others.

Monitoring programs should facilitate the identification of irregularities that could point at non-conformance situations and trigger mitigating actions. Information extracted from monitoring data will generally be uncertain due to measurement or interpretation errors and the sparsity or limited resolution of data, requiring the use of (also uncertain) models to fill the gaps (Harp et al., 2019). Conformance verification activities involve the comparison between modelled behavior of the CO_2 and its 'actual' behavior, which must be inferred from monitoring data, and they must therefore take this uncertainty into account.

No clear guidelines or frameworks currently exist for identifying monitoring tasks and setting monitoring performance requirements in a conformance verification context (Bourne et al., 2014). However, experience with monitoring of industrial-scale CO_2 storage operations has been gathered in a number of projects that include Sleipner (e.g. Arts et al., 2008; Furre et al., 2017), Snøvhit (Maldal and Tappel, 2004), In-Salah (Mathieson et al., 2011), Otway (Jenkins et al., 2015) and Quest (Bourne et al., 2014). Decisions about which monitoring technologies to adopt to address identified risks in these projects have generally been based on site-specific monitoring technology feasibility studies that employ qualitative or scenario-based approaches for risk assessment involving collective expert judgement (Bourne et al., 2014). A method often used in support of such approaches is the bowtie method which aims to map and rank threats, consequences, and safeguards. Safeguards include monitoring that can detect irregularities, and a decision logic to interpret the monitoring data and suggest control measures. The possible monitoring technologies are ranked by experts, and an ultimate selection is made by balancing benefits against costs. Applications of such a risk-based framework for developing site-specific measurement, monitoring and verification plans were presented by Bourne et al. (2014) for Quest, by Dean and Tucker (2017) for the Peterhead CCS project (Spence et al., 2014) and by Metcalfe et al. (2017) for the White Rose project. This framework aims to produce a systematic risk assessment and to establish monitoring performance targets that will reduce storage risks to a desired level. As part of this approach extensive modelling based on variation of identified uncertain model aspects was performed to assess the risk of CO_2 migration beyond spill points or storage site boundaries. Active seismic surveys were planned to establish a baseline for repeat surveys and detect lateral movement of the CO_2, and to identify changes in the formations above the storage formation. Revision of the monitoring plans would be necessary in the case of unexpected plume shape or migration velocity, injection pressures, or other deviations from the modelled behavior. Mismatches between dynamic models and monitoring data could then lead to model updates, additional monitoring or corrective measures. It may not always be clear a priori what the character of mismatches would need to be in order to trigger one or more of these actions. The current practice appears to be to conduct a qualitative assessment of such mismatches, leading to an expert-based judgement of required actions.

Quantitative assessment of risks, and the design of monitoring strategies that address them, is often considered complex and computationally demanding since it requires extensive evaluation of mathematical models that capture the full range of possible behavior of the system, identification of the ranges of possible values of all parameters that influence system behavior, and a derivation of risk under uncertainty. However, there are important benefits to such an approach. Perhaps the most important are the possibility to properly account for all uncertainties and their interdependencies, both in models and data (without the need to rank them based on extensive sensitivity experiments), and the possibility to assess the contribution of different monitoring technologies in quantitative terms.

Several examples of quantitative workflows related to risk management in CO_2 storage operations can be found in the literature (see Section 2). This includes applications of optimization approaches to above-zone monitoring well placement aimed at minimizing the time to leak detection (Sun et al., 2013; Cameron, 2013; Yonkofski et al., 2016, 2017; Jeong et al., 2018) and workflows for leakage location estimation and uncertainty reduction based on monitoring data (Sun and Nicot, 2012; Jung et al., 2013; Jung et al., 2015; Hu et al., 2015; Chen et al., 2018; Chen et al., 2020). Some of these methods were also applied to the FutureGen 2.0 candidate storage site (Vermeul et al., 2016; Bacon et al., 2019) illustrating the feasibility of the application of quantitative life-cycle management workflows to realistic settings.

Chadwick and Noy (2015) considered modeling-monitoring convergence as an indicator of the possibility to demonstrate conformance. This is based on their observation that model predictions tend to become more reliable when models are in sufficient agreement with historic data. Following this idea, Harp et al. (2017) proposed an approach for identifying robust pressure management strategies based on decision gap theory. Rather than representing the uncertainty probabilistically, the methodology aims to quantify the degree that one can be incorrect in the characterization of the system and still ensure that performance (conformance) criteria are met. This could be viewed as a metric for comparing alternative monitoring strategies. This approach was adopted by Harp et al. (2019) in a follow-up study that demonstrated an application based on pressure monitoring at a single well where a conformance criterion was defined in terms of the pressure at the well. The workflow uses derivatives of observables with respect to uncertain model parameters which will not be generally available for more complex simulators and measurements or for conformance criteria that are not directly observable. The approach appears to require that a critical value for the conformance criterion is not exceeded in any of the model realizations, which seems quite restrictive. Also, it is not entirely clear how the approach should be extended to applications with multiple conformance criteria and large numbers of uncertain parameters, such as heterogeneous permeability fields.

As an alternative to the workflow of Harp et al. (2019), we develop and demonstrate a generic conformance analysis approach that can be applied to cases with arbitrary numbers and types of uncertainties and arbitrary combinations of monitoring techniques. The approach is inspired by ensemble-based workflows for quantitative evaluation of monitoring strategies for oil field management (e.g. Le and Reynolds, 2013; 2014; He et al., 2020, Barros et al., 2016; Barros, 2018). These studies have suggested that such approaches are indeed capable of providing useful comparisons and rankings of alternative strategies under realistically complex uncertainty scenarios. Here we will therefore propose a quantitative ensemble-based workflow that can be applied to objectively assess the current and future conformance state of the storage system. The workflow can also be used on top of practical risk identification and mitigation frameworks as currently used, to produce objective assessments of the validity of simulation models, the effectiveness of measurement acquisition strategies, and ultimately support decisions about contingency monitoring or corrective measures.

We demonstrate an application of the workflow prior to the start of injection with the aim of evaluating possible monitoring strategies in terms of their capability to correctly assess the conformance of the site at any time in the future.

The remainder of this paper is organized as follows. In the next section we review the literature on quantitative workflows for evaluation of monitoring designs in subsurface applications. In Section 3 we define terminology related to conformance that will be used in this document. In Section 4 we introduce our proposed quantitative workflow for conformance assessment. Section 5 provides a relatively simple example application of the workflow to illustrate the main steps as well as an application to a complex case based on a real potential storage site. In Section 6 we discuss possible extensions and alternative applications of the workflow that could be useful in support of CO_2 storage operations. Finally, the main conclusions are provided in Section 7.

2. Quantitative evaluation of monitoring plans

Early work on quantitative workflows for the assessment of monitoring strategies goes back to at least the mid-1970s, fueled by legislation in the US aimed at protecting groundwater resources from contamination originating from e.g. landfills or agricultural pesticide use. Later applications have included rainfall gauge placement throughout a watershed, river monitoring systems, water distribution networks, and sewer systems. Work in this domain includes studies on pollution source identification (Mahar and Datta, 1997; Datta et al., 2009, 2013), plume characterization (e.g. Kim and Lee, 2007; Balbarini, 2017) and operational plume containment strategies (Tiedeman and Gorelick, 1993). An early application of dynamical simulations for the purpose of optimizing monitoring network design was presented by Meyer and Brill (1988). Loaiciga (1989) presented a workflow consisting of 2 stages: parameter estimation and network optimization, similar to recently proposed Value Of Information workflows for hydrocarbon reservoir management applications (e.g. Barros et al., 2016). The use of ensemble methods for parameter estimation and uncertainty quantification and multi-objective optimization was proposed by Kollat et al. (2011). The multiple objectives included monitoring costs and indicators of monitoring performance, e.g. the number of failures to detect a tracer and the error in characterizing the plume. An ANOVA-based probabilistic collocation-based Kalman Filter was combined with sequential optimal design by Man et al. (2016; 2019).

In the CO_2 storage domain, Sun et al. (2013) used optimization to design a pressure-based monitoring system that minimizes the number of undetected leakage events and the leaked volume. Jeong et al. (2018) extended this optimization framework to more realistic cases, including more elaborate physics models, integrating regulatory constraints and accounting for geological uncertainty. One of the main drawbacks of these studies concerns the way that detection is determined from data without any statistical basis, which is important when models and measurements are not perfect. Cameron (2013) defined a practical workflow for the optimization of sensor placement, using history matching to reduce the geological uncertainty in a storage site, including uncertainty on leak location and leaked fractions. Similar concepts were pursued by e.g. Jung et al., (2013; 2015), He et al. (2018) and Chen et al. (2020) where the latter introduced the use of ensemble methods for reducing the joint uncertainty in model parameters and the respective model responses as represented by up to 100 model realizations, but without explicitly relating the metrics of uncertainty to quantifiable notions of conformance.

In groundwater applications, a number of studies have focused on so-called data-worth (related to Value Of Information) evaluations (or experimental design strategies), aimed at identifying observation strategies that reduce uncertainty (e.g. Tiedeman et al., 2003; Neuman et al., 2012; Leube et al., 2012; Xue et al., 2014). Siade et al. (2017) used D-optimality and minimax objectives to identify robust designs for computationally challenging large-scale applications. Wang et al. (2018) applied an EnKF framework to a real-world case study and considered various information measures for state estimation and parameter estimation respectively. An application of a different metric based on entropy theory to an urban drainage system was presented by Yazdi (2018).

In oil and gas field management applications, Le and Reynolds (2013; 2014) considered the expected reduction in uncertainty as a metric to identify an optimal surveillance operation. Specifically, they proposed to use the concept of mutual information to provide a measure of the strength of the relationship between the observables and a quantity of interest, e.g. oil production. Surveillance operations that deliver a high mutual information are expected to lead to the smallest posterior uncertainty in the quantity of interest. He et al. (2016) suggested the use of proxy models and rejection sampling to avoid the large computational cost that would be associated with performing explicit history matching for many full-complexity reservoir simulation models. In subsequent work He et al. (2018) proposed an ensemble variance analysis method to quantify uncertainty reduction and combined this with a decision-tree analysis of Value Of Information (VOI).

Sato (2011) assessed the VOI in the context of the monitoring of CO_2 storage. His-examples focused on applications considering simple decision scenarios (e.g., “go”, “no-go”, or “gather more information”), simple uncertainties (i.e., one uncertain parameter) and arbitrarily pre-defined likelihood functions, without directly addressing conformance aspects of the problem. Lüth et al. (2015) evaluated the agreement between model predictions and time-lapse seismic data in terms of various measures associated with the geometry of the CO_2 plume developed in a test storage site. However, these studies have not addressed the full quantitative conformance verification problem as defined here, which goes beyond establishing consistency between models and data.

Barros et al. (2016), Barros (2018) developed a fully Bayesian VOI methodology to assess the value of future measurements in closed-loop reservoir management, i.e. under the assumption that the models and production strategies are updated every time that new measurements become available. The workflow considered the value of measurements in terms of the expected impact on the overall economics of the project, taking all uncertainties and nonlinear effects into account. One of the key features of this workflow is the use of an ensemble of model realizations to generate multiple possible outcomes of the monitoring process, which enables quantification of the expected value of a monitoring strategy prior to its implementation in the field. In the workflow presented here we will adopt a similar approach as will be discussed in more detail in Section 4.

3. Conformance assessment

In this section we define the terms and concepts that will be used in the conformance verification workflow. A central concept in CO_2 storage is *containment*, meaning the (permanent) retainment of all CO_2 within safe boundaries after injection. These boundaries could be the bounding faults and storage seals of an entire storage complex, possibly consisting of multiple compartments or layers, or alternatively, internal boundaries, defined by internal faults or spill point contours. We define *conformance* to mean that the actual system behavior satisfies two conditions: (1) consistency with modelled behavior (also referred to as concordance) and (2) compliance with regulations and requirements as defined in the storage permit (performance). Initial modelled (expected) behavior, i.e. before the start of injection, will presumably always suggest the second condition, because otherwise the storage permit would not be granted. However, if the models, or the operational plan that was the basis for predictions made with the model(s), are updated based on data gathered during injection, this may no longer be the case. It is possible that, after such an update, revised model predictions will suggest the possibility that the second condition is violated. In such a situation, consistency between actual and modelled behavior would no

longer be sufficient. A *conformance indicator* is a quantity of interest based on which statements about the behavior of the site can be derived. It refers to a specific physical quantity at a given location and time. It may apply to both properties (e.g., leak, rock strength, fault seal) or the dynamic states (e.g., pore pressure, fluid phase distribution, temperature) of the storage site. For example, Chadwick and Noy (2015) considered plume footprint area, maximum lateral migration distance of CO_2 from the injection point, area of CO_2 accumulation trapped at top reservoir, volume of CO_2 accumulation trapped at top reservoir, area of all CO_2 layers summed, and spreading co-efficient. We note that conformance indicators may be quantities that cannot be observed directly. For instance, due to technical limitations of currently available monitoring technologies, it is not possible to directly measure the properties and state of the subsurface storage site everywhere. Therefore, conformance assessment implies inferring the conformance indicators from indirect information. For example, if the permit requires the CO_2 to remain within specified spatial boundaries, the location of the front of the CO_2 plume could be used as an indicator. An alternative indicator is average reservoir pressure. Neither one of these two quantities can be measured directly but must be inferred from measured data that provide indirect information. Furthermore, we can obviously not directly measure quantities of interest related to the behavior of the storage site in future times. In such cases it must be derived from model-based predictions where we must ensure that the models are at least consistent with available data. *Conformance criteria* are used to convert qualifications such as consistency and compliance into quantitative statements and assessments by classifying deviations from expected behavior as acceptable or not acceptable. In order to delineate the concept of conformance in quantitative terms, conformance criteria should cover three key aspects: (i) the definition of quantifiable conformance indicators, (ii) a range of acceptable values for each indicator and (iii) the confidence thresholds required to classify the state as conformance or non-conformance. Examples of (i, ii) include maximum allowed reservoir pressure and storage complex boundaries that the CO_2 plume may not exceed. These limits will typically arise from safety, integrity or regulatory constraints. The confidence thresholds (iii) concern the probability levels required by the stakeholders to confirm conformance, or trigger mitigating actions in the case of a non-conformance assessment. For critical safety indicators these confidence requirements should ideally be specified in accordance with regulatory directives. For other indicators the choice of confidence thresholds could in principle also be subject to the risk-attitude of the operator and experts involved in the decision-making process for the storage site. Chadwick and Noy (2015) recommend that, given the difficulty in achieving unique or perfect matches between modelled and observed data, regulators should set conformance criteria at realistic levels, focusing on progressive reduction of uncertainty with time and demonstration that the fundamental storage processes are sufficiently well understood. The assessed probability of occurrence of undesired situations might increase gradually over time, in which case various levels of conformance and mitigation actions could be applied, following for example a traffic-light system. Furthermore, comprehensive definitions of conformance for a particular storage site might require the combination of multiple conformance criteria. Even in these more complex cases, however, the quantitative approach proposed in this work is applicable. *Conformance statements*, such as "the storage site is in a state of conformance" must be based on information that is extracted from model simulations and on monitoring data gathered prior to and during operation of the site. Both the models and the data are uncertain, due to modelling and measurement error respectively. Data may be very local (in the case of wells) or have low resolution (in the case of seismic data) and therefore provide an incomplete and indirect picture of the actual behavior of the CO_2. Statements such as the one above should therefore be modified to, for example, "the probability that the storage site is in a state of conformance is 90%". (If conformance regulations are formulated directly in terms of measurements, e.g. bottom-hole pressure data, the relevant information is direct, and the assessment will be much simpler.) In order to arrive at what we will refer to as a *conformance assessment* (for example, 'conformance', or 'non-conformance') from a conformance statement, probability thresholds must be applied. Assessments are the basis for subsequent decisions to take risk-mitigating measures. For example, if the assessment is 'non-conformance', a decision to halt injection may be logical. If the assessment is 'probable conformance', it may be decided to perform contingency monitoring, i.e. gather additional measurements. With *conformance verification* we will mean the activity aimed at establishing if a situation of conformance exists at any moment during operation of the site and will remain to exist in the future. This is the process that will lead to a conformance statement and assessment. If the verification aims to establish the state of conformance at the current time, it will involve an integration of information extracted from past and current monitoring data and the use of a model if some sort of extrapolation is required to relevant quantities that are not directly observed. This latter process is known as history matching (or data assimilation) and aims to estimate the states and parameters of the model that are consistent with the model equations as well as with all available measurements. The future conformance state of the system can subsequently be predicted by simulating the model forward in time.

