Citation (APA)
Serrao Seabra, G. S. S. (2026). *Improving uncertainty quantification of geological carbon storage with data assimilation*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:8182a209-ba87-426c-a326-de3da83e7dcd

IMPROVING UNCERTAINTY QUANTIFICATION OF GEOLOGICAL CARBON STORAGE WITH DATA ASSIMILATION

Gabriel Serrao Seabra

IMPROVING UNCERTAINTY QUANTIFICATION OF GEOLOGICAL CARBON STORAGE WITH DATA ASSIMILATION

IMPROVING UNCERTAINTY QUANTIFICATION OF GEOLOGICAL CARBON STORAGE WITH DATA ASSIMILATION

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof.dr.ir. H. Bijl,
chair of the Board for Doctorates,
to be defended publicly on 2 April 2026 at 10:00 hours

by

Gabriel SERRAO SEABRA

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. F.C. Vossepoel,	Technische Universiteit Delft, promotor
Dr. D. V. Voskov,	Technische Universiteit Delft, promotor
Independent members:	
Prof. dr. ir. M. Verlaan,	Technische Universiteit Delft
Prof. dr. ir. E.C. Slob,	Technische Universiteit Delft
Prof. dr. R.G. Hanea,	University of Stavanger and Equinor, Norway
Dr. F.O.L. Falcão,	Petrobras, Brazil
Dr. ir. D.S. Draganov,	Technische Universiteit Delft, reserve member

Keywords: Machine Learning, Data Assimilation, Geological Carbon Storage, Geomechanics, Proxy Models

Printed by: Gildeprint

ISBN 978-94-6496-535-3

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

I have approximate answers, and possible beliefs,
and different degrees of certainty about different things,
but I'm not absolutely sure of anything and there are many things I
don't know anything about.

Richard Feynman

CONTENTS

SUMMARY

Geological Carbon Storage (GCS) is an important component of strategies to reduce atmospheric CO_2 concentrations. The long-term security of stored CO_2, however, depends on a deep understanding of the subsurface. The rock formations used for storage are complex and varied, and our measurements are sparse, which makes it difficult to predict how the CO_2 plume will migrate and how underground pressures will change. These uncertainties create real risks: the stored CO_2 could leak, the injection could trigger small earthquakes, or the ground could move enough to damage surface infrastructures. To manage a GCS project safely and effectively, we need models that can predict these coupled flow and geomechanical effects and, more importantly, to understand and quantify the uncertainties present in each estimate of the process. The quality of our forecasts of the CO_2 plume behavior depends on how well we can define the model's uncertain parameters with the measurements available. This thesis presents a methodology that integrates physics-based simulation, data assimilation, and machine learning to improve uncertainty quantification for GCS. The work aims to deliver practical procedures that help quantify uncertainty in model predictions, guide the design of effective monitoring programs, and increase confidence in the long-term security of stored CO_2.

The research in this thesis starts with physics-based analysis and builds toward more complex methods that use machine learning. The initial work, presented in Chapter 2, investigates the value of geomechanical data for monitoring GCS projects, showing that surface deformation contains information about subsurface pressure changes and CO_2 plume migration. To make this information useful within a data assimilation context, which requires many forward model evaluations, we developed an efficient geomechanical proxy based on the nucleus of strain concept. As detailed in Chapter 3, we connect this proxy to the Ensemble Smoother with Multiple Data Assimilation (ESMDA), a well-established Data Assimilation method that iteratively updates an ensemble of reservoir models based on prior knowledge and observations. This integrated setup provides the basis for one of our most important contributions, which we explore in Chapter 4: a quantitative way to evaluate and design monitoring strategies using information theory. By using Shannon entropy (measured in nats) as a quantitative measure of information content and uncertainty, we could objectively compare different monitoring setups. We discovered that the observation frequency can be as important as their location. For the complex Brugge model, collecting data at multiple times from a simple linear array of sensors increased the uncertainty reduction from 20% to 63%—a threefold improvement. This suggests that in some cases, it may be more cost-effective to measure more frequently than to install more sensors. We demonstrate the influence of the reservoir's shape in determining the best sensor layout. For a symmetric reservoir, a sparse grid of 36 sensors gave us more information than a dense line of 51 sensors. These results provide a solid, quantitative basis for designing better, more cost-effective monitoring programs for GCS projects.

Two interconnected problems constrain the practical application of data assimilation in GCS. The first is the high computational cost of reservoir simulation. A single high-fidelity simulation can take many hours, one full 3D coupled flow-geomechanics simulation of a realistic field from the Brazilian Pre-Salt can easily last from 1 to 3 days, making in some cases the hundreds or thousands of runs needed for ensemble methods impractical. This forces the use of small ensembles, which introduces the second problem: properly representing complex, non-Gaussian geology. Many GCS sites are in formations like channelized reservoirs, where standard assimilation techniques like ESMDA, which are based on Gaussian assumptions, perform poorly. With small ensembles, these methods generate spurious correlations that cause the ensemble to collapse, complicating the reconstruction of reservoir geology and leading to a severe underestimation of uncertainty. One common solution, distance-based localization, is often ineffective because it is blind to geological connectivity. At the same time, variational methods like Randomized Maximum Likelihood (RML), which are more suitable to estimate non-Gaussian geology, have been difficult to apply because they require model gradients, which are typically only available with a dedicated adjoint model.

This thesis introduces two new hybrid methods to resolve these issues, which are the focus of Chapter 5. The first, Surrogate-based Hybrid ESMDA (SH-ESMDA), uses a fast ML surrogate for the intermediate assimilation steps, reducing the total computation time by more than 50%. We ensure physical consistency of our results by using the DARTS simulator for the final step of the data assimilation process. The second, Surrogate-based Hybrid RML (SH-RML), solves the adjoint problem by using the ML surrogate to compute the required gradients via automatic differentiation. This approach makes powerful variational optimization methods available for a wider class of simulators. To address complex geology, we also developed a new AI-enhanced localization technique, which is the contribution of Chapter 6. This procedure uses a diffusion model to generate a large "super-ensemble" of 5,000 realistic geological models. A lightweight ML proxy, trained on a small, high-fidelity "working ensemble," is then applied to this super-ensemble to get a much better estimate of the true covariance. This allows us to create localization weights that follow the actual geological connectivity instead of simple distance.

We tested these new methods on both conceptual and realistic reservoir models and found they offered clear improvements. By assimilating surface deformation data (Chapter 3), we reduced the uncertainty in the predicted CO_2 plume's location by 53.1% in a conceptual case and 18.5% in the complex Brugge field model. This demonstrates that geomechanical data, specifically vertical surface displacements on the order of 1–10 mm, enable quantitative constraint of subsurface permeability distributions and plume predictions, providing information that complements traditional well-based measurements. Our AI-enhanced localization also demonstrated strong quantitative performance. In our channelized reservoir tests (Chapter 6), the method preserved 15–40% more of the original ensemble variance than conventional techniques (achieving normalized variance NV = 0.86 vs. 0.71 for standard methods) while maintaining comparable data-match quality in terms of pressure RMSE. By maintaining higher posterior variance, our estimates avoided the overconfident uncertainty quantification (underestimation of true uncertainty) typical of standard ensemble methods with small sample sizes. Just as importantly, the updated permeability fields in the reservoir model retained their sharp, clear channel structures, a critical feature needed for the reservoir model to accurately predict CO_2 plume migration pathways.