In the following we will introduce a quantitative workflow that aims to provide a conformance assessment by systematic treatment of uncertainties in both models and measured data, building on the concepts discussed above. The individual elements of the workflow are discussed in further detail in the next section, after which they are combined into a single workflow.

4. Methodology

4.1. Uncertainty quantification

Uncertainty in the expected behavior of CO_2, and in the behavior of the site as a whole, during and after injection can be related to many factors. The main reservoir-related factors are associated with the properties of the reservoir rock such as permeability and porosity, the presence of conducting or sealing fractures, faults, and baffles, the distribution of saturation and temperature of fluids and gases already present in the reservoir, and the physical-chemical interaction of the CO_2 with these fluids and the rock under all conditions encountered during operation and after abandonment of the site. A practical quantitative way of dealing with these types of uncertainties is Monte Carlo simulation. In this approach the distribution of relevant quantities of interest is derived from a large collection of simulation results. A benefit is that this approach can capture co-variation of uncertain parameters as is typical in subsurface models where spatial variability in rock properties are associated with processes on large time and spatial scales. All uncertain inputs to these simulations are sampled from distributions that are informed by prior knowledge and are jointly referred to simply as 'the prior'. The simulation model is used to propagate these uncertainties to the system state at future times. This approach has sometimes been considered too computationally demanding because of the relatively large number of simulations (samples) that is required to provide a complete characterization of the distributions of relevant quantities of interest. Furthermore, a re-assessment of uncertainty should be performed, in principle, whenever new measurement data has been gathered. We will use an efficient ensemble-based method for history matching that enables such a re-assessment in a low-dimension uncertainty subspace spanned by the prior ensemble of realizations of uncertain parameters.

4.2. Ensemble-based data assimilation

A prediction of the future state of conformance can be made by forward simulation of a model and evaluation of the conformance

indicator(s) from the simulated CO_2 behavior. Whenever new informative data are collected, they can be used to update the model used in that prediction in order to increase its reliability, under the assumption that a model that is consistent with past data will produce more accurate predictions. In order to take uncertainty fully into account, we use an ensemble of models, based on many sample realizations of the uncertain model parameters. History matching many models manually, or even with traditional computer-assisted approaches, is too complex and time demanding. A computationally efficient Bayesian method for large-ensemble history matching is the Ensemble Smoother with Multiple Data Assimilations (ES-MDA) proposed by Emerick and Reynolds (2012). We use this method in our experiments.

Consider a set of model parameters collected in a column vector $\mathbf{m} = (m_1, m_2, \cdots, m_n)^T$. We can generate N realizations of these parameters by drawing samples of each individual parameter from an appropriate distribution (the prior), where we will assume that the parameters are distributed according to a joint Gaussian, $N(\overline{\mathbf{m}}, \mathbf{C_m})$, where $\overline{\mathbf{m}}$ is the vector with expected values and $\mathbf{C_m}$ is a covariance matrix. The N sample vectors can be collected in a matrix $\mathbf{M} = [\mathbf{m}^1, \mathbf{m}^2, \ldots, \mathbf{m}^N]$ that defines the prior ensemble of models. Each model can be simulated to predict a set of model outputs $\mathbf{z}_t = \boldsymbol{g}(\mathbf{x}_0, t, \mathbf{m})$ at some time t, based on an initial state $\mathbf{x}_0$ and model operator $\boldsymbol{g}$. We assume that measurements are available of quantities that can be associated with the model output quantities, $\mathbf{y}_t = \boldsymbol{y}(\mathbf{z}_t)$. The measurements themselves are collected in the data vector $\mathbf{d}_t = \mathbf{y}_t^{\text{true}} + \mathbf{e}_t$, where $\mathbf{y}_t^{\text{true}}$ is the vector with true values of the quantities in $\mathbf{y}_t$, and $\mathbf{e}_t$ is a vector with data errors distributed according to $N(0, \mathbf{C_e})$. After simulation of the model ensemble we can construct the ensemble matrix $\mathbf{Y} = [\mathbf{y}^1, \mathbf{y}^2, \cdots, \mathbf{y}^N]_t$. We can also define $\mathbf{E} = [\mathbf{e}^1, \mathbf{e}^2, \ldots, \mathbf{e}^N]_t$ as the ensemble matrix with random data error samples drawn from the data error distribution, and the ensemble matrix $\mathbf{D} = [\mathbf{d}^1, \mathbf{d}^2, \ldots, \mathbf{d}^N]_t$ where $\mathbf{d}^i = \mathbf{d} + \mathbf{e}^i$. We have dropped the time index on the matrices to simplify notation. ES-MDA performs an iterative series of updates of the ensemble matrix $\mathbf{M}$ following Ensemble Kalman Filter theory (Evensen, 2009) to produce an updated sample $\mathbf{M}_{i+1}$

$$\mathbf{M}_{i+1} = \mathbf{M}_i + \mathbf{M}_i \mathbf{Y}_i^{'} \left[\left(\mathbf{Y}_i \mathbf{Y}_i^{'} \right) + \alpha_i \mathbf{C}_\mathbf{e} \right]^{-1} [\mathbf{D} + \mathbf{E}_i - \mathbf{Y}_i]$$

where i indicates the iteration index, and α_i is an inflation factor. Iteration is initiated with $\mathbf{M}_0 = \mathbf{M}$, the prior distribution, and the ensemble obtained at the final iteration will define the 'posterior' distribution. Any informative data will result in a reduction of the uncertainty as characterized by the spread of the posterior. Conditions on the choice of α_i are discussed in Emerick and Reynolds (2012) and Le et al., 2016Le et al. (2016). A typical ensemble size N for large-scale model problems is 100. The full ensemble of models needs to be re-simulated after each update. However, the models can, in principle, all be simulated simultaneously when a high-performance computing system is available. Many successful applications of ES-MDA to history matching of complex reservoirs have been reported in the literature, including history matching to geophysical data (e.g. Emerick and Reynolds, 2012; Leeuwenburgh and Arts, 2014; Zhang and Leeuwenburgh; 2018). A recent application of this methodology to reducing uncertainty in CO_2 storage risk was reported by Chen et al. (2020).

4.3. Quantitative evaluation of conformance monitoring plans

We will now derive an accuracy metric to quantify the expected contribution of future measurements for the purpose of conformance verification that can be used in a workflow for comparison of alternative monitoring plans. Since it considers future candidate measurement gathering strategies, it relies entirely on simulated measurement data. The workflow builds on the previously introduced ensemble approach for uncertainty representation and the Bayesian framework to reduce these uncertainties to arrive at a probabilistic assessment of conformance. Since it considers future candidate measurement gathering strategies, it relies entirely on simulated measurement data.

Prior ensembles of model scenarios $\mathbf{M} = [\mathbf{m}^1, \mathbf{m}^2, \ldots, \mathbf{m}^N]$ and their predicted outputs $\mathbf{Z} = [\mathbf{z}^1, \mathbf{z}^2, \ldots, \mathbf{z}^N]$ are assumed to characterize the uncertainty prior to any monitoring activities. We assume that the state of conformance $C^i = C(\mathbf{z}^i) = \{0, 1\}$ can be derived from the model predictions. With the ensemble of models, we can therefore associate an ensemble of conformance predictions $\mathbf{C} = [C^1, C^2, \ldots, C^N]$. If N_c model predictions indicate conformance, the probability of conformance is estimated to be $p_c = N_c/N$. Given a probability threshold p_α, we can decide to assess the state of the system to be in conformance (i.e. $C = 1$) if $p_c \geq p_\alpha$, and in non-conformance ($C = 0$) if $1 - p_c \geq p_\alpha$.

Let's now assume that the true state of conformance C^{true} is known. We could then compare the predicted conformance state C with the true state and determine if our assessment could be qualified as a true positive ($C = C^{\text{true}} = 0$), false positive ($C = 0,\ C^{\text{true}} = 1$), true negative ($C = C^{\text{true}} = 1$) or false negative ($C = 1,\ C^{\text{true}} = 0$). Since in reality the true state of the storage system is unknown, we could instead consider a set of plausible true models $\mathbf{m}^\tau$ for $\tau = 1, \ldots, N_\tau$, their simulated predictions $\mathbf{z}^\tau$, and their known conformance states $C^\tau = C(\mathbf{z}^\tau)$. For all these plausible truths, we can determine an assessment, and subsequently determine the overall fraction of correct (true) and incorrect (false) assessments. This will be the accuracy metric that we will use in the following. Note that these fundamental metrics could easily be combined into concepts such as precision and recall (or sensitivity), or into a combined F1 score, if preferred.

We now want to associate this accuracy metric with a candidate monitoring plan. For each plausible true model, we can simulate measurement data as $\mathbf{d}^\tau = \mathbf{y}^\tau + \mathbf{e}^\tau$. The prior ensemble of models can be conditioned to this data as described in section 4.2 and revised conformance states at any arbitrary time t can be determined after simulation of the posterior models from their output $\mathbf{Z}_{\text{post}}^\tau$. These steps can be repeated for alternative candidate monitoring plans, enabling a comparison in terms of accuracy from the resulting values of the accuracy metric. Note that since this analysis involves the use of random numbers $\mathbf{e}^\tau$ to represent the random character of the measurement process (see Section 4.2), the analysis should ideally be repeated N_d times to filter out the random effects.

The overall workflow is depicted in Fig. 1.

Similar to the VOI assessment workflow proposed by Barros et al. (2016; Barros, 2018), the workflow for associating a conformance metric with a candidate monitoring strategy is applied to an ensemble $\mathbf{M}_\tau$ of N_τ model realizations where each member has an associated probability. Typically, we use equal probabilities, but we may select the models such that the occurrence of conformance or non-conformance situations across the ensemble reflects our prior confidence p_c^{prior}. Simulation of each member $\mathbf{m}^\tau$ results in predicted observables to which random measurement noise is added. This could be repeated N_d times to arrive at a total of $N_\tau \times N_d$ datasets representing plausible outcomes of the measurement process. A second ensemble $\mathbf{M}$ consisting of N equally probable models can be history matched to each of these datasets individually, resulting each time into an ensemble of updated (posterior) conformance predictions $\mathbf{C}^{\text{post}}$, from which a single posterior probability of conformance p_c^{post}can be derived. Since the conformance state is known for each of the N_τ data-generating models, we can also evaluate if the ultimate conformance assessment would be correct or not. The overall conformance metric for a single monitoring configuration is calculated from all $N_\tau \times N_d$ evaluations. For monitoring configurations generating many data points in time or space, the impact of differences between random data noise realizations may be expected to be negligible relative to the impact of the differences between model realizations and between controllable monitoring design parameters and choosing $N_d = 1$ should therefore be acceptable in most applications.

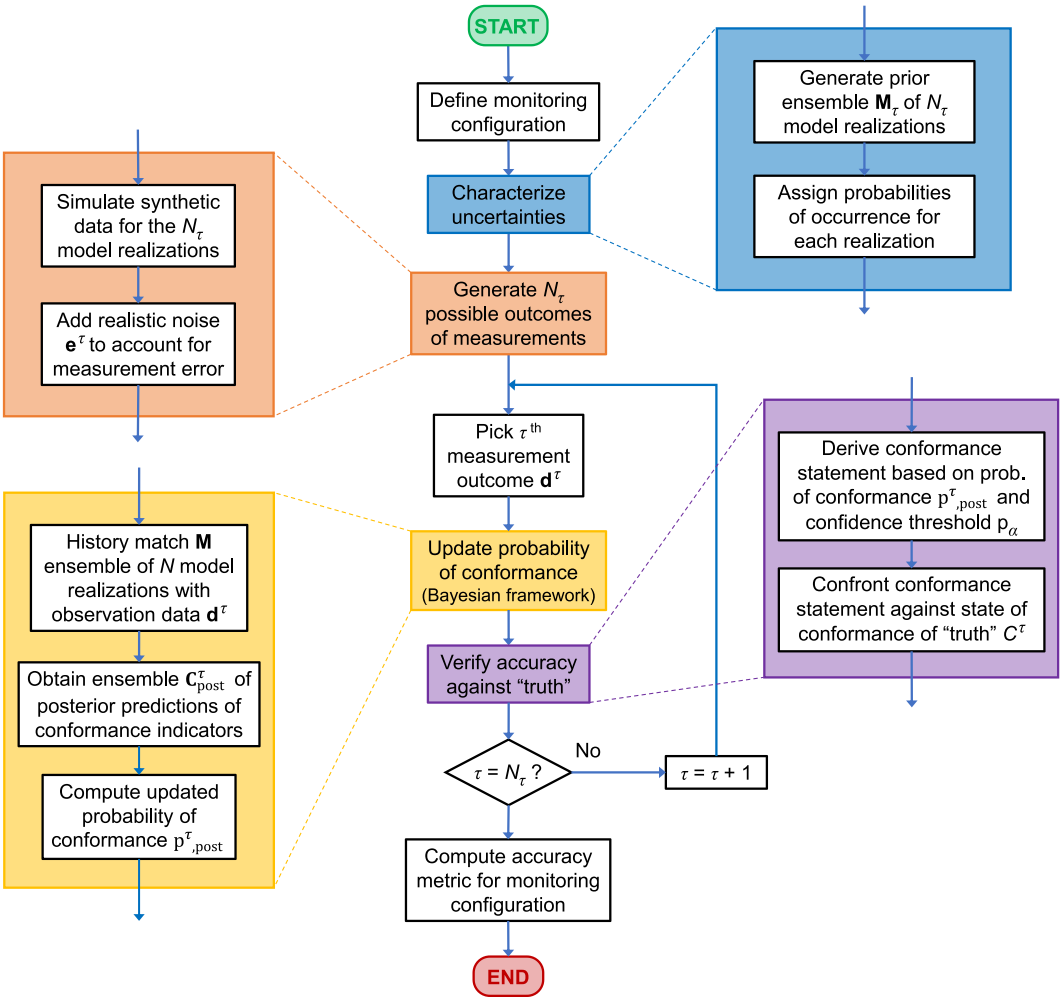

Fig. 1. Proposed workflow for quantifying the expected accuracy to be achieved with a candidate monitoring configuration.

5. Experiments

5.1. Conceptual 2D model example

In a first illustrative example, we consider a 2D model of a storage aquifer with dimensions 1380 × 1380 × 10 m discretized into 69 × 69 × 1 = 4761 active grid blocks. The model has two vertical wells: an injection well operated with a prescribed rate target of 1×10^5 m^3/day (at standard conditions) and maximum allowed injection pressure of 200 bar, and a brine discharge well which is opened 180 days after the start of injection and is operated at a fixed bottom-hole pressure of 80 bar. The reservoir is initially at 83 bar pressure and fully saturated with water. The reservoir simulations are performed with OPM-Flow (The Open Porous Media Initiative, 2021, Rasmussen et al., 2020) under the assumption that CO_2 is immiscible in water. The geological uncertainty in spatial permeability and porosity distributions is characterized by an ensemble of $N_\tau = 100$ model realizations (Fig. 2). The heterogeneities in the permeability and porosity fields have a direct impact on the propagation of the CO_2 plume. The notion of conformance adopted here is associated with regulatory safety bounds for the extent of the CO_2 plume after 1800 days of injection. A (user-defined or regulatory) internal boundary is defined to limit the area where it is safe for the injected CO_2 plume to develop during the storage life cycle (indicated by the yellow box in Fig. 2). The compliance with, or violation of, this boundary

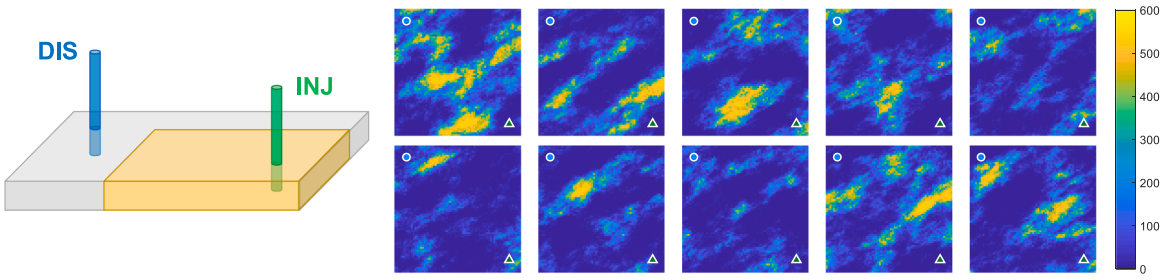

Fig. 2. Synthetic test case study with one CO_2 injection well and one brine discharge well (left). The yellow box indicates the region within the conformance boundary. 10 randomly selected realizations of the permeability field (right); the color bar indicates the permeability values in mD. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

determines conformance or non-conformance, respectively (see Fig. 3). Out of the prior ensemble of 100 realizations, 71 correspond to conformance and 29 to non-conformance. Fig. 4 depicts the simulated CO_2 fronts for the prior ensemble considered at five intermediate times and at the time of interest for conformance verification ($t = 1800$ days). Any determination of conformance based on monitoring at early times (i.e., before the time of interest) will depend not only on the acquired monitoring data but also on the impact of this data on the predictions of the posterior (history matched) models, which will still carry some (reduced) degree of uncertainty.

The ES-MDA history matching method is used to incorporate information from bottom-hole pressure (BHP) measurements and plume tracking measurements such as time-lapse seismic surveys (Leeuwenburgh and Arts, 2014) into the models and improve their predictions. We use ensembles of $N = N_\tau - 1 = 99$ model realizations within the history matching step, where the remaining model is used as the truth model to generate measurement data. Each model acts as the truth once. Following the methodology described in Section 4.3, the model predictions are then used to determine the probability of conformance (non-conformance) by identifying the fraction of model realizations for which the CO_2 plume remains inside (outside) of the conformance boundary. Some of these results have been presented by the authors in their previous conference abstract (Barros et al., 2018b).

5.1.1. BHP monitoring

We quantify the performance (in achieving accurate conformance assessment) of $M = 3$ strategies for BHP monitoring at the injector at 30-day intervals. The monitoring strategies vary in terms of the precision of the measurements. Fig. 5 shows 100 realizations of the generated synthetic BHP data at the injection well for the 3 candidate levels of noise, $\sigma_{BHP} = \{1, 2, 4\}$ bar, obtained by simulating the $N_\tau = N = 100$ models and adding random noise. In this study we consider the availability of BHP measurements until $t = 900$ days to evaluate the value of early pressure signals for conformance verification at a later time. We observe that, while most of the conforming realizations (green lines) are associated with low BHP's at the injector (≤ 110 bar), no clear distinction is observed between conforming and non-conforming realizations among the cases with higher BHPs.

The results obtained with our proposed quantitative conformance assessment workflow applied to these BHP measurements are depicted in Fig. 6. Overall low conformance verification accuracy is achieved by monitoring BHP only, with less than 25% of the conformance assessments being accurate in the case with measurements of highest precision ($\sigma_{BHP} = 1$ bar). We observe that the accuracy is slightly degraded for increasing levels of noise in the data, which is an expected result. We also note that significantly lower accuracy is achieved for the identification of non-conformance (true positives) than for conformance (true negatives).We attribute this to the relatively easy discrimination of conformance and non-conformance situations for the large number of cases with low injection BHP (which are predominantly conforming), while for cases with high injection BHP, including almost all non-conforming cases, discrimination is more difficult. Note that the fraction of true and false conformance assessments displayed in Fig. 6 (right panel) correspond to the sum of true positive / negative and false positive / negative curves in the left plot.

This analysis indicates that, for this particular case, BHP measurements are not expected to provide sufficient evidence to make accurate statements about conformance of the CO_2 plume in the storage aquifer (the fraction of accurate assessments is lower than the fraction of inaccurate assessments). The BHP response at the wells seems to be too weakly (or indirectly) related to the propagation of the injected CO_2 through the porous media, which in this example is determined by the (uncertain) geological heterogeneities. Next, we investigate the feasibility of assessing CO_2 plume conformance (as defined in this case study) based on front-tracking measurements assumed to be derived from time-lapse seismic monitoring surveys delivering interpretations of the saturation field. Such data is expected to provide more direct and relevant evidence for reducing the uncertainty on the footprint of the CO_2 plume than the BHP measurements considered until now.

5.1.2. Impact of acquisition time

We quantify the performance of $M = 5$ monitoring strategies consisting of a single seismic survey each but with varying acquisition time. The candidate survey acquisition times considered are $t_{survey} = \{300, 600, \ldots, 1500\}$ days. Fig. 7 depicts the simulated CO_2 front position at the $M = 5$ candidate survey times for one model out of an ensemble of $N_\tau = 100$ model realizations . The survey observations are assumed to have a measurement precision of 0.5 (grid-block units) in the identification of the front position. To illustrate the quality of the history matching of the front-tracking measurements, the predicted CO_2 fronts for the posterior ensembles obtained with three different truths and $t_{survey} = 600$ days are shown in Fig. 8. We observe an excellent match of the predicted front position at the time of the assimilated survey in all three cases. For two of these cases (truths 4 and 14) the match helps to guide the predictions of the front positions towards the true values at the time of interest (1800 days). However, there are may also be cases (truth 12 is an example) where the predictions at the time of interest may deviate from the true behavior despite a good match at $t_{survey} = 600$ days, leading to false positive or false negatives.

Fig. 9 shows the results in terms of the fraction of false positives and negatives before and after assimilating the surveys from different times. Later surveys are more effective in reducing the number of wrong

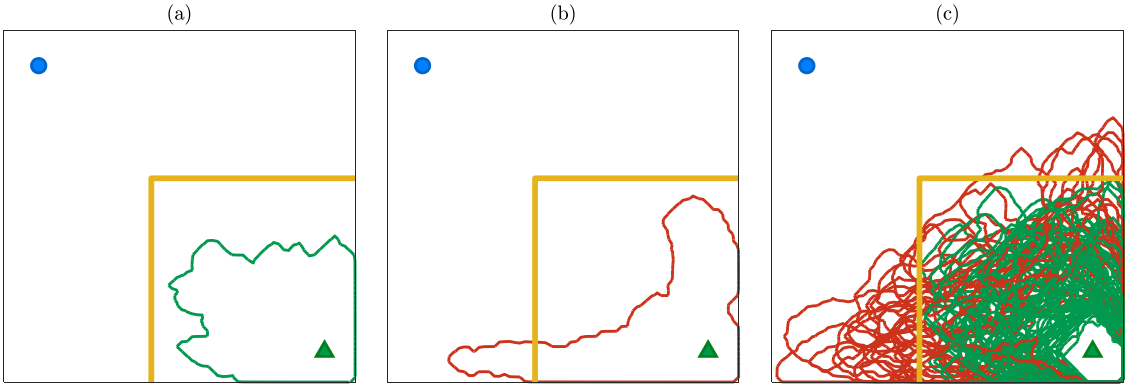

Fig. 3. Representation of CO_2 plume conformance in a case with a user-defined or regulatory internal boundary (yellow line): (a) example of conformant plume, (b) example of non-conformant plume and (c) all plumes colored according to conformance state..; green lines correspond to "negatives" (i.e., indication of conformance) and red lines to "positives" (i.e., indication of non-conformance). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

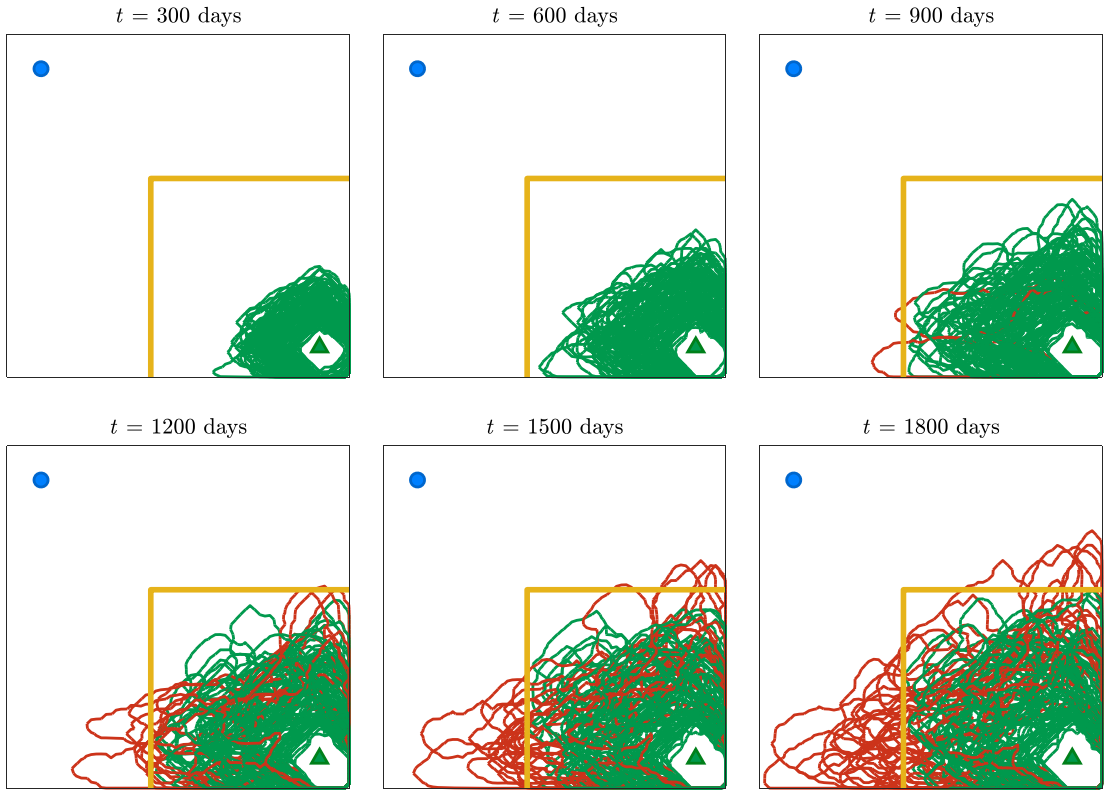

Fig. 4. Development of the CO_2 plume conformance over time based on simulation of the initial ensemble of model realizations (using the same color scheme as in Fig. 3).