In summary, this thesis demonstrates quantifiable uncertainty reduction in GCS through the integration of physics-based models, data assimilation, and machine learning. We developed and validated practical tools that quantify uncertainty in subsurface characterization. The geomechanical proxy-ESMDA framework (Chapters 3–4) achieved entropy reductions of 53.1% (2D conceptual) and 18.5% (Brugge field) when assimilating surface displacement measurements. The hybrid ML methods (Chapter 5) reduced computational time by over 50% while maintaining physical consistency, enabling feasible ensemble-based inference for complex systems. The diffusion-based localization (Chapter 6) preserved 15–40% more ensemble variance in channelized reservoirs than standard distance-based methods.

Important limitations remain. Extension to full 3D is necessary to capture vertical flow, gravity effects, and layered heterogeneity absent in our 2D cases. The geomechanical proxy assumes homogeneous rock properties, limiting applicability to heterogeneous formations. The ML-enhanced localization was tested only with pressure data; multi-physics observations (seismic, electromagnetic, InSAR) would test robustness across different sensitivity patterns. Finally, validation against field data from operational sites is needed to assess performance under real-world conditions with noise, gaps, and model-reality mismatches. By providing tools to quantify and manage uncertainty, these contributions help advance GCS from a promising concept to a predictable engineering practice for addressing climate change.

SAMENVATTING

Geologische koolstofopslag (GCS) is een essentieel onderdeel van de wereldwijde strategie om atmosferische CO_2-concentraties te verminderen. De langetermijnveiligheid van opgeslagen CO_2 hangt echter af van ons begrip van de ondergrond. De gesteenteformaties die voor opslag worden gebruikt, zijn complex en heterogeen, en metingen ervan zijn schaars, waardoor het moeilijk is te voorspellen hoe de CO_2-pluim zal migreren en hoe de ondergrondse vloeistofdruk zal veranderen bij CO_2-injectie. Deze onzekerheden creëren reële risico's: de opgeslagen CO_2 zou kunnen lekken, de injectie zou kleine aardbevingen kunnen veroorzaken, of de grond zou zodanig kunnen bewegen dat infrastructuur beschadigd raakt. Om een GCS-project veilig en effectief te beheren, hebben we modellen nodig die gekoppelde stromings- en geomechanische effecten kunnen voorspellen en, nog belangrijker, die ons helpen om de onzekerheden in elke schatting van het proces te begrijpen en te kwantificeren. De kwaliteit van onze prognoses van het gedrag van een geïnjecteerde CO_2-pluim hangt af van hoe goed we onzekere modelparameters kunnen bepalen met de beschikbare metingen. Dit proefschrift presenteert een methodologie die op fysica gebaseerde simulatie, data-assimilatie en machine learning integreert om onzekerheden geassocieerd met GCS beter te kwantificeren. In dit werk worden praktische procedures geformuleerd die helpen bij het kwantificeren van onzekerheid in modelvoorspellingen en bij het ontwerp van effectieve monitoringprogramma's en vergroot daarbij de langetermijnveiligheid van opgeslagen CO_2.

Het onderzoek in dit proefschrift begint met op fysica gebaseerde analyse en onderzoekt het gebruik van complexe methoden met machine learning. Het eerste werk, gepresenteerd in Hoofdstuk 2, onderzoekt de waarde van geomechanische gegevens voor het monitoren van GCS-projecten en toont aan dat oppervlakteverplaatsingsinformatie bevat over ondergrondse drukveranderingen en CO_2-pluimmigratie. Om deze informatie bruikbaar te maken binnen een data-assimilatiecontext, die een groot aantal voorwaartse modelevaluaties vereist, hebben we een efficiënte geomechanisch proxymodel ontwikkeld gebaseerd op het nucleus of strain-concept. Zoals gedetailleerd in Hoofdstuk 3, verbinden we deze proxy met de Ensemble Smoother with Multiple Data Assimilation (ESMDA), een gevestigde data-assimilatiemethode die iteratief een ensemble van reservoirmodellen bijwerkt op basis van a-priori kennis en waarnemingen. Deze geïntegreerde setup vormt de basis voor een van onze belangrijkste bijdragen, die we verkennen in Hoofdstuk 4: een kwantitatieve manier om monitoringstrategieën te evalueren en te ontwerpen met behulp van informatietheorie. Door Shannon-entropie (gemeten in nats) te gebruiken als kwantitatieve maat voor informatie-inhoud en onzekerheid, kunnen we verschillende monitoringopstellingen objectief vergelijken. We ontdekken dat de waarnemingsfrequentie net zo belangrijk kan zijn als de locatie van de waarnemingen. Voor het complexe Brugge-model verhoogt het verzamelen van gegevens op meerdere tijdstippen van een eenvoudige lineaire array van sensoren de onzekerheidsreductie van 20% naar 63%—een drievoudige verbetering. Dit suggereert dat het in sommige gevallen kosteneffectiever kan zijn om vaker te meten dan om meer sensoren te installeren. We demonstreren de invloed van de vorm van het reservoir bij het bepalen van de beste sensor-lay-out. Voor een symmetrisch reservoir geeft een raster van 36 sensoren ons meer informatie dan 51 sensoren die op een rij zijn geplaatst. Deze resultaten bieden een solide, kwantitatieve basis voor het ontwerpen van betere, kosteneffectievere monitoringprogramma's voor GCS-projecten.

Twee onderling verbonden problemen beperken de praktische toepassing van data-assimilatie in GCS. Het eerste is de hoge computationele kosten van reservoirsimulatie. Een enkele high-fidelity simulatie kan vele uren duren; één volledige 3D-gekoppelde stromings-geomechanica simulatie van een realistisch veld uit de Braziliaanse Pre-Salt kan bijvoorbeeld gemakkelijk 1 tot 3 dagen duren, waardoor in sommige gevallen de honderden of duizenden het simuleren van runs die nodig zijn voor ensemble-methoden onpraktisch wordt. Dit dwingt ons tot het gebruik van kleine ensembles, wat het tweede probleem met zich meebrengt: het adequaat representeren van complexe, niet-Gaussiaanse geologie. Veel GCS-locaties bevinden zich in formaties met heterogeniteiten zoals kanaalvormige hogedoorlaatbaarheid gebieden, waar standaard assimilatietechnieken zoals ESMDA, die gebaseerd zijn op Gaussiaanse aannames, slecht presteren. Met kleine ensembles genereert een methode als ESMDA schijnbare correlaties die leiden tot zogenaamd ensemble collapse, en waardoor de reconstructie van de reservoirgeologie wordt gecompliceerd en de onzekerheid ernstig wordt onderschat. De standaardoplossing, op afstand gebaseerde lokalisatie, is vaak ineffectief omdat deze blind is voor geologische connectiviteit. Tegelijkertijd zijn variationele methoden zoals Randomized Maximum Likelihood (RML), die meer geschikt zijn om niet-Gaussiaanse geologie te schatten, moeilijk toe te passen omdat ze modelgradiënten vereisen, die doorgaans alleen beschikbaar zijn met een specifiek adjoint-model.