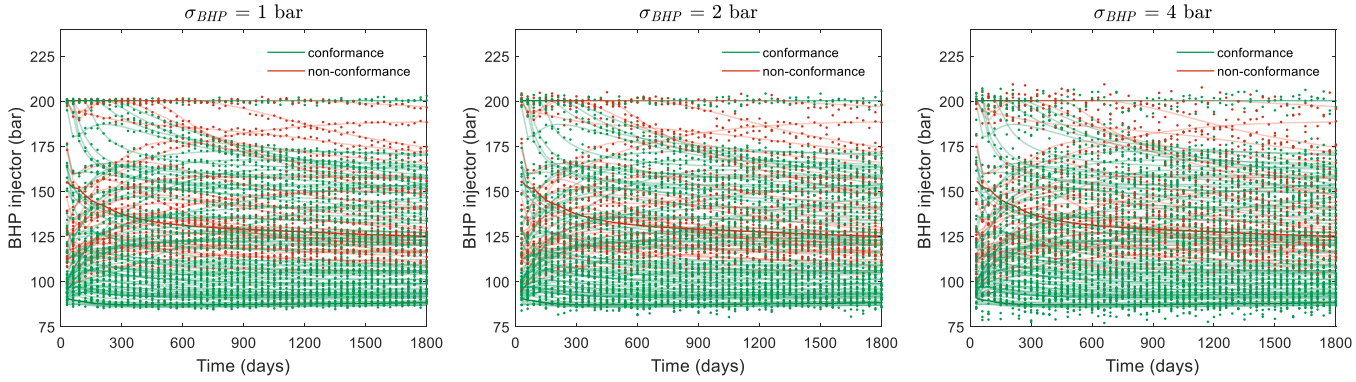

Fig. 5. Simulated BHP profiles and measurements at the CO_2 injection well for the 2D example case. Green lines correspond to simulated BHP profiles for conformance scenarios and red lines correspond to non-conformance scenarios. Green and red dots depict simulated outcomes of the measurement process, obtained by adding noise to the simulated profiles: $\sigma_{BHP} = 1$ bar (left), $\sigma_{BHP} = 2$ bar (middle) and $\sigma_{BHP} = 4$ bar (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

predictions. This is an expected result given the fact that the later the surveys are acquired closer to the time of interest for conformance verification purposes (1800 days). However, from a decision-making point of view, later surveys imply less flexibility available for the operators to react to the new information. Therefore, depending on the type of actions being considered and on the impact of prediction reliability on the decision-making process (e.g. acceptability of, or attitude towards, certain risks), earlier surveys may be more valuable.

5.1.3. *Impact of spatial coverage*

Next, we quantify the performance of $M = 4$ monitoring strategies consisting of a single survey acquired after 1500 days. The survey designs considered have different spatial coverages (Fig. 10), ranging from full reservoir coverage, or sparse two-dimensional lines with different orientations, to sample collection at grid locations.

Fig. 11 compares the four configurations in terms of true and false positives and negatives (like in Fig. 9). The very low number of false negatives indicates that, with the acquisition of the full coverage survey (a), our models will be very reliable when predicting scenarios where conformance takes place. On the other hand, even with the acquisition of a very informative survey indicating conformance at $t_{survey} = 1500$ days, the updated models are not able to rule out the risk of non-conformance at 1800 days entirely. This could be explained by the relatively limited information that is provided by the front position about areas of the reservoir through which the front has not yet passed. The configurations with partial coverage of the reservoir are not able to reduce the conformance prediction errors as much as the survey with full coverage. Despite having the same degree of sparsity, configuration

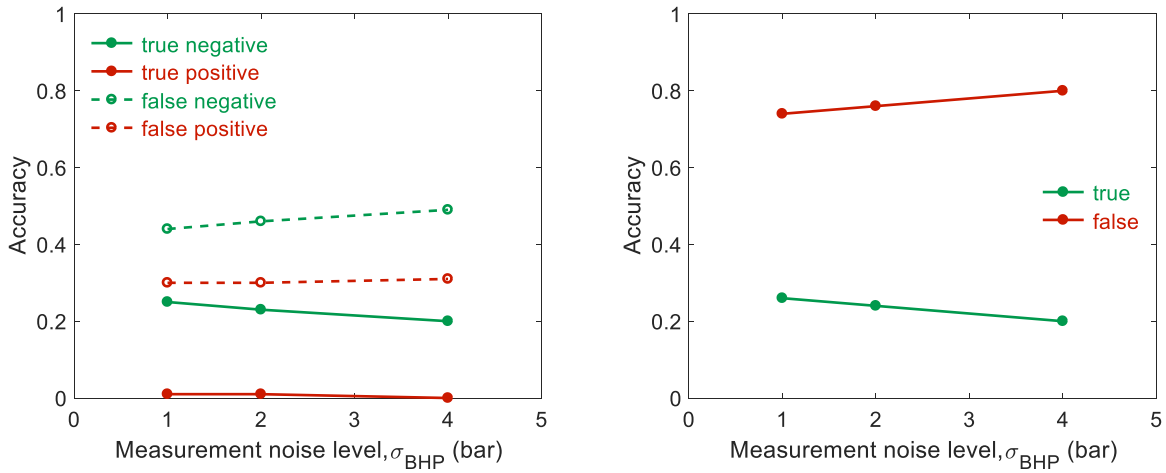

Fig. 6. Accuracy of conformance statements for different levels of noise in BHP measurements: fraction of true and false positives and negatives (left) and true (accurate) and false (inaccurate) assessments (right). In this and other experiments described in this section, $p_\alpha = 1$.

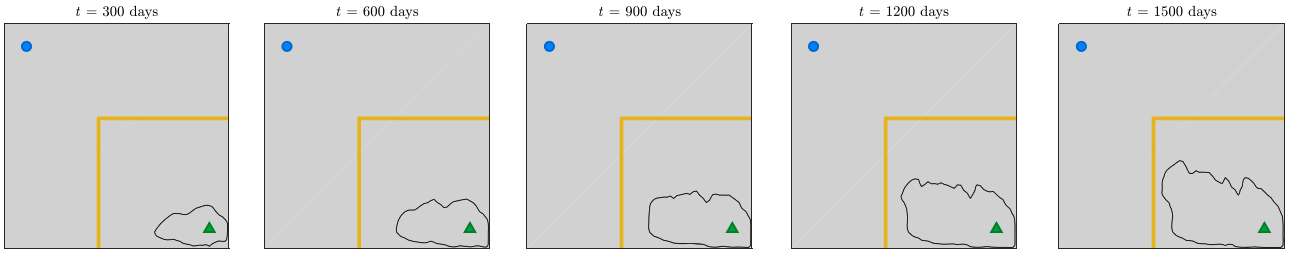

Fig. 7. Five different times of the time-lapse monitor survey. The grey pixels indicate the locations (grid blocks) in the reservoir where the CO_2 plume can be tracked, the black line indicates the plume position for one of the model realizations. The positions of the CO_2 injection and brine production wells are indicated by the green triangle and blue circle, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(c) results in lower errors than configuration (b) in terms of false negatives, indicating a more favorable orientation of the two-dimensional lines. We attribute that to the typical shape of the CO_2 plume in this example, i.e. the ensemble of fronts (Fig. 4) being almost parallel to the sparse lines in configuration (c) and perpendicular to the line connecting the injection and extraction wells, which increases the amount of information available when the front crosses one of the sparse lines. However, the experiments would need to be repeated with larger N_r to entirely exclude the possible influence of chance. The sparsest configuration (d) is not as effective as the others but is still somewhat useful for conformance verification, i.e. better than not acquiring any information and relying on prior knowledge only.

To gain more understanding of the effect of coverage on the achieved accuracy, we consider a new set of $M = 5$ monitoring configurations consisting of a single survey, again acquired at $t_{survey} = 1500$ days, but with varying degrees of sparsity for a fixed orientation of the survey lines (Fig. 12). The results obtained with our quantitative workflow show how fast accuracy in conformance verification is degraded with increasing sparsity of the monitoring survey (Fig. 13).With fewer probed locations in the reservoir it becomes increasingly difficult to obtain evidence of the footprint of the plume to calibrate the ensemble of models, reduce uncertainties and improve accuracy.

We also applied the quantitative approach to assess the usefulness of front-tracking monitoring for plume conformance verification as a function of the reliability of the front positions inferred from time-lapse geophysical. The results obtained confirmed the intuition that increasing front detection errors are expected to lower the accuracy of conformance assessments. However, these results also indicated a more modest decrease in accuracy as a function of the front detection error compared to the previous aspects analyzed (spatial coverage, acquisition time), indicating that even geophysical technologies leading to less than perfect resolution of the front position could still be useful.

5.2. Real field example

The Smeaheia aquifer is a candidate CO_2 storage site in the Norwegian sector of the North Sea (Gassnova, 2016; Statoil, 2016). The aquifer is bounded by two main faults with North-South orientation and includes the Alpha anticlinal structure with capacity for CO_2 storage. The main uncertainty for this site is its connectivity to the neighboring Troll field from which gas is produced. The main risk to be avoided concerns the migration of CO_2 after injection from the Alpha structure to the Beta structure, a second anticline (separated from Alpha through a spill point) where a risk of leakage is associated with uncertain bounding fault properties. This migration could occur if the extent of the CO_2 plume exceeds the spill point limits of the Alpha structure. This will depend in large part on the properties of the CO_2 in the aquifer (e.g., density), which are varying as a function of the pressure in the reservoir. Here we model immiscible CO_2 injection in the gaseous form. In case of high connectivity to the neighboring Troll field, depletion in that field will lead to pressure lowering also in the Smeaheia aquifer, which would cause expansion of the CO_2, thereby facilitating spilling of the CO_2 across the spill point of the Alpha structure.

A high-resolution model of the Smeaheia reservoir was previously constructed for a feasibility study (Gassnova, 2016) and was the basis for an upscaled version that is used here. The model contains 27 layers which are defining 11 vertical zones which we assume to be heterogeneous following varying geostatistical properties (we note that the used geo-statistics are not based on real data or geological information). The model contains three wells, including one injection well located in the Alpha structure. The connection to the neighboring Troll field is modelled through a pressure boundary condition imposed at the

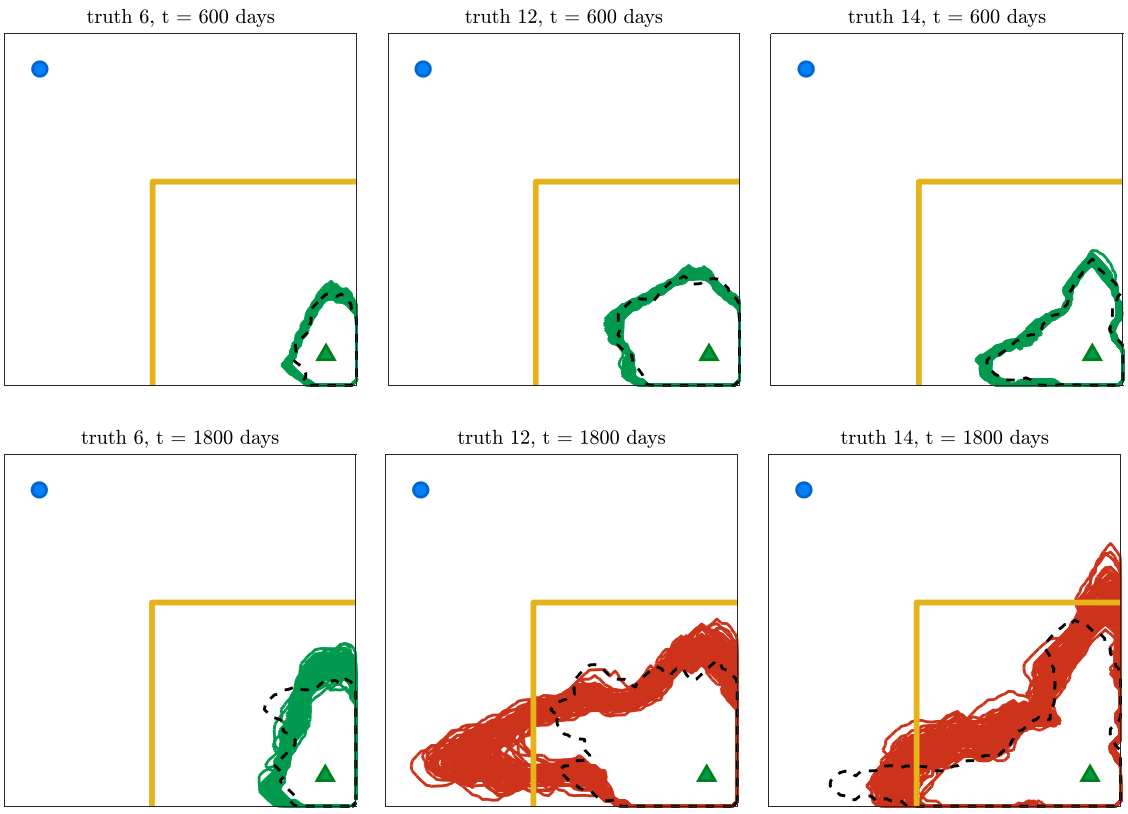

Fig. 8. Posterior predictions as result of history matching of front-tracking measurements at $t = 600$ days for three different truth realizations (ensemble members 6, 12 and 14). Posterior ensemble at $t = 600$ days (= time of assimilated data) (top) and posterior ensemble predictions at the time of interest for conformance verification $t = 1800$ days. The same color scheme (for conformance and non-conformance) as in Fig. 3 is used here, and the dashed black line corresponds to the CO_2 front in the respective truth realizations. These results can be compared to the predictions prior to the history matching displayed in Fig. 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

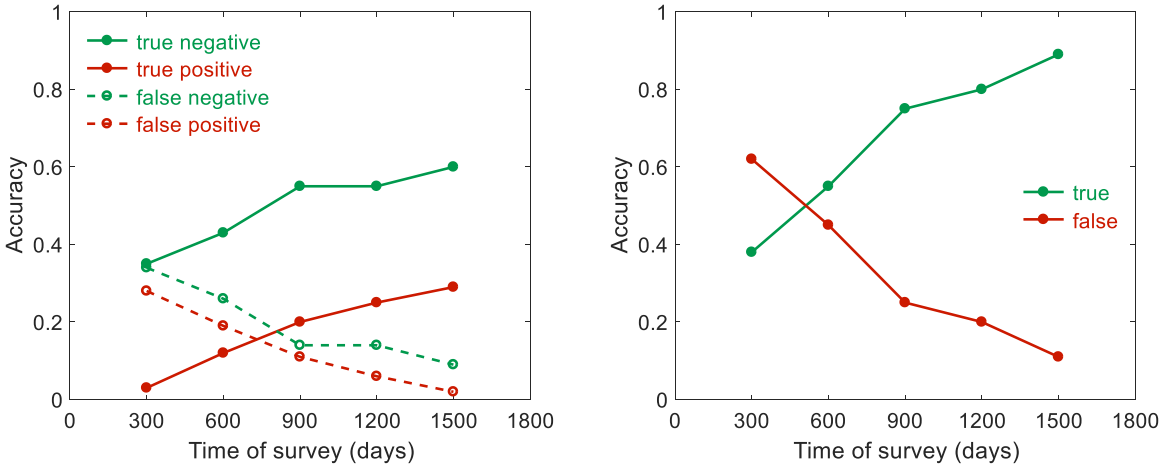

Fig. 9. Accuracy of conformance statements for different survey times: fraction of true and false positives and negatives (left) and true (accurate) and false (inaccurate) assessments (right).

southern boundary of the model by introducing artificial vertical production wells (WP1 and WP2 depicted in Fig. 14) with a prescribed fixed bottom-hole pressure. The simulated period includes the 25-year injection period with 3 Mt of CO_2 injected per year, followed by a period of 5 years during which the injection well is shut in. Changes in the reservoir during this period are associated with redistribution of the injected CO_2, primarily due to depletion of the Troll field. Fig. 14 shows, from a top view, the Smeaheia reservoir model, depicting the porosity (left) and permeability (middle) fields as well as the CO_2 (gas saturation) plume at the end of the simulated period. The simulations were performed with the open-source black-oil simulator OPM-Flow (Rasmussen et al., 2020).

In order to characterize the uncertainty in the connectivity to the depleting Troll field, we use ensembles of $N_\tau = 50$ model realizations created by sampling the pressure boundary condition from a uniform distribution between bounds of 70 and 120 bar. As in the previous example, we consider 50 ensembles of $N = N_\tau - 1 = 49$ model realizations within the history matching step. Fig. 15 shows the extent of the simulated CO_2 plume in the first layer of a sector of the reservoir at

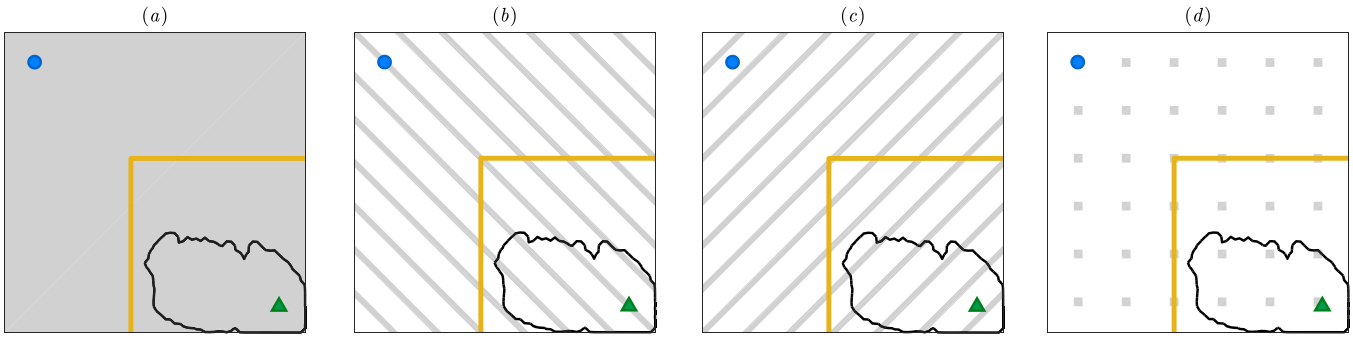

Fig. 10. Four survey configurations with different reservoir coverages or orientations. The grey area (a), lines (b and c) and pixels (d) indicate the locations (grid blocks) in the reservoir where the saturation can be observed. The positions of the CO_2 injection and brine production wells are indicated by the green triangle and blue circle, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

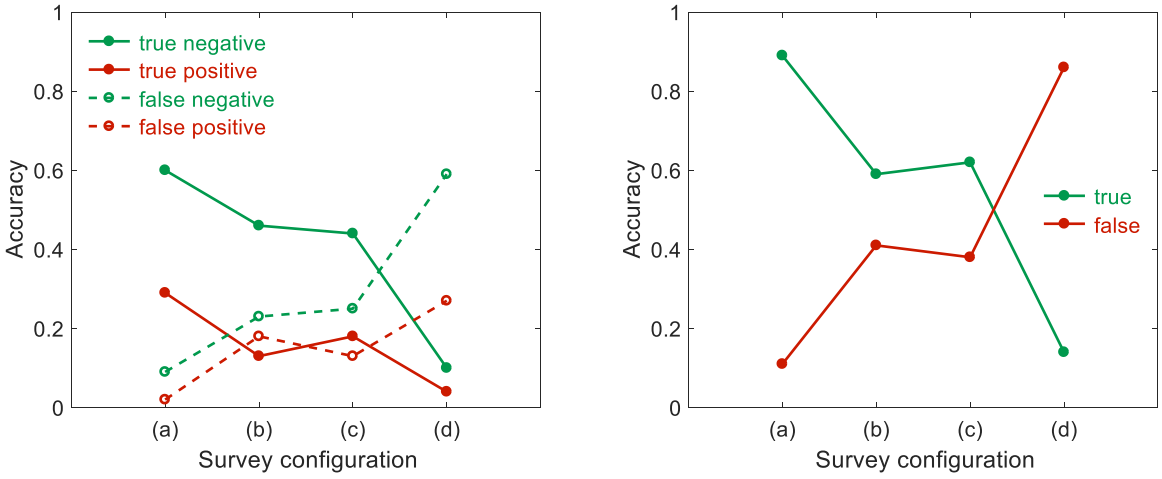

Fig. 11. Accuracy of conformance statements for the four different survey configurations from Fig. 9: fraction of true and false positives and negatives (left) and true (accurate) and false (inaccurate) assessments (right).

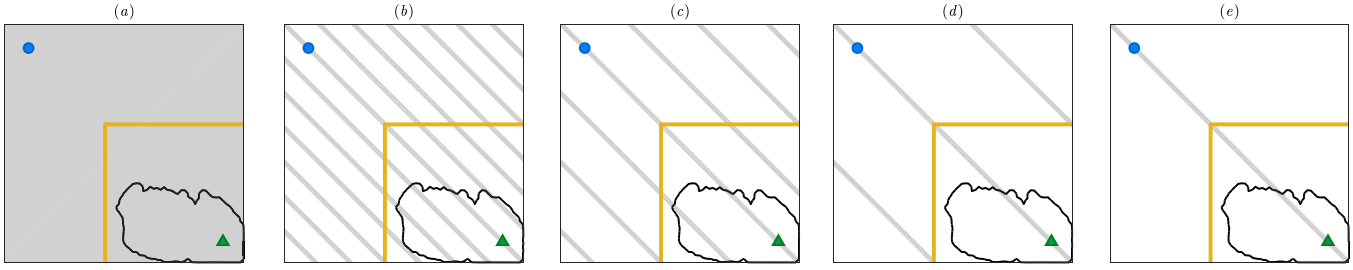

Fig. 12. Five survey configurations with different degrees of sparsity. The grey pixels indicate the locations (grid blocks) in the reservoir where the CO_2 plume can be tracked. The positions of the CO_2 injection and brine production wells are indicated by the green triangle and blue circle, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the end of the simulated period for five of the generated scenarios. In three of these five depicted scenarios (top) CO_2 migrates to the Beta structure area, indicated by the trespassing of the yellow dashed line. In the other scenarios, CO_2 remains within the limits of the Alpha structure. We will refer to these groups of scenarios as non-conformance and conformance scenarios respectively. The degree of connectivity to the Troll field also has an impact on the bottom-hole pressure (BHP) response at the injection well. Fig. 16 depicts the simulated BHP response (lines) and possible outcomes of (noisy) BHP monitoring measurements at the CO_2 injection well (dots) for the 50 generated model scenarios. We observe a segregation between green and red lines, which indicates that non-conformance scenarios are associated with stronger connection to the depleting Troll field and thus lower BHPs at the injector. But we also see that the discrimination between conformance and non-conformance scenarios becomes less obvious in the presence of measurement noise, in particular for scenarios close to the boundary between conformance and non-conformance. Note that the conformance assessment here refers to the state of the reservoir at the end of the simulated period (i.e., 5 years after the end of CO_2 injection) while the BHP measurements are acquired only during the injection period.

5.2.1. *Results: uncertainty in connectivity to depleting field*

We applied our proposed quantitative workflow to quantify the effectiveness of the BHP measurements at the injection well for accurate conformance assessment. In particular, we are interested in comparing

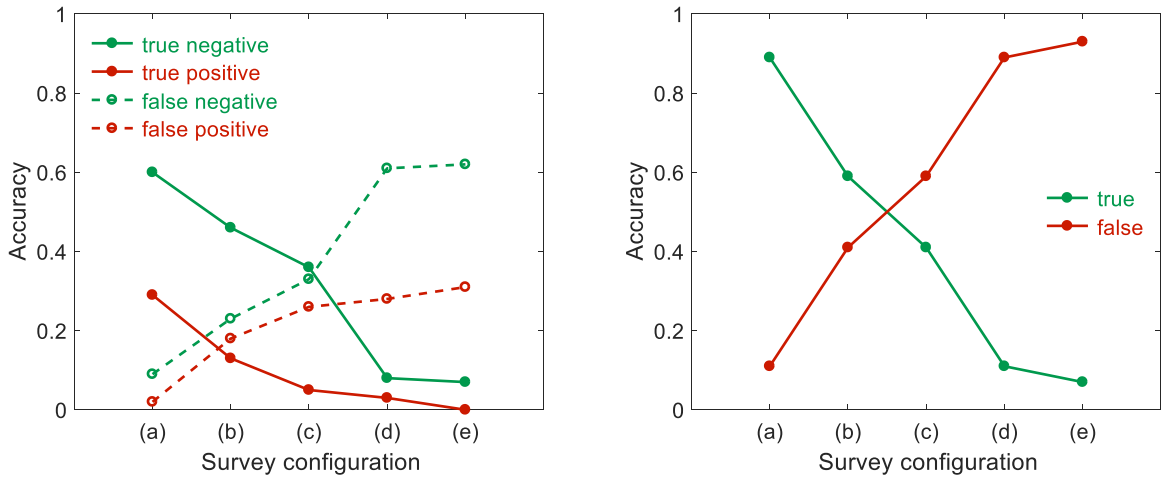

Fig. 13. Accuracy of conformance statements for the survey configurations from Fig. 12 with five different degrees of sparsity for a fixed orientation: fraction of true and false positives and negatives (left) and true (accurate) and false (inaccurate) assessments (right).

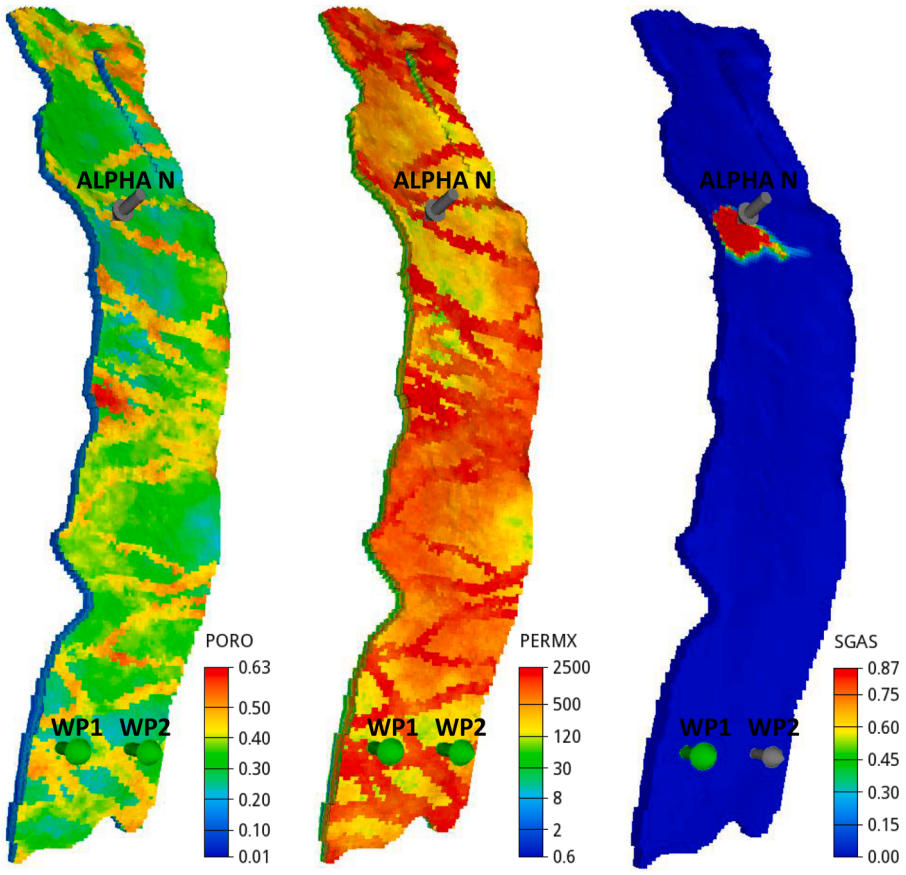

Fig. 14. Top view of the Smeaheia reservoir model: porosity (left) and permeability (middle) fields and CO_2 plume at the end of the simulated period (right). CO_2 is injected in Alpha structure and connection to neighboring field in depletion is modelled by two artificial production wells in the southern portion of the model.

the expected accuracy of the conformance assessments based on varying monitoring durations, in order to determine the minimum duration required to achieve sufficiently reliable discrimination of conformance and non-conformance cases. We consider five different monitoring durations, $T_{BHP} = \{1, 5, 10, 15, 25\}$ years, with BHP measurement errors of $\sigma_{BHP} = 2$ bar (standard deviation) gathered every 6 months. Fig. 17 shows the results of one of the history matches performed as part of the workflow.