Dit proefschrift introduceert twee nieuwe hybride methoden om deze problemen op te lossen, en die vormen de focus van Hoofdstuk 5. De eerste, Surrogate-based Hybrid ESMDA (SH-ESMDA), gebruikt een snelle ML-surrogaat voor de tussenliggende assimilatiestappen, waardoor de totale rekentijd met meer dan 50% wordt verminderd. We waarborgen fysische consistentie van onze resultaten door de DARTS-simulator te gebruiken voor de laatste stap van het data-assimilatieproces. De tweede, Surrogate-based Hybrid RML (SH-RML), neemt de behoefte aan een adjoint-model weg door de ML-surrogaat te gebruiken om de vereiste gradiënten te berekenen via automatische differentiatie. Deze benadering maakt krachtige variationele optimalisatiemethoden beschikbaar voor een bredere klasse van simulatoren. Om complexe geologie aan te pakken, hebben we ook een nieuwe AI-versterkte lokalisatietechniek ontwikkeld, die gepresenteerd wordt in Hoofdstuk 6. Deze procedure gebruikt een diffusiemodel om een groot "super-ensemble" van 5.000 realistische geologische modellen te genereren. Een lichtgewicht ML-proxy, getraind op een klein, high-fidelity "werkensemble", wordt vervolgens toegepast op dit super-ensemble om een veel betere schatting van de werkelijke covariantie te krijgen. Dit stelt ons in staat om lokalisatiegewichten te creëren die de werkelijke geologische connectiviteit volgen, en zich dus niet enkel op fysieke afstand baseren.

We hebben deze nieuwe methoden getest op zowel conceptuele als realistische reservoirmodellen en vonden dat ze duidelijke betere schattingen gaven. Door het assimileren van oppervlakteverplaatsingsgegevens (Hoofdstuk 3) reduceerden we de onzekerheid in de voorspelde locatie van de CO_2-pluim met 53,1% in een conceptueel model en 18,5% in het complexe Brugge-veldmodel. Dit toont aan dat geomechanische gegevens, specifiek verticale oppervlakteverplaatsingen in de orde van 1–10 mm, kwantitatieve beperking van ondergrondse permeabiliteitsdistributies en pluimvoorspellingen mogelijk maken, en informatie bieden die traditionele put-gebaseerde metingen aanvult. Onze AI-versterkte lokalisatie presteerde ook in kwantitatieve zin goed. In onze reservoirtests met hogedoorlaatbaarheids kanalen (Hoofdstuk 6) behield deze methode 15–40% meer van de oorspronkelijke ensemblevariantie dan conventionele technieken (genormaliseerde variantie $NV = 0{,}86$ vs. 0,71 voor standaardmethoden) terwijl de match met de data kwalitatief vergelijkbaar was, gemeten in termen van RMSE van de vloeistofdruk. Door hogere a-posteriori variantie te behouden in het ensemble, hadden onze schattingen niet te leiden onder overmoedige onzekerheidskwantificatie (onderschatting van werkelijke onzekerheid) die typisch is voor standaard ensemble-methoden met kleine ensembles. Net zo belangrijk was dat in de bijgewerkte permeabiliteitsvelden in het reservoirmodel hun scherpe, duidelijke kanaalstructuren behouden bleven, nodig om nauwkeurig de migratiewegen van de CO_2-pluim te voorspellen.

Samenvattend demonstreert dit proefschrift kwantificeerbare onzekerheidsreductie in GCS door de integratie van op fysica gebaseerde modellen, data-assimilatie en machine learning. We hebben praktische tools ontwikkeld en gevalideerd die onzekerheid in ondergrondse karakterisering kwantificeren. Het geomechanische proxy-ESMDA framework (Hoofdstukken 3–4) bereikte entropiereducties van 53,1% (2D conceptueel) en 18,5% (Brugge veld) bij het assimileren van oppervlakteverplaatsingsmetingen. De hybride ML-methoden (Hoofdstuk 5) reduceerden de rekentijd met meer dan 50% terwijl fysische consistentie behouden bleef, wat haalbare ensemble-gebaseerde inferentie voor complexe systemen mogelijk maakt. De diffusie-gebaseerde lokalisatie (Hoofdstuk 6) behield 15–40% meer ensemblevariantie in kanaalvormige hoge-doorlaatbaarheid structuren in reservoirs dan standaard op afstand gebaseerde lokalisatiemethoden.

Belangrijke beperkingen blijven bestaan. Uitbreiding naar volledige 3D schattingen is noodzakelijk om verticale stroming, zwaartekrachteffecten en gelaagde heterogeniteit vast te leggen die afwezig zijn in onze 2D-gevallen. De geomechanische proxy veronderstelt homogene gesteente-eigenschappen, wat de toepasbaarheid op heterogene formaties beperkt. De ML-versterkte lokalisatie werd alleen getest met vloeistofdrukgegevens; andere waarnemingen (seismische, elektromagnetische, InSAR waarnemingen, bijvoorbeeld) zouden ons in staat stellen om de robuustheid van de methoden te testen. Ten slotte is validatie met operationele waarnemingen uit het veld nodig om prestaties te beoordelen onder real-world omstandigheden met ruis, waarnemingshiaten en waarnemingen die niet matchen met de modelwaarden. Door tools te bieden om onzekerheid te kwantificeren en te beheren, helpen de bijdragen uit dit proefschrift mogelijkerwijs GCS vooruit van een veelbelovend concept naar een praktische toepassing voor het aanpakken van klimaatverandering.

RESUMO

O Armazenamento Geológico de Carbono (AGC) é um componente essencial das estratégias para reduzir as concentrações atmosféricas de CO_2. A segurança a longo prazo do CO_2 armazenado, no entanto, depende de uma compreensão adequada da subsuperfície. As formações rochosas usadas para armazenamento são complexas e variadas, e nossas medições são esparsas, o que torna difícil prever como a pluma de CO_2 migrará e como as pressões subterrâneas mudarão. Essas incertezas criam riscos reais: o CO_2 armazenado pode vazar, a injeção pode desencadear pequenos terremotos, ou o solo pode se mover o suficiente para danificar infraestruturas superficiais. Para gerenciar um projeto de AGC de forma segura e eficaz, precisamos de modelos que possam prever esses efeitos acoplados de fluxo e geomecânicos e, mais importante, entender e quantificar as incertezas presentes em cada estimativa do processo. A qualidade de nossas previsões do comportamento da pluma de CO_2 depende de quão bem podemos definir os parâmetros incertos do modelo com as medições disponíveis. Esta dissertação apresenta uma metodologia que integra simulação baseada em física, assimilação de dados e aprendizado de máquina para melhorar a quantificação de incertezas para AGC. O trabalho visa fornecer procedimentos práticos que ajudem a quantificar incertezas em previsões de modelos, orientem o design de programas de monitoramento eficazes e aumentem a confiança na segurança a longo prazo do CO_2 armazenado.