The results obtained from the conformance workflow are depicted in Fig. 18 in terms of accuracy as a function of monitoring duration. These results are for a confidence threshold of $p_\alpha = 0.95$, i.e. assuming a decision maker that would only be satisfied with probabilistic conformance statements exceeding 95% certainty. Similar to the results from the previous case study, we observe here again that the fractions of true negatives and positives tend to converge to values close to the proportion of conformance and non-conformance cases in the 50 scenarios considered, indicating that the BHP measurements are able to provide, in time, sufficient evidence for almost perfect conformance

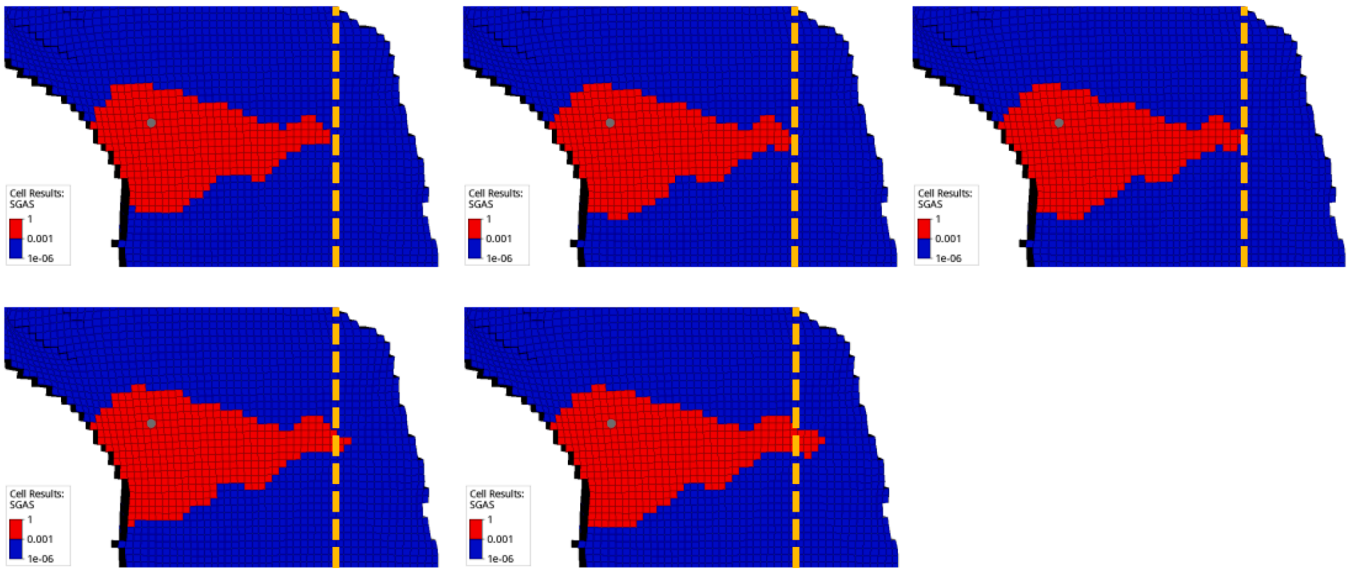

Fig. 15. Simulated CO_2 plume for 5 realizations of the Smeaheia model with varying connectivity to the depleting Troll field. The 3 top scenarios show conformance, defined here as containment of the CO_2 in the Alpha structure. Migration of CO_2 to the Beta structure on the right is indicated by the plume crossing the yellow line in the 2 bottom non-conformance scenarios. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

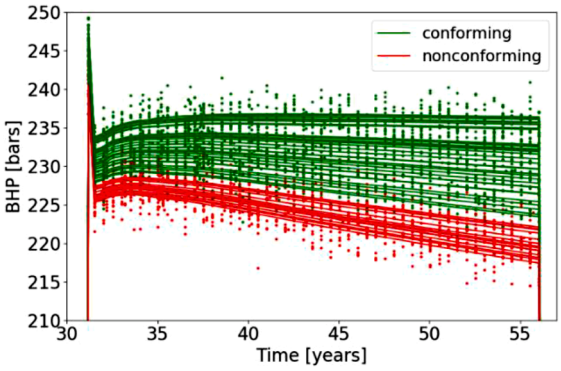

Fig. 16. Simulated BHP profiles and measurements at the CO_2 injection well for the Smeaheia case with varying connectivity to the depleting Troll field. Green lines correspond to simulated BHP's in the conformance scenarios and red lines to the non-conformance scenarios, where the CO_2 plume migrates to the Beta structure. Green and red dots depict simulated outcomes of the measurement process, obtained by adding noise to the simulated profiles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

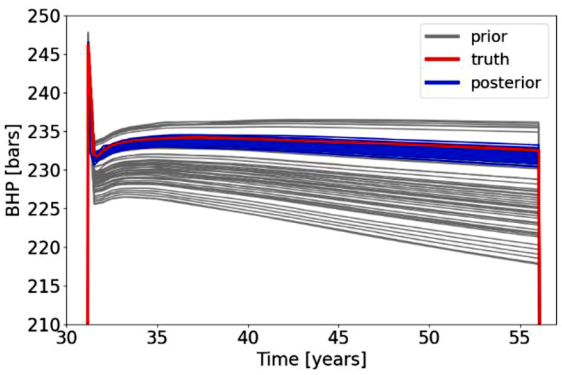

Fig. 17. Bottom-hole pressure predictions at the injection well as result of history matching of BHP measurements during $T_{BHP} = 15$ years for one of the plausible truth realizations. Grey lines were obtained with the prior ensemble, blue lines with the posterior ensemble and the red line with the truth realization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

discrimination. Moreover, we note that there does not seem to be an increase in the expected accuracy after $\Delta T_{BHP} > 15$ years. The results in terms of false positives and negatives also suggest that inaccurate determination of non-conformance is a little more likely to be avoided than inaccurate confirmation of conformance.

In order to gain understanding of the sensitivity of these results to the choice of the confidence threshold p_α, we have repeated the analysis with a fixed monitoring duration of 15 years for varying p_α (Fig. 19). These results illustrate the expected accuracy given the conformance problem and monitoring data for varying preferences (or risk attitude) of the decision maker. We observe that decision makers with less rigorous confidence requirements ($p_\alpha < 0.75$) would correctly determine 100% of the conformance assessments while a more strict confidence threshold ($p_\alpha = 1$) would result in a smaller fraction of correct conformance assessments. This occurs because, in this particular case, posterior probabilities of conformance (or non-conformance, depending on the respective truth model being conforming or non-conforming) are always larger than 75%, while only for a few truth models and realizations of monitoring data posterior probabilities close to 100% are achieved. We note that these results may be case dependent and will depend on the ability to discriminate conformance and non-conformance scenarios given the monitoring data, and possibly on the proportions of these scenarios in the initial ensemble.

5.2.2. *Results: impact of geological uncertainty*

Next, we evaluate the impact of geological uncertainty on the accuracy of conformance assessments based on BHP monitoring. In order to do this analysis, we have varied the permeability values in the top layer of the model. The geological heterogeneities were kept fixed, and we modified the permeability values by applying a multiplier. We consider three cases with increasing uncertainty on this multiplier, i.e. three ensembles of $N = 50$ model realizations were created with $\pm$ 10%, $\pm$ 25% and $\pm$ 50% ranges respectively with respect to the base case described in previous section. Note that the ensembles of 50 models characterize both the uncertainty in permeability of the top layer and the additional uncertainty in the connectivity to the depleting Troll

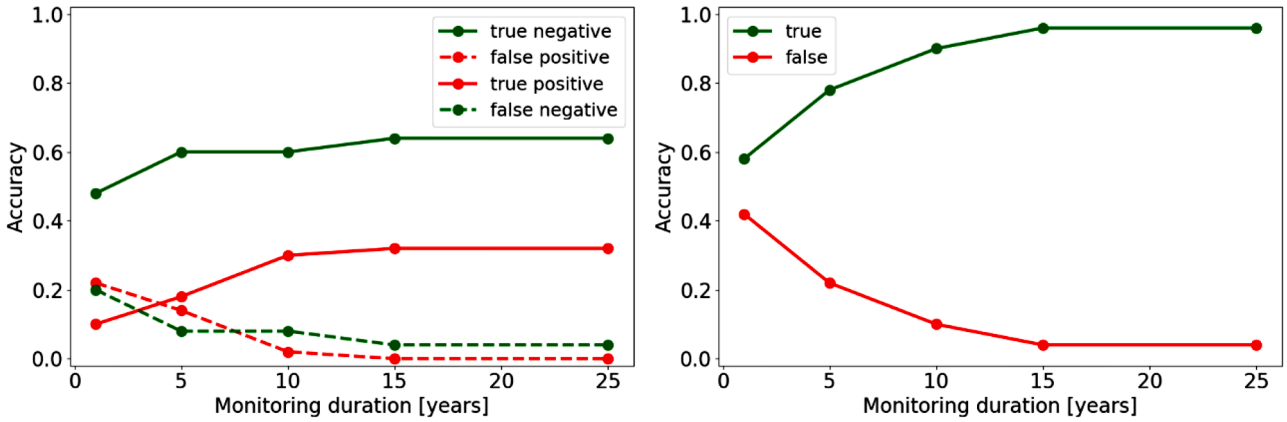

Fig. 18. Accuracy of conformance assessment as a function of the duration of BHP monitoring: fraction of true and false statements for a confidence threshold $p_\alpha = 0.95$ in the presence of uncertainty in connectivity to the neighboring depleting field.

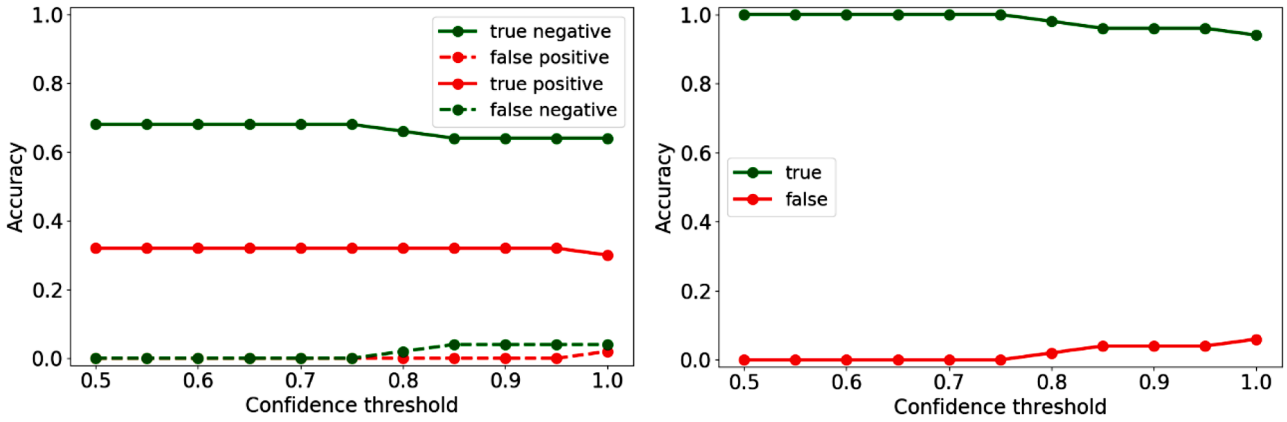

Fig. 19. Accuracy of conformance assessment as a function of the confidence threshold (p_α): fraction of true and false statements for a BHP monitoring duration of 15 years in the presence of uncertainty in connectivity to the neighboring depleting field.

field. Fig. 20 shows the BHP responses of these three ensembles, which can be compared with those of Fig. 16 for the base case ensemble. We observe a much broader overlapping of conforming and non-conforming model scenarios for increasing ranges of uncertainty on permeability, which suggests that conformance discrimination will be more difficult based on BHP measurements only in the presence of geological uncertainty.

We repeated our proposed quantitative conformance assessment workflow for the new ensembles of models using a fixed monitoring duration of 15 years. Fig. 21 depicts the results, showing how accuracy varies as a function of the uncertainty in permeability. Once again, the results are consistent with intuition: increasing uncertainty in permeability leads to an increasing fraction of inaccurate conformance statements due to increasing ambiguity in the BHP monitoring data. These results confirm what was observed in the previous case, where uncertainty on geological heterogeneities was assumed and BHP monitoring performed much worse than front-tracking monitoring. In a separate study, Barros et al. (2020) reached similar conclusions for the case used in this section while utilizing a simplified version of the workflow to assimilate front-tracking measurements. These examples illustrate how the ability to quantify the degradation or improvement of conformance assessment accuracy could provide crucial input for decisions on the deployment and design of additional data gathering activities in cases with large geological uncertainty.

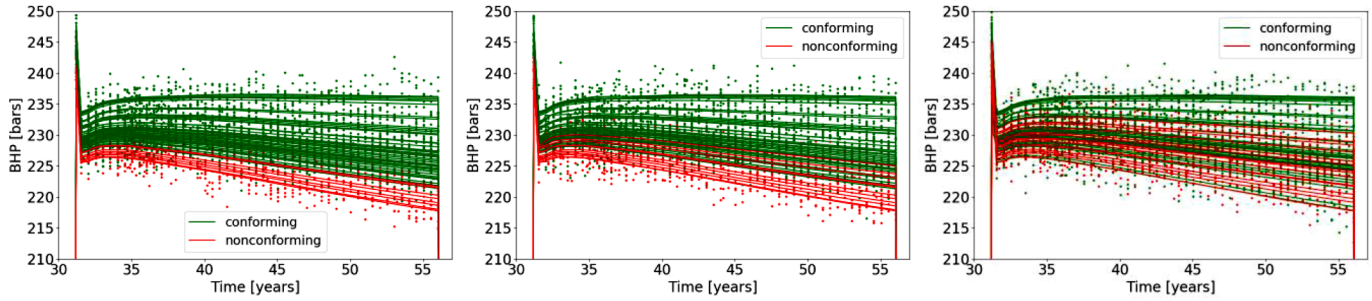

Fig. 20. Simulated BHP profiles and measurements at the CO_2 injection well for the Smeaheia case with varying connectivity to the depleting Troll field and increasing uncertainty on the permeability of the first layer of the model: ± 10% (left), ± 25% (middle) and ± 50% (right).