A pesquisa nesta dissertação começa com análise baseada em física e avança para métodos mais complexos que usam aprendizado de máquina. O trabalho inicial, apresentado no Capítulo 2, investiga o valor dos dados geomecânicos para monitorar projetos de AGC, mostrando que a deformação superficial contém informações sobre mudanças de pressão subsuperficial e migração da pluma de CO_2. Para tornar essa informação útil dentro de um contexto de assimilação de dados, que requer muitas avaliações de modelos diretos, desenvolvemos um proxy geomecânico eficiente baseado no conceito de núcleo de deformação. Como detalhado no Capítulo 3, conectamos este proxy ao Ensemble Smoother with Multiple Data Assimilation (ESMDA), um método de Assimilação de Dados bem estabelecido que atualiza iterativamente um conjunto de modelos de reservatório baseado em conhecimento prévio e observações. Esta configuração integrada fornece a base para uma de nossas contribuições mais importantes, que exploramos no Capítulo 4: uma maneira quantitativa de avaliar e projetar estratégias de monitoramento usando teoria da informação. Ao usar entropia de Shannon (medida em nats) como uma medida quantitativa de conteúdo de informação e incerteza, pudemos comparar objetivamente diferentes configurações de monitoramento. Descobrimos que a frequência de observação pode ser tão importante quanto sua localização. Para o modelo complexo de Brugge, coletar dados em múltiplos tempos de um array linear simples de sensores aumentou a redução de incerteza de 20% para 63%—uma melhoria de três vezes. Isso sugere que, em alguns casos, pode ser mais econômico medir com mais frequência do que instalar mais sensores. Demonstramos a influência da forma do reservatório na determinação do melhor layout de sensores. Para um reservatório simétrico, uma grade esparsa de 36 sensores nos deu mais informação do que uma linha densa de 51 sensores. Esses resultados fornecem uma base sólida e quantitativa para projetar programas de monitoramento melhores e mais econômicos para projetos de AGC.

Dois problemas interconectados restringem a aplicação prática da assimilação de dados em AGC. O primeiro é o alto custo computacional da simulação de reservatório. Uma única simulação de alta fidelidade pode levar muitas horas, uma simulação 3D completa acoplada de fluxo-geomecânica de um campo realista do Pré-Sal brasileiro pode facilmente durar de 1 a 3 dias, tornando em alguns casos impraticáveis as centenas ou milhares de execuções necessárias para métodos de conjunto. Isso força o uso de pequenos conjuntos, o que introduz o segundo problema: representar adequadamente geologia complexa e não-gaussiana. Muitos sítios de AGC estão em formações como reservatórios canalizados, onde técnicas de assimilação padrão como ESMDA, que são baseadas em pressupostos gaussianos, têm desempenho ruim. Com pequenos conjuntos, esses métodos geram correlações espúrias que fazem o conjunto colapsar, complicando a reconstrução da geologia do reservatório e levando a uma subestimação severa da incerteza. A solução padrão, localização baseada em distância, é frequentemente ineficaz porque é cega à conectividade geológica. Ao mesmo tempo, métodos variacionais como Randomized Maximum Likelihood (RML), que são mais adequados para estimar geologia não-gaussiana, têm sido difíceis de aplicar porque requerem gradientes do modelo, que normalmente só estão disponíveis com um modelo adjunto dedicado.

Esta dissertação introduz dois novos métodos híbridos para resolver essas questões, que são o foco do Capítulo 5. O primeiro, Surrogate-based Hybrid ESMDA (SH-ESMDA), usa um substituto de ML rápido para as etapas intermediárias de assimilação, reduzindo o tempo total de computação em mais de 50%. Garantimos a consistência física de nossos resultados usando o simulador DARTS para a etapa final do processo de assimilação de dados. O segundo, Surrogate-based Hybrid RML (SH-RML), resolve o problema adjunto usando o substituto de ML para calcular os gradientes necessários via diferenciação automática. Essa abordagem torna métodos de otimização variacional poderosos disponíveis para uma classe mais ampla de simuladores. Para abordar geologia complexa, também desenvolvemos uma nova técnica de localização aprimorada por IA, que é a contribuição do Capítulo 6. Este procedimento usa um modelo de difusão para gerar um grande “super-conjunto” de 5.000 modelos geológicos realísticos. Um proxy de ML leve, treinado em um pequeno “conjunto de trabalho” de alta fidelidade, é então aplicado a este super-conjunto para obter uma estimativa muito melhor da verdadeira covariância. Isso nos permite criar pesos de localização que seguem a conectividade geológica real em vez de distância simples.

Testamos esses novos métodos em modelos de reservatório tanto conceituais quanto realísticos e descobrimos que ofereciam melhorias claras. Ao assimilar dados de deformação superficial (Capítulo 3), reduzimos a incerteza na localização prevista da pluma de CO_2 em 53,1% em um caso conceitual e 18,5% no modelo complexo do campo de Brugge. Isto demonstra que dados geomecânicos, especificamente deslocamentos verticais de superfície na ordem de 1–10 mm, permitem restrição quantitativa das distribuições de permeabilidade subsuperficial e previsões de pluma, fornecendo informação que complementa medições tradicionais baseadas em poços. Nossa localização aprimorada por

IA também demonstrou forte desempenho quantitativo. Em nossos testes de reservatório canalizado (Capítulo 6), o método preservou 15–40% mais da variância original do conjunto do que técnicas convencionais (alcançando variância normalizada NV = 0,86 vs. 0,71 para métodos padrão) enquanto mantinha qualidade comparável de ajuste aos dados em termos de RMSE de pressão. Ao manter maior variância posterior, nossas estimativas evitaram a quantificação de incerteza excessivamente confiante (subestimação da incerteza real) típica de métodos de conjunto padrão com tamanhos de amostra pequenos. Igualmente importante, os campos de permeabilidade atualizados no modelo de reservatório retiveram suas estruturas de canal nítidas e claras, uma característica crítica necessária para que o modelo de reservatório preveja com precisão os caminhos de migração da pluma de CO_2.

Em resumo, esta dissertação demonstra redução quantificável de incerteza em AGC através da integração de modelos baseados em física, assimilação de dados e aprendizado de máquina. Desenvolvemos e validamos ferramentas práticas que quantificam incerteza na caracterização subsuperficial. O framework proxy geomecânico-ESMDA (Capítulos 3–4) alcançou reduções de entropia de 53,1% (2D conceitual) e 18,5% (campo Brugge) ao assimilar medições de deslocamento de superfície. Os métodos híbridos de ML (Capítulo 5) reduziram o tempo computacional em mais de 50% mantendo consistência física, possibilitando inferência viável baseada em conjunto para sistemas complexos. A localização baseada em difusão (Capítulo 6) preservou 15–40% mais variância de conjunto em reservatórios canalizados do que métodos padrão baseados em distância.

Importantes limitações permanecem. A extensão para 3D completo é necessária para capturar fluxo vertical, efeitos gravitacionais e heterogeneidade em camadas ausentes em nossos casos 2D. O proxy geomecânico assume propriedades de rocha homogêneas, limitando aplicabilidade a formações heterogêneas. A localização aprimorada por ML foi testada apenas com dados de pressão; observações multi-físicas (sísmicas, eletromagnéticas, InSAR) testariam robustez através de diferentes padrões de sensibilidade. Finalmente, validação contra dados de campo de sítios operacionais é necessária para avaliar desempenho sob condições reais com ruído, lacunas e incompatibilidades modelo-realidade. Ao fornecer ferramentas para quantificar e gerenciar incerteza, essas contribuições ajudam a avançar o AGC de um conceito promissor para uma prática de engenharia previsível para abordar as mudanças climáticas.

1

INTRODUCTION

1.1. MOTIVATION

CLIMATE change is reshaping how we produce and use energy. Atmospheric CO_2 has climbed to levels not seen in human history, and the warming that follows is already affecting ecosystems, infrastructure, and public health. The Intergovernmental Panel on Climate Change (IPCC) reports that keeping warming to 1.5°C requires steep and rapid cuts in emissions. In that context, Geological Carbon Storage (GCS)—often referred to as Carbon Capture and Storage (CCS)—is one of the tools we can use at scale. The International Energy Agency estimates that by 2050, CCS must handle billions of tonnes of CO_2 annually to meet global climate targets (IEA, 2022).

GCS stores CO_2 in deep subsurface formations after capture at industrial sources or from the air. The idea is straightforward, but implementing it at a large scale before widespread deployment is not. We need to select sites carefully, operate safely, and monitor the subsurface and surface for decades, ensuring it meets stringent safety and performance criteria. Technical challenges include caprock integrity, CO_2 migration and leakage risks, induced seismicity, and ground deformation. Societal concerns—such as public acceptance, regulatory compliance, and liability—are directly influenced by our ability to manage these technical risks. The technical challenges are hampered by limited knowledge of the subsurface, which is heterogeneous and only partially observed through sparse measurements (March et al., 2018; Witter et al., 2019).

Injecting large volumes of CO_2 triggers coupled flow-geomechanical-thermal-chemical processes across time and space: multiphase flow, rock deformation, geochemical reactions, and heat transfer. Pressure build-up can produce surface uplift that may affect infrastructure onshore or seafloor stability offshore (Vilarrasa et al., 2019). Dissolution of CO_2 into brine can drive convective mixing that improves storage security but is hard to predict at the field scale. These interacting processes, operating from pore to basin scales, motivate methods that combine physics, data, and clear uncertainty quantification. Small-scale heterogeneity and facies-scale capillary-entry contrasts critically affect residual and capillary trapping and overall plume evolution (Gershenzon et al., 2014).

We need computational approaches that bring together diverse data sources, quantify uncertainty, and make reliable predictions for decision-making. Data assimilation, rooted in Bayesian inference, offers a way to combine models and observations in a consistent manner (Evensen et al., 2022; Tarantola, 2005). Yet high-fidelity multiphysics simulations are expensive, geological priors are often non-Gaussian, and monitoring data are limited. This thesis develops frameworks that address these obstacles by combining data assimilation with machine learning and with efficient geomechanical surrogates. As an example of ML surrogates for subsurface flow processes in energy transition applications (including but extending beyond GCS), attention U-Nets have been used for image-to-image groundwater flow prediction with salient-region focus and strong runtime reductions relative to numerical solvers (Taccari et al., 2022).

1.2. LITERATURE REVIEW AND RESEARCH GAPS

1.2.1. GEOMECHANICAL MONITORING IN GCS

Surface deformation measurements can help reveal what is happening underground during CO_2 injection. The In Salah project in Algeria, for example, recorded centimeter-scale uplift linked to injection (Shi et al., 2013), prompting the use of coupled flow–geomechanics models to interpret such signals (Rutqvist, 2012). Most studies rely on InSAR, which is powerful but best suited for onshore regions with limited vegetation.

Several open questions remain. Sensors such as tiltmeters (Wang et al., 2020b), GPS (Wang, 2013), and fiber optics (Rucci et al., 2012) provide precise data, but we still lack clear rules for where and how densely to place them. Monitoring often stops with injection, even though post-injection deformation can be informative for long-term containment. Sensitivity of surface motion to rock properties and injection choices has not been mapped out across settings (Verma et al., 2021). Recent work questions the assumption of perfectly sealing caprocks, suggesting composite confining systems with discontinuities (Bump et al., 2023). We show that pressure dissipation in such systems can keep surface deformation evolving well after injection ends, which opens a window for long-term monitoring strategies.

1.2.2. DATA ASSIMILATION FOR SUBSURFACE CHARACTERIZATION

Ensemble-based methods, such as the Ensemble Smoother with Multiple Data Assimilation (ESMDA) (Emerick & Reynolds, 2013a), are widely used because they can handle nonlinearity and return uncertainty estimates. Early work on conditioning permeability fields to pressure data established the basis for modern history matching (S. Oliver et al., 1996). Bringing geomechanical data into these workflows is promising but still not common, in part because fully coupled simulations are costly. Studies in depleting gas reservoirs show the value of surface displacement data (Candela et al., 2022b); the same idea applies to GCS but has fewer demonstrations. Efficient semi-analytical geomechanical surrogates (Geertsma, 1973; Segall, 1992) offer a path to make such integrations practical. As an example within the energy transition, data assimilation has been applied to improve understanding of naturally fractured reservoirs (Seabra et al., 2022).

Handling non-Gaussian geological priors remains difficult. Channelized systems, in particular, violate Gaussian assumptions. Variational methods such as Randomized Maximum Likelihood (RML) (S. Oliver et al., 1996) can better respect non-Gaussian priors but often require gradients that are hard to obtain for complex simulators. Comparative studies across methods and geological settings are still limited for GCS.

Another gap is a clear way to compare monitoring data types and decide how to combine them. Prior work points to the value of surface deformation (Ringrose et al., 2013; Shi et al., 2019), but systematic, quantitative comparisons are scarce. Information theory offers tools to measure uncertainty reduction (Shannon, 1948) and has seen initial use in subsurface settings (Boso & Tartakovsky, 2018; Kim & Vossepoel, 2024). We build on these ideas in a GCS context.

1.2.3. MACHINE LEARNING INTEGRATION

Machine learning is increasingly used to speed up forward simulations and to support inference. Neural operators, such as the Fourier Neural Operator (FNO), learn solution operators for PDEs and have shown strong accuracy and generalization (Li et al., 2021; Lu et al., 2021), with early applications to CO_2 storage (Tang et al., 2022; Wen et al., 2022). Most studies focus on faster forward runs; tighter integration with data assimilation is less common. Practical issues include matching parameterizations between ML surrogates and physics-based models (Witte et al., 2023) and the cost of generating training data. Recent results suggest that modest datasets (on the order of 1,000–2,000 samples) can suffice for certain tasks (Wen et al., 2021), but this varies with geology and flow regime. On the prior side, generative models—GANs (Goodfellow et al., 2020), VAEs (Kingma & Welling, 2022), and score-based diffusion models (Song & Ermon, 2019; Song et al., 2020)—can help create realistic geological ensembles at scale. Low-fidelity models with machine-learning error correction can also reduce optimization cost, for example in well placement, while remaining close to high-fidelity outcomes (Tang & Durlofsky, 2022). A broader review of surrogate modeling practices—variable screening, design of experiments, combining multiple surrogates, and sequential sampling—provides complementary guidance for building efficient surrogates (Viana et al., 2021).