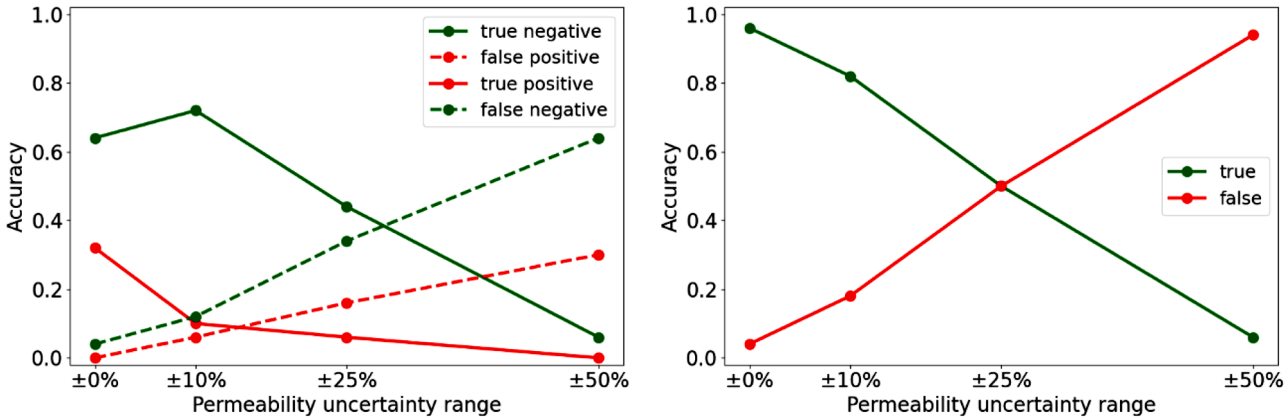

Fig. 21. Accuracy of conformance assessment as a function of the uncertainty range on the permeability of the first layer of the Smeaheia model: fraction of true and false statements for a BHP monitoring duration of 15 years including also uncertainty on connectivity to depleting Troll field.

6. Discussion

The results described in the previous section show how our proposed procedure for quantitative conformance assessment can be used to evaluate and compare different monitoring strategies, providing the key elements to support decisions on the optimal design of monitoring plans for CO_2 storage projects. In addition, this type of analysis can be used to determine possible technical limits of existing monitoring technologies in specific applications, in particular regarding the level of confidence that can be achieved for conformance assessments, which is crucial for communication with regulatory bodies and other stakeholders. A quantitative approach can also be useful to support field development planning (e.g., well placement, drilling sequence), reservoir management actions (e.g., allocation of injection rates and pressures) and monitoring design decisions (i.e., supported by workflows like the one proposed here) to reduce the risk of developing non-conformance situations. This could be achieved by incorporating the described workflow into an optimization framework that would aim to maximize conformance assessment accuracy and the minimization of the probability of non-conformance. This will be the subject of follow-up work.

The workflow as presented here uses Bayesian inference methods to condition full-physics simulation models to measured data. Such an approach can be computationally costly and time consuming because of the need for repeated simulation of a large ensemble of models as part of an iterative assisted history matching workflow. Typically, the number of reservoir simulations required by the proposed workflow is a function of the number of model realizations (N), the number of plausible truths (N_r), the number of iterations of the history matching process (N_{iter}) and the number of monitoring configurations that are being evaluated (M). For instance, in the second case study, we performed $N_{\text{sim}} = N \times N_r \times N_{\text{iter}} \times M = 49 \times 50 \times 4 \times 5 = 49,000$ simulations to evaluate the 5 different monitoring durations. However, there are multiple opportunities for parallelization within the workflow. This more computationally expensive approach is necessary in cases where detailed analysis and understanding of physical processes and the relation of predictions to model uncertainties are required. In a routine monitoring setting, however, several approximations and simplifications could be considered. Barros et al. (2018a); (2020) proposed a simplified approach that associates a dummy variable to each model realization that could be associated with conformance and non-conformance qualifications through a threshold value. Because it does not involve the updating of model parameters, repeated simulation of the model ensemble throughout the iterative history matching procedure is avoided. Such approach may be feasible in cases where prior scenarios can be ranked or logically grouped based on a single conformance indicator or uncertain parameter, which may not always be possible. This approach has similarities with the data-space inversion (DSI) methodology (Sun and Durlofsky, 2017, 2019). Also in DSI, the parameters of the ensemble of models themselves are not updated. Instead, the simulations of the prior ensemble of models are used to estimate statistical relationships between the observations and the predictions of the quantities of interest. These relationships are then leveraged to derive updated predictions directly from the measured data without the need to update model parameters and rerun reservoir simulations.

Instead of conventional statistical approaches to update dynamic predictions (or derived conformance assessments) advanced nonlinear interpolation models could be trained to predict conformance classifiers. Barros and Boullenger (2020) trained convolutional neural networks (CNNs) in a supervised learning approach based on a large set of realizations of simulated monitoring data labelled as conformance or non-conformance and evaluated their conformance classification predictive performance on a test dataset derived from an ensemble of plausible truth scenarios. Similar to the simulation of only the prior ensemble in DSI-like methods, the training simulations, based on uncertainty and monitoring scenarios are performed only once. Another potential approach to speed-up our proposed workflow for monitoring design is to construct predictive reduced order models (ROMs) through machine learning or statistical techniques based on simulations performed upfront (Chen et al., 2018; Xiao et al., 2021).

As discussed previously, the quantitative approach proposed in this paper requires quantifiable criteria for conformance to be defined, in the form of clear conformance indicators, their acceptable limits and the confidence requirements. However, there may be cases where practitioners conceptualize conformance in less absolute terms. For example, conformance is often communicated as the situation where the observed behavior is consistent with (i.e., remains within acceptable deviations from) model predictions. In the absence of a quantitative framework to help determine what is acceptable or not, the task of defining conformance will become more subjective and will make it more difficult to generalize, as the line between acceptable and unacceptable will strongly depend on the context and the interpretation of experts. Still, if these more subjective views of conformance can be translated into a set of multiple indicators and acceptable limits, the approach presented in this work would be applicable.

The underlying assumption of our proposed workflow is that all uncertainty can be characterized by initial ensembles of models which should include both conformance and non-conformance scenarios. (Similarly, risk-based approaches rely on the idea that all possible risks are identified on forehand.) The limitation is an inability to characterize so-called unknown unknowns. However, there are ways, to identify the presence of such unknown uncertainties. For ensemble methods, statistical tools such as rank histograms, reliability diagrams, the ranked probability score, the relative operating characteristic, etc. can be used to assess the consistency between the ensemble predictions and data

(Hamill, 2001; Smith, 2001). Machine learning techniques for anomaly (or outlier) detection (Russo et al., 2019) can also help identify data samples that fall outside the range of "normal" expected behavior. Barros and Boullenger (2021) have used semi-supervised learning approaches based on auto-encoder neural networks to detect non-conformance as outliers out of an open set, i.e. without the need of a priori conceiving a complete set of plausible scenarios of non-conformance. Detected inconsistencies could point to unidentified model shortcomings and should initiate a re-assessment of prior uncertainties. Such an approach would be consistent with the concept of model maturation (Joosten et al., 2014) in which unexpected model updates should lead to the identification of modelling inadequacies through interdisciplinary dialogue between experts.

Finally, the adopted approach allows for extension of the workflow towards evaluating the impact of monitoring strategies in the context of decisions aimed at improving performance of the system or mitigating risks. The evaluation of the contribution of monitoring data is then not merely based on an estimated reduction of uncertainty in some uncertainty system property, but on the actual decision that would be based on a forecast with that reduced uncertainty. While we do not address the impact of possible conformance assessments on subsequent decisions on mitigation measures here, the possibility to do so is a strong motivation to base our workflow on the same framework that was developed previously for VOI assessments (Barros et al., 2016). A first exploration of an application of conformance assessments in such a decision context was presented by Barros et al. (2021).

Summary and conclusions

We have presented a novel model-based framework for quantitative conformance verification. We have provided a formal notion of conformance for use in quantitative approaches along with the required ingredients to define useful quantifiable conformance criteria. The workflow builds on concepts used in state-of-the-art subsurface reservoir management practices, such as ensemble-based uncertainty quantification and history matching. We have demonstrated how the developed methodology can be used to objectively evaluate and compare candidate monitoring configurations in terms of the expected contribution to producing accurate conformance assessments. In particular, our approach enables a-priori (i.e., prior to deployment) evaluation of monitoring strategies in the presence of uncertainties, by simulating and utilizing multiple realizations of plausible outcomes of monitoring activities. Two case studies were presented that illustrate how the results obtained with our proposed methodology can be interpreted to provide insight into the design of monitoring plans.

The first conceptual example showed how the proposed approach can be used to quantify, in terms of conformance verification quality, the contribution of monitoring strategies with various geophysical survey configurations (timing and sparsity) in the presence of geological uncertainties. It was demonstrated that decisions for specific monitoring strategies always require a balance between cost (represented indirectly here by the amount and quality of gathered information) and reliability (accuracy of assessments) and that a quantitative modelling-based approach can facilitate such a decision.

In the second case study, based on the Smeaheia candidate storage site, we analyzed the usefulness of BHP monitoring during the CO_2 injection period to infer the future state of conformance of the CO_2 plume at a time after closure of the storage site. In particular we have investigated the impact of the duration of BHP monitoring on the accuracy of discriminating between conformance and non-conformance cases in the presence of realistically complex uncertainties. The results indicated that, in the presence of uncertainty on the connectivity to the nearby depleting field, BHP measurements were able, in time, to provide sufficient evidence of the behavior of the storage site and enable highly accurate conformance statements. When additional uncertainty in the geology of the top layer of the storage aquifer was considered, we observed that BHP measurements were no longer able to achieve the same results, with steadily degrading accuracy of conformance assessments for increasing degrees of uncertainty in the permeability.

Finally, we have discussed several options that could be useful in facilitating deployment of the workflow in practical and even (quasi) real-time conformance monitoring applications by either reducing the computational cost associated with repeated simulation of the full-complexity reservoir models, or by training regression or classification models a priori. These approaches will be the subject of follow-up research.

CRediT authorship contribution statement

E.G.D. Barros: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Supervision, Visualization, Writing – original draft, Writing – review & editing. **O. Leeuwenburgh:** Conceptualization, Methodology, Formal analysis, Investigation, Supervision, Writing – original draft, Writing – review & editing. **S.P. Szklarz:** Methodology, Software, Validation, Formal analysis, Visualization, Writing – review & editing.

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.

Acknowledgments

This work has been produced with support from the SINTEF-coordinated Pre-ACT project (Project no. 271497) funded by RCN (Norway), Gassnova (Norway), BEIS (UK), RVO (Netherlands), and BMWi (Germany) and co-funded by the European Commission under the Horizon 2020 programme, ACT Grant Agreement no. 691712. We also acknowledge the industry partners for their contributions: Total, Equinor, Shell, TAQA. The Smeaheia model was provided by the partners of the Northern Lights consortium (Equinor, Shell, Total and Gassnova). The results and interpretations discussed in this presentation do not necessarily reflect the views of the data owners.

References

Arts, R., Chadwick, A., Eiken, O., Thibeau, S., Nooner, S., 2008. Ten years' experience of monitoring CO_2 injection in the Utsira Sand at Sleipner, offshore Norway. First Break 26 (1), 91–96.

Bacon, D.H., Yonkofski, C.M.R., Brown, C.F., Demirkanli, D.I., Whiting, J.M., 2019. Risk-based post injection site care and monitoring for commercial-scale carbon storage: reevaluation of the FutureGen 2.0 site using NRAP-Open-IAM and DREAM. Int. J. Greenhouse Gas Control 90, 102784. https://doi.org/10.1016/j.ijggc.2019.102784.

Balbarini, N., 2017. Modelling Tools For Integrating geological, Geophysical and Contamination Data For Characterization of Groundwater plumes. PhD Thesis. Technical University of Denmark.

Barros, E.G.D., Van den Hof, P.M.J., Jansen, J.D., 2016. Value of information in closed-loop reservoir management. Comput. Geosci. 20 (3), 737–749. https://doi.org/10.1007/s10596-015-9509-4.

Barros, E.G.D., 2018. Value of Information in Closed-Loop Reservoir management. PhD Thesis. Delft University of Technology. https://doi.org/10.4233/uuid:9667dc41-c736-47e6-b818-78c7c50fb08d.

Barros, E.G.D., Leeuwenburgh, O., Carpentier, S., Wilschut, F., Neele, F., 2018a. Quantifying the Efficiency of Surveillance Strategies for Subsurface CO_2 Storage Applications. In: Paper presented at the 14th Greenhouse Gas Control Technologies Conference (GHGT-14). Melbourne, Australia. October 21-26. https://ssrn.com/abstract=3365588.

Barros, E.G.D., Leeuwenburgh, O., Carpentier, S.F.A., Wilschut, F., Neele, F.P., 2018b. Quantifying Efficiency Of Field-Wide Geophysical Surveys For Verifying CO_2 Plume Conformance During Storage Operations. Paper presented at the 5th EAGE CO2 Geological Storage Workshop, Utrecht, The Netherlands. https://doi.org/10.3997/2214-4609.201802991. November 21-23.

Barros, E.G.D., Leeuwenburgh, O., Boullenger, B., 2020. Practical Quantitative Monitoring Strategy Assessment For Conformance Verification of CO_2 Storage Projects. Paper presented at the EAGE Annual Online Conference & Exhibition 2020. https://doi.org/10.3997/2214-4609.202011450. Online, December 8-11.

Barros, E.G.D., Boullenger, B., 2020. Quantitative Conformance Assessment in CO_2 Storage Reservoirs Under Geological Uncertainties Using Convolutional Neural Network Classifiers. Paper presented at the 1st EAGE Geoscience and Engineering in Energy Transition Conference (GET 2020). https://doi.org/10.3997/2214-4609.202021062. November 16-18.

Barros, E.G.D., Boullenger, B., 2021. Early Detection of Non-Conformance in Monitoring of CO_2 Storage Reservoirs Using Auto-Encoder Neural Networks. Paper presented at the 3rd EAGE Workshop on Practical Reservoir Monitoring (PRM), Online, March 1-3. https://doi.org/10.3997/2214-4609.202131019.

Barros, E.G.D., Romdhane, A., Bergmo, P.E.S., Leeuwenburgh, O., Grimstad, A.A., 2021. Quantitative Decision Analysis for CO_2 Storage Conformance Management: a Synthetic Case Study at Smeaheia, North Sea. In: Proceedings of the 15th Greenhouse Gas Control Technologies Conference 15-18 March 2021 https://doi.org/10.2139/ssrn.3814679.

Bourne, S., Crouch, S., Smith, M., 2014. A risk-based framework for measurement, monitoring and verification of the Quest CCS Project, Alberta, Canada. Int. J. Greenhouse Gas Control 26, 109–126.

Cameron, D.A., 2013. PhD Thesis. Stanford University.

Chadwick, R.A., Noy, D.J., 2015. Underground CO2 storage: demonstrating regulatory conformance by convergence of history-matched modeled and observed CO2 plume behavior using Sleipner time-lapse seismics. Greenhouse Gases 5 (3), 305–322. https://doi.org/10.1002/ghg.1488.