As part of this thesis, we released the open-source Python package `dageo` for data assimilation in geosciences (Werthmüller et al., 2025). It implements ESMDA with localization, Gaussian random-field generation, and simulator integration, and includes a 2D single-phase reservoir simulator for pressure and well responses (useful for teaching and rapid prototyping). The package provides documentation and examples for learning ESMDA concepts and applying them in practice; the code used in this thesis is available at https://github.com/tuda-geo/dageo.

1.2.4. MONITORING STRATEGY OPTIMIZATION

Monitoring designs often reflect regulations and practitioner judgment rather than explicit optimization. Jenkins (2015) shows how projects can lack quantitative tools for weighing trade-offs among options. The Value of Information (VOI) concept is well established in petroleum settings (Eidsvik et al., 2015), but GCS needs adaptations because the goal is containment, not production, and the time horizon is long. Information-theoretic measures (Kim & Vossepoel, 2024) can be adapted to quantify how much uncertainty a given monitoring setup reduces. We use these ideas to compare spatial sensor density with temporal sampling frequency and to evaluate mixed data types.

1.2.5. COMPUTATIONAL CHALLENGES AND SOLUTIONS

High-fidelity multiphysics models are expensive. A single run may take days, which limits ensemble-based studies at realistic scales. Analytical and semi-analytical surrogates can reduce cost by orders of magnitude, with the nucleus of strain approach (Geertsma, 1973; Segall, 1992) as a example. Extensions exist (van Wees et al., 2019), but broad validation for GCS is still developing. Reduced-order models have helped in other fields, yet strong advection and sharp CO_2 fronts challenge linear reduction techniques (Chaturantabut & Sorensen, 2010). Neural operators offer another option, but we must quantify and manage surrogate error when they are used inside inference workflows.

1.3. RESEARCH OBJECTIVES AND CONTRIBUTIONS

This thesis advances uncertainty quantification in GCS by combining geomechanical monitoring, data assimilation, and machine learning into practical workflows. The work addresses the following research questions:

1. How can surface deformation measurements inform subsurface characterization during and after CO_2 injection?
2. What is the relative value of spatial sensor density versus temporal sampling frequency for reducing uncertainty in plume predictions?
3. Can machine learning surrogates accelerate data assimilation while preserving physical consistency and geological realism?
4. How can generative models improve prior ensembles and localization strategies for non-Gaussian geological systems?

To address these questions, our contributions are:

- I study ground deformation during and after CO_2 injection, show where and when surface measurements carry relevant information, and discuss their use for long-term monitoring.
- I build an integrated workflow that couples a reservoir simulator (DARTS) with a semi-analytical geomechanical surrogate. This makes it feasible to use surface deformation data in ensemble-based inference at scale.
- I develop an information-theoretic approach to compare monitoring designs, quantifying the trade-off between spatial sensor density and measurement frequency.
- I integrate neural operators and transformer-based surrogates with ensemble methods, using fast surrogates for proposal and physics-based models for final predictions, with uncertainty carried through.
- I combine score-based diffusion priors with ML-enhanced localization in ensemble methods to better preserve ensemble variance and geological realism in channelized reservoirs.

1.4. THESIS ORGANIZATION

The thesis is structured as follows:

Chapter 2: Geomechanical monitoring insights. We analyze a Brazilian offshore reservoir case and show how ground deformation reflects subsurface pressure and CO_2 plume evolution. We identify informative sensor locations that are not restricted to wellheads and demonstrate the value of sensitivity analysis for monitoring and risk management.

Chapter 3: Integrating geomechanical surrogates with data assimilation. We couple semi-analytical geomechanics with ESMDA and show how surface deformation measurements improve subsurface characterization beyond well data. We validate the approach on two datasets and measure uncertainty reduction using entropy.

Chapter 4: Monitoring design using information theory. We evaluate monitoring options by combining CO_2 flow simulations, geomechanical surrogates, and data assimilation. We show that frequent measurements in time can match or exceed the value of dense spatial arrays. The method provides a quantitative way to compare designs.

Transition to ML-supported workflows: Chapters 3 and 4 use physics-based surrogates inside ensemble methods. The next chapters extend this idea by bringing machine learning surrogates into the loop.

Chapter 5: AI-supported data assimilation and uncertainty quantification. We integrate Fourier Neural Operators (FNOs) and a Transformer U-Net (T-UNet) with ES-MDA and RML. Fast surrogates propose updates, and high-fidelity simulators deliver final states, combining efficiency with physical consistency.

Chapter 6: Diffusion priors with ML-enhanced localization for data assimilation in GCS. We use large ensembles (5,000+ members) generated by score-based diffusion models and apply ML-enhanced localization to preserve ensemble variance in channelized geology. The approach improves data match and keeps geological structure.

Chapter 7: Conclusions. We synthesize the insights, their implications for GCS practice, and concrete directions for follow-up work.

Taken together, these chapters deliver methods, results, and practical guidance that help make CO_2 storage more predictable and more observable under uncertainty. The aim is to support decisions about where to monitor, how to combine data types, and how to run inference at field scale without losing sight of physics or geology.

2

UNVEILING VALUABLE GEOMECHANICAL MONITORING INSIGHTS: EXPLORING GROUND DEFORMATION IN GEOLOGICAL CARBON STORAGE

This chapter investigates flow and geomechanical responses to CO_2 injection using numerical simulations of a Brazilian offshore reservoir. The study demonstrates that ground deformation provides valuable monitoring insights, with centimeter-scale uplift patterns revealing subsurface pressure evolution and CO_2 plume behavior. Our findings show that maximum uplift locations shift away from injection wells as pressure dissipates through the caprock, challenging conventional wellhead-focused monitoring strategies. Sensitivity analyses on rock mechanical properties and injection rates establish the relationship between operational parameters and surface deformation, providing guidance for monitoring design and risk assessment in geological carbon storage projects.

Parts of this chapter have been published in Gabriel Serrão Seabra, Marcos Vitor Barbosa Machado, Mojdeh Delshad, Kamy Sepehrnoori, Denis Voskov, Femke C. Vossepoel, Unveiling Valuable Geomechanical Monitoring Insights: Exploring Ground Deformation in Geological Carbon Storage, *Applied Sciences* **14** (10), 4069 (2024) (Seabra et al., 2024a). Changes have been applied to make the text and figures consistent with the thesis.

2.1. INTRODUCTION

A CCS project involves CO_2 capture from high-emission industries and injecting it into geological formations, such as aquifers and depleted hydrocarbon reservoirs. One of the most critical barriers to long-term and large-volume CO_2 storage in geological formations is proof of safe and reliable storage. For that, CO_2 can take advantage of different trapping mechanisms in porous media (Delshad et al., 2013; Han et al., 2010; Nghiem et al., 2009; Rackley, 2017). For example, free CO_2 migration is controlled by the structural and stratigraphic trapping exerted by the caprock during the short term, encompassing the injection time, known as a primary mechanism. During this stage, the caprock plays an essential role in the security of the geological storage operation due to CO_2's buoyancy, which can move it up to the surface in onshore fields or the seafloor in the case of offshore storage sites. In the Frio CO_2 field demonstration project conducted in the U.S. (Hovorka, 2009), CO_2 was injected in a deeper zone below several well-known shale seals. However, some authors (Bump et al., 2023) are currently reviewing the need for a caprock to control the plume rising because (i) there is no prescriptive regulation concerning geologic seals and (ii) a composite confining system, e.g., a set of discontinuous barriers, can create long and tortuous paths that attenuate the mobile CO_2 saturation.

During the mid and long term, part of that mobile CO_2 will be dissolved in water (solubility trapping) as time passes, especially in low-salinity brine under high-pressure and low-temperature conditions (Duan & Sun, 2003; Portier & Rochelle, 2005). Molecular diffusion and brine solubility are the main physical phenomena that facilitate the dissolution of CO_2 into a brine. As CO_2 dissolves into saline water, the density of the resulting CO_2-dissolved brine increases. This density difference between the CO_2-dissolved brine and pure brine leads to the sinking of the denser CO_2-dissolved brine, while the pure brine rises (Farajzadeh et al., 2009; Neufeld et al., 2010). Density-driven convective mixing accelerates the rate of mass transfer of CO_2, promotes its dissolution, and enhances the stability and safety of the geological storage (Elenius et al., 2015; Taheri et al., 2021).

In addition to the solubility trapping, saturation changes caused by the rising plume can lead to more CO_2 being trapped as a residual phase due to relative permeability and capillary hysteresis (Nghiem et al., 2009; Qi et al., 2009; Spiteri et al., 2005). On saline aquifers, thermally enhanced dissolution of the CO_2 occurs (Lyu & Voskov, 2023). Beyond that, in a few cases, some CO_2 can be trapped as minerals because of the resulting pH of the brine and the mineralogy of the rock (Gunter et al., 2004; Xu et al., 2014). Therefore, those additional mechanisms increase storage security when the buoyant CO_2 is immobile in the pore space or no longer exists as a free phase (generally, as supercritical CO_2).

Geological CO_2 storage also involves a range of geomechanical risks associated with associated physical phenomena. These risks arise from the interaction between the injected CO_2 and the geological formations, leading to changes in rock mechanics, fluid dynamics, and chemical reactions. Some concerns include induced seismicity due to subsurface stress alterations, caprock integrity challenges from pressure-induced deformation, and changes in rock and fault strength owing to CO_2-related geochemical reactions. The integrity of wells is at risk due to CO_2 corrosion, and the long-term effects of mineral dissolution or precipitation add complexity to the stability and safety of storage sites (Nagel, 2001; Song et al., 2023; Wouters et al., 2021). Among these, excessive ground displacement emerges as an issue. For onshore projects, such ground displacement can lead to considerable damage to houses and buildings, as in the case of the Groningen gas field in The Netherlands (Wouters et al., 2021). In offshore environments, the primary concern lies with the impact on subsea infrastructure and surface facilities, such as the Ekofisk oil field in the North Sea (Nagel, 2001), where subsidence led to engineering interventions to mitigate the adverse effects. This study specifically focuses on ground deformation as the principal risk under investigation.

During the injection process and even afterward, the injection of large volumes of CO_2 into deep geological formations can cause significant pressure buildup and change stresses on the reservoir and the surrounding rocks. The pressure buildup induced by fluid injection can lead to expansion of the storage formation in proportion to the pressure increase. This expansion pattern can cause an uplift of the caprock, as observed in some actual GCS projects (Morris et al., 2011; Vilarrasa et al., 2019). The dissolution of CO_2 into water forms carbonic acid, which lowers the pH and can result in the dissolution of rock. Vaporization of water can also lead to the precipitation of minerals (Gharbi et al., 2013; Guyant et al., 2015; Schütz et al., 2011; Snippe et al., 2020). These geochemical reactions have the potential to alter the geomechanical properties of rocks and faults, which depends on the position of the displacement fronts (Ahusborde et al., 2024; Machado et al., 2023a). Therefore, injection conditions that may initially appear geomechanically stable could become unstable in the long term if geochemical reactions weaken the strength of the rocks or faults (Kim et al., 2018). Furthermore, in the event of fault reactivation, induced seismicity is possible, which has been associated with the risk of generating seismic activity due to stress changes caused by CO_2 injection (Braim et al., 2023; Harvey et al., 2021).

In the pursuit of advancing GCS technologies, the importance of conducting comprehensive uncertainty analyses on geomechanical parameters cannot be overstated. These analyses are essential for accurately characterizing the behaviors of geological formations under CO_2 injection scenarios. The variability inherent in rock properties, coupled with the dynamic nature of CO_2 injection processes, introduce uncertainties that must be addressed to ensure the efficacy and safety of CO_2 sequestration efforts. By evaluating the sensitivity of model outcomes to changes in rock mechanical properties, caprock permeability, and injection rates, researchers can identify critical thresholds and optimize monitoring and mitigation strategies. For instance, a study (Verma et al., 2021) on sensitivity analysis of geomechanical constraints in CO_2 storage highlights the impact of rock properties and injection rate on pressure buildup and CO_2 plume migration, underscoring the importance of accurately characterizing these parameters to predict storage site performance. Similarly, work on minimizing geomechanical risks under geological uncertainty through controlled CO_2 injection into storage reservoirs (Zheng et al., 2023) demonstrates the role of understanding and managing uncertainties in geomechanical properties to ensure caprock integrity and minimize leakage risks. Moreover, a study (AL-Ameri et al., 2016) on the long-term effects of CO_2 sequestration on rock mechanical properties further illustrates the potential changes in rock integrity over time due to CO_2–rock interactions, highlighting the need for long-term monitoring and model calibration to account for these changes. The research (Ahusborde et al., 2024) on uncertainty and global sensitivity analysis of CO_2 storage capacity prediction in deep saline aquifers also emphasizes the importance of identifying and managing uncertainties in geomechanical and other parameters to ensure accurate storage capacity estimates and safe CO_2 sequestration.

In the context of modeling vertical displacements due to reservoir activities, various monitoring technologies offer distinct advantages and limitations. High-precision tiltmeters are a method for monitoring direct surface deformation in onshore and offshore targets. With an accuracy equivalent to detecting an uplift of 0.01 mm over a typical distance of 10 km, tiltmeters provide precise measurements of ground surface movement (Davis et al., 2000), especially for subsidence mapping purposes in hydrocarbon fields or aquifers due to groundwater removal (Morris et al., 2011). Fiber Optic Sensing Technology offers high-resolution, continuous deformation profiles over long distances, making it suitable for detailed spatial mapping in both onshore and offshore fields. Other methods are most likely to be applied to onshore fields, like Interferometric Synthetic Aperture Radar (InSAR), which excels in monitoring large areas with millimeter accuracy, though its application is limited offshore or in dense forests. The Global Positioning System (GPS) provides three-dimensional displacement data, which are important for comprehensive movement analysis across both local and regional scales. Laser Scanning (LiDAR) generates precise topographical models, aiding in the detection of subtle ground deformations. Each method contributes uniquely to the comprehensive monitoring and analysis of ground movement, enabling a more nuanced understanding of the geomechanical effects of reservoir activities (Vilarrasa et al., 2019). A typical lower boundary used for ground displacement detection is 1 mm/year (Bouqueta et al., 2022). By providing precise measurements of ground surface movement, these monitoring methods enable the use of ground deformation as a surface tool to monitor the redistribution of CO_2 and the diffusion of pressure into the subsurface.