Chen, B., Harp, D.R., Lin, Y., Keating, E.H., Pawar, R.J., 2018. Geologic CO_2 sequestration monitoring design: A machine learning and uncertainty quantification based approach. Applied Energy 225, 332–345. https://doi.org/10.1016/j.apenergy.2018.05.044.

Chen, B., Harp, D.R., Lu, Z., Pawar, R.J., 2020. Reducing uncertainty in geologic CO_2 sequestration risk assessment by assimilating monitoring data. IJGGC 94, 102926. https://doi.org/10.1016/j.ijggc.2019.102926.

Datta, B., Chakrabarty, D., Dhar, A., 2009. Optimal dynamic monitoring network design and identification of unknown groundwater pollution sources. Water Resour. Manage. 23, 2031–2049. https://doi.org/10.1007/s11269-008-9368-z.

Datta, B., Prakash, O., Campbell, S., Escalada, G., 2013. Efficient Identification of Unknown Groundwater Pollution Sources Using Linked Simulation-Optimization Incorporating Monitoring Location Impact Factor and Frequency Factor. Water Resour. Manage. 27, 4959–4976. https://doi.org/10.1007/s11269-013-0451-8.

Dean, M., Tucker, O., 2017. A risk-based framework for Measurement, Monitoring and Verification (MMV) of the Goldeneye storage complex for the Peterhead CCS project, UK. Int. J. Greenhouse Gas Control 61, 1–15.

Emerick, A.A., Reynolds, A.C., 2012. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. Comput. Geosci. 16 (3), 639–659.

European Parliament and Council of the European Union (EC) (2009). Directive 2009/31/EC of the EC of 23 April 2009 On the Geological Storage of Carbon Dioxide. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32009L0031.

Evensen, G., 2009. Data Assimilation – The Ensemble Kalman Filter, 2nd ed. Springer, Berlin. https://doi.org/10.1007/978-3-540-38301-7.

Furre, A.-K., Eiken, O., Alnes, H., Vevatne, J.N., Kiær, A.F., 2017. 20 years of monitoring CO_2-injection at Sleipner. Energy Procedia 114, 3916–3926. https://doi.org/10.1016/j.egypro.2017.03.1523.

Gassnova, 2016. Feasibility Study For Full-Scale CCS in Norway. Internal report, restricted access.

Hamill, T., 2001. Interpretation of rank histograms for verifying ensemble forecasts. Month. Weather Rev. 129, 550–560.

Harp, D.R., Stauffer, P.H., OMalley, D., Jiao, Z., Egenolf, E.P., Miller, T.A., Martinez, D., Hunter, K.A., Middleton, R.S., Bielicki, J.M., Pawar, R., 2017. Development of robust pressure management strategies for geologic CO_2 sequestration. Int. J. Greenhouse Gas Control 64, 43–59. https://doi.org/10.1016/j.ijggc.2017.06.012.

Harp, D.R., Oldenburg, C.M., Pawar, R., 2019. A metric for evaluating conformance robustness during geologic CO_2 sequestration operations. Int. J. Greenhouse Gas Control 85, 100–108. https://doi.org/10.1016/j.ijggc.2019.03.023.

He, J., Sarma, P., Bhark, E., Tanaka, S., Chen, B., Wen, X.H., Kamath, J., 2018. Quantifying expected uncertainty reduction and value of information using ensemble-variance analysis. SPE J. 23 (2), 428–448.

Hu, L., Bayer, P., Brauchler, R., 2015. Detection of carbon dioxide leakage during injection in deep saline formations by pressure tomography. Water Resources Research 52 (7), 5676–5686. https://doi.org/10.1002/2015WR018420.

Jenkins, C., Chadwick, A., Hovorka, S.D., 2015. The state of the art in monitoring and verification - ten years on. Int. J. Greenhouse Gas Control 40, 312–349.

Jeong, H., Sun, A.Y., Zhang, X., 2018. Cost-optimal design of pressure-based monitoring networks for carbon sequestration projects, with consideration of geological uncertainty. Int. J. Greenhouse Gas Control 71, 278–292.

Joosten, G.J.P., Altintas, A., van Essen, G., Van Doren, J., Gelderblom, P., van den Hoek, P., Foreste, K., 2014. Reservoir model maturation and assisted history matching based on production and 4D seismic data. In: Paper SPE-170604-MS presented at the SPE Annual Technical Conference and Exhibition. Amsterdam. https://doi.org/10.2118/170604-MS, 27-29 October.

Jung, Y., Zhou, Q., Birkholzer, J.T., 2013. Early detection of brine and CO_2 leakage through abandoned wells using pressure and surface-deformation monitoring data: Concept and demonstration. Advances in Water Resources 62, 555–569. https://doi.org/10.1016/j.advwatres.2013.06.008.

Jung, Y., Zhou, Q., Birkholzer, J.T., 2013. On the detection of leakage pathways in geological CO_2 storage systems using pressure monitoring data: Impact of model parameter uncertainties. Advances in Water Resources 84, 112–124. https://doi.org/10.1016/j.advwatres.2015.08.005.

Kim, K.-.H., Lee, K.-.K., 2007. Optimization of groundwater-monitoring networks for identification of the distribution of a contaminant plume. Stoch. Environ. Res. Risk Assess. 21 (6), 785–794. https://doi.org/10.1007/s00477-006-0094-x.

Kollat, J.B., Reed, P.M., Maxwell, R.M., 2011. Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics. Water Resour. Res. 47 (2) https://doi.org/10.1029/2010WR009194.

Le, D.H., Reynolds, A.C., 2014. Optimal choice of a surveillance operation using information theory. Comput. Geosci. 18 (3–4), 505–518.

Le, D.H., Reynolds, A.C., 2014. Estimation of Mutual Information and Conditional Entropy for Surveillance Optimization. SPE Journal 19 (4), 648–661. https://doi.org/10.2118/163638-PA.

Le, D.H., Emerick, A.A., Reynolds, A.C., 2016. An Adaptive Ensemble Smoother With Multiple Data Assimilation for Assisted History Matching. SPE Journal 21 (6), 2195–2207. https://doi.org/10.2118/173214-PA.

Leeuwenburgh, O., Arts, R., 2014. Distance parameterization for efficient seismic history matching with the Ensemble Kalman Filter. Comput. Geosci. 18 (3–4), 535–548. https://doi.org/10.1007/s10596-014-9434-y.

Leube, P.C., Geiges, A., Nowak, W., 2012. Bayesian assessment of the expected data impact on prediction confidence in optimal sampling design. Water Resour. Res. 48, W02501. https://doi.org/10.1029/2010WR010137.

Loaiciga, H.A., 1989. An optimization approach for groundwater quality monitoring network design. Water Resour. Res. 25 (8), 1771–1782. https://doi.org/10.1029/WR025i008p01771.

Lüth, S., Ivanova, A., Kempka, T., 2015. Conformity assessment of monitoring and simulation of CO_2 storage: a case study from the Ketzin pilot site. Int. J. Greenhouse Gas Control 42, 329–339.

Mahar, P.S., Datta, B., 1997. Optimal monitoring network and groundwater pollution source identification. J. Water Resour. Plann. Manage. 123 (4), 199–207.

Maldal, T., Tappel, I.M., 2004. CO_2 underground storage for Snøhvit gas field development. Energy 29 (9), 1403–1411.

Man, J., Zhang, J., Li, W., Zeng, L., Wu, L., 2016. Sequential ensemble-based optimal design for parameter estimation. Water Resour. Res. 52 (10), 7577–7592. https://doi.org/10.1002/2016WR018736.

Man, J., Zheng, Q., Wu, L., Zeng, L., 2019. Improving parameter estimation with an efficient sequential probabilistic collocation-based optimal design method. J. Hydrol. (Amst) 569, 1–11. https://doi.org/10.1016/j.jhydrol.2018.11.056.

Metcalfe, R., Thatcher, K., Towler, G., Paulley, A., Eng, J., 2017. Sub-surface risk assessment for the Endurance CO_2 store of the White Rose project. UK. Energy Procedia 114, 4313–4320. https://doi.org/10.1016/j.egypro.2017.03.1578.

Mathieson, A., Midgely, J., Wright, I., Saoula, N., Ringrose, P., 2011. Salah CO_2 Storage JIP: CO_2 sequestration monitoring and verification technologies applied at Krechba, Algeria. Energy Procedia 4, 3596–3603. https://doi.org/10.1016/j.egypro.2011.02.289.

Meyer, P.D., Brill Jr., E.D., 1988. A method for locating wells in a groundwater monitoring network under conditions of uncertainty. Water Resour. Res 24 (8), 1277–1282. https://doi.org/10.1029/WR024i008p01277.

Neuman, S.P., Xue, L., Ye, M., Lu, D., 2012. Bayesian analysis of data-worth considering model and parameter uncertainties. Adv. Water Resour. 36, 75–85. https://doi.org/10.1016/j.advwatres.2011.02.007.

Oldenburg, C.M., 2018. Are we all in concordance with the meaning of the word conformance, and is our definition in conformity with standard definitions? Greenhouse Gas. Sci. Technol. 8, 210–214. https://doi.org/10.1002/ghg.1773.

The Open Porous Media Initiative, https://opm-project.org/. 2021.

Rasmussen, A.F., Sandve, T.H., Bao, K., Lauser, A., Hove, J., Skaflestad, B., Klöfkorn, R., Blatt, M., Rustad, A.B., Sævareid, O., Lie, K.-A., Thune, A., 2020. The open porous media flow reservoir simulator. Comput. Math. Appl. https://doi.org/10.1016/j.camwa.2020.05.014.

Russo, S., Disch, A., Blumensaat, F., Villez, K., 2019. Anomaly detection using deep autoencoders for in-situ wastewater systems monitoring data. In: Proceedings of the 10th IWA Symposium on Systems Analysis and Integrated Assessment (Watermatex 2019). Copenhagen, Denmark. September 1-4.

Sato, K., 2011. Value of information analysis for adequate monitoring of carbon dioxide storage in geological reservoirs under uncertainty. Int. J. Greenhouse Gas Control 5, 1294–1302.

Siade, A.J., Hall, J., Karelse, R.N., 2017. A practical, robust methodology for acquiring new observation data using computationally expensive groundwater models. Water Resour. Res. 53 (11), 9860–9882. https://doi.org/10.1002/2017WR020814.

Smith, L.A., 2001. Disentangling uncertainty and error: on the predictability of nonlinear systems. In: Alistair, E.Mees (Ed.), Nonlinear Dynamics and Statistics. Birkhauer Press, pp. 31–64. Ed.

Spence, B., Horan, D., Tucker, O., 2014. The peterhead-goldeneye gas post-combustion CCS project. Energy Procedia 63, 6258–6266. https://doi.org/10.1016/j.egypro.2014.11.657.

Statoil, 2016. Selected Extracts from Statoil internal Report On Subsurface Evaluation of Smeaheia as Part of 2016 Feasibility study On CO_2 Storage in the Norwegian Continental Shelf. Internal report, restricted access.

Sun, A.Y., Nicot, J-P., 2012. Inversion of pressure anomaly data for detecting leakage at geologic carbon sequestration sites. Advances in Water Resources 44, 20–29. https://doi.org/10.1016/j.advwatres.2012.04.006.

Sun, A.Y., Nicot, J.P., Zhang, X., 2013. Optimal design of pressure-based, leakage detection monitoring networks for geologic carbon sequestration repositories. Int. J. Greenhouse Gas Control 19, 251–261.

Sun, W., Durlofsky, L.J., 2017. A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems. Math. Geosci. 49 (6), 679–715.

Sun, W., Durlofsky, L.J., 2019. Data-space approaches for uncertainty quantification of CO_2 plume location in geological carbon storage. Adv. Water Resour. 123, 234–255. https://doi.org/10.1016/j.advwatres.2018.10.028.

Tiedeman, C., Gorelick, S.M., 1993. Analysis of uncertainty in optimal groundwater contaminant capture design. Water Resour. Res. 29 (7), 2139–2153. https://doi.org/10.1029/93WR00546.

Tiedeman, C.R., Hill, M.C., D'Agnese, F.A., Faunt, C.C., 2003. Methods for using groundwater model predictions to guide hydrogeologic data collection, with application to the Death Valley regional ground-water flow system. Water Resour. Res. 39 (1), 1010. https://doi.org/10.1029/2001WR001255.

Vermeul, V.R., Amonette, J.E., Strickland, C.E., Williams, M.D., Bonneville, A., 2016. An overview of the monitoring program design for the FutureGen 2.0 CO2 storage site. IJGGC 51, 193–206. https://doi.org/10.1016/j.ijggc.2016.05.023.

Wang, Y., Shi, L., Zha, Y., Li, X., Zhang, Q., Ye, M., 2018. Sequential data-worth analysis coupled with ensemble Kalman filter for soil water flow: a real-world case study. J. Hydrol. (Amst) 564, 76–88. https://doi.org/10.1016/j.jhydrol.2018.06.059.

Xiao, C., Leeuwenburgh, O., Lin, H.-X., Heemink, A., 2021. Conditioning of deep-learning surrogate models to image data with application to reservoir characterization. Knowl. Based Syst. 220, 106956 https://doi.org/10.1016/j.knosys.2021.106956.

Xue, L., Zhang, D., Guadagnini, A., Neuman, S.P., 2014. Multimodel Bayesian analysis of groundwater data worth. Water Resour. Res. 50 (11), 8481–8496. https://doi.org/10.1002/2014WR015503.

Yazdi, J., 2018. Water quality monitoring network design for urban drainage systems, an entropy method. Urban Water J. 15 (3), 227–233. https://doi.org/10.1080/1573062X.2018.1424215.

Yonkofski, C.M.R., Gastelum, J.A., Porter, E.A., Rodriguez, L.R., Bacon, D.H., Brown, C. F., 2016. An optimization approach to design monitoring schemes for CO_2 leakage detection. Int. J. Greenhouse Gas Control 47, 233–239. https://doi.org/10.1016/j.ijggc.2016.01.040.

Yonkofski, C.M.R., Davidson, C.L., Rodriguez, L.R., Porter, E.A., Bender, S.R., Brown, C. F., 2017. Optimized, Budget-constrained Monitoring Well Placement Using DREAM. Energy Procedia 114, 3649–3655. https://doi.org/10.1016/j.egypro.2017.03.1496.

Zhang, Y., Leeuwenburgh, O., 2017. Image-oriented distance parameterization for ensemble-based seismic history matching. Comput Geosci 21, 713–731. https://doi.org/10.1007/s10596-017-9652-1.