In this context, the primary objective of this study is to assess the technical feasibility of the proposed idea by employing a comprehensive numerical simulation framework that combines fluid flow and geomechanics in an offshore reservoir case study. Therefore, the objectives of this study can be described as follows:

- Study the ongoing dissipation of pressure in the reservoir even after the CO_2 plume has stabilized;
- Analyze the continued ground deformation due to CO_2 movement and pressure dissipation after the cessation of CO_2 injection;
- Investigate how the uplift behavior can be utilized for long-term monitoring of CO_2 redistribution;
- Provide technical arguments supporting the feasibility of monitoring the ground uplift caused by CO_2 storage in offshore targets, even after the injection has ceased;
- Examine the role of uplift in monitoring scenarios where the CO_2 plume is confined but the pressure continues to dissipate;
- Conduct sensitivity analyses on geomechanical properties and injection rates to enhance the conclusions of our base case.

The concept of modeling surface deformation data to monitor CO_2 storage activity in porous media is not new. It has been successfully demonstrated in projects like the In Salah CCS project in Algeria, where the geomechanical response at the ground surface to subsurface CO_2 injection was effectively reproduced using a coupled reservoir simulation–geomechanical model (Shi et al., 2013). However, it is worth noting that in their study, the authors relied on interferometric synthetic–aperture radar (InSAR) data to calibrate the numerical model (Zhang et al., 2022). In contrast, our approach aims to validate the feasibility of directly measuring ground uplift as a monitoring tool beyond the InSAR data, which are restricted to onshore and non-covered areas. This research seeks to contribute to the advancement of the field by exploring a supplementary method that can potentially offer more cost-effective and accessible monitoring solutions. This also has the potential to aid in accurately determining the optimal placement of the monitoring instrument within the so-called Area of Review. The Area of Review is a regulatory term that encompasses the storage site, focusing on the edges of the injected CO_2 plume and areas of elevated pressure (Bump et al., 2023). It necessitates a monitoring strategy informed by a deep understanding of the geomechanical state of the reservoir, rather than assumptions based on the CO_2 injection locations alone. By incorporating robust geomechanical analyses, we can optimize the surveillance of CO_2 sequestration sites, thereby enhancing the fidelity of monitoring systems and maintaining the integrity of long-term carbon storage solutions. This, in turn, can support the exploration of alternatives, such as drilling new wells to produce brine, to relieve the pressure buildup associated with such surface events (Buscheck et al., 2012; Kim et al., 2023). The technical evidence to be explored in this study will also provide new insights into the magnitude of geomechanical events and their occurrence timeframe, supporting the application of several ground deformation monitoring tools for GCS.

The following aspects are outside the scope of this study:

- Accurately representing enhanced CO_2 dissolution at the reservoir scale, as performed by Lei and Luo (2021) and Singh et al. (2019). This assumption is based on the fact that the gridblocks used in our coupled geomechanics and numerical flow simulations may not accurately capture the dissolution of CO_2 in brine. Our study primarily focuses on ground deformation; therefore, a simplified representation of the dissolved CO_2 was adopted to enable the simulation of a field-scale model. This model encompasses the caprock and the surrounding rocks, allowing for a comprehensive sensitivity analysis regarding the geomechanical properties.

- Consideration of impurities and free water content in the CO_2 stream injected.

- Modeling the dry-out effect due to water vaporization with CO_2 injection or another injectivity issue, which can be found in Machado et al. (2023b).

- CO_2 leakage through wells with poor cement jobs, as pointed out by Gholami et al. (2021), which could be the most important reason for migration and leakage.

- Evaluating other monitoring techniques, such as time-lapse seismic surveys or microseismic methods.

- Evaluating data assimilation or inversion methods of ground displacements for CO_2 plume tracking.
- Modeling geochemical reactions between CO_2 and rock minerals, as performed by Farshidi et al. (2013) and Machado et al. (2024).
- Modeling caprock wettability changes.
- Modeling of fault activation and fracture propagation induced by CO_2 injection.
- Evaluating the impact of temperature on geomechanical behavior.
- Assessing small-scale rock microstructures influencing CO_2 migration and storage.

The study is organized as follows. Section 2.2 describes basic petrophysical modeling for both sandstone and shale, which is important for the understanding of the geological context of the study area. Section 2.3 describes the modeling of CO_2 entrapment. Section 2.4 covers the geomechanical modeling of the CO_2 injection. Section 2.5 delves into the methodology and results, showcasing the numerical simulation outcomes and discussing the implications of the ground deformation phenomena observed. It also presents a sensitivity analysis to assess the impact on the geomechanical behavior of the reservoir and the caprock system with wide variations in parameters. Lastly, the study is concluded by summing up the findings and giving an overview of contributions that the study makes to the literature on GCS.

2.2. PETROPHYSICAL MODELING FOR SANDSTONE AND SHALE

The saline aquifer geological model in this study, including geometry and porosity/permeability distributions, was derived from a benchmark model, referred to as UNISIM-I (Avansi & Schiozer, 2015), constructed based on structural, facies, and petrophysical data obtained from the Namorado Oil Field in the Campos Basin, Brazil. The 3D model, including the over, lateral, and underburden, consists of 114 × 66 × 56 gridblocks with lateral dimensions of 170×170 m^2 and variable thickness (Figure 2.1), discretized using a corner point grid. Table 2.1 provides a summary of the mean values for petrophysical properties assigned to the model. The model represents a heterogeneous sandstone reservoir, and the associated siliceous shaly caprock was assumed to have homogenous properties, depicted in Table 2.1. The shale properties are typical for this type of rock (Gercek, 2007; Molina et al., 2017). Figure 2.1 shows the porosity distribution in the model. The low-porosity zone in blue represents the shale layers, and it highlights the area of the CO_2 injection well, positioned on a high-porosity lower flank of the reservoir. This specific positioning, a practice observed in GCS projects to optimize the injection (Kuk et al., 2020; Shchipanov et al., 2022), leverages gravitational forces to enhance CO_2 dispersion and retention within the high-porosity zones, thereby optimizing injection. The blue zones surrounding this area represent the shale layers with low porosity, serving as a caprock to contain the CO_2. This placement is critical to ensure the containment of the CO_2 within the target zone. The left figure presents a 3D view of the geological model. The right figure provides a slice through the model, emphasizing the depth of the reservoir and allowing for a focused examination of the geological features and porosity at a specific subsurface level.

The dynamic properties used in this model are based on the characterization proposed by Machado et al. (2024), and it will be discussed in the next paragraphs. The relative permeability curves for CO_2 and brine in sandstone and shale were obtained from Bennion and Bachu (2008). The CO_2-brine capillary pressure curves were obtained from a J-function fitted to the data from Abdoulghafour et al. (2020) for the sandstone and from Bennion and Bachu (2008) for the shale. Figure 2.2 and Figure 2.3 show the final drainage curves used in the simulation for the aquifers and their caprock, respectively. The imbibition curves are generated according to the hysteresis model, which is detailed in the next section.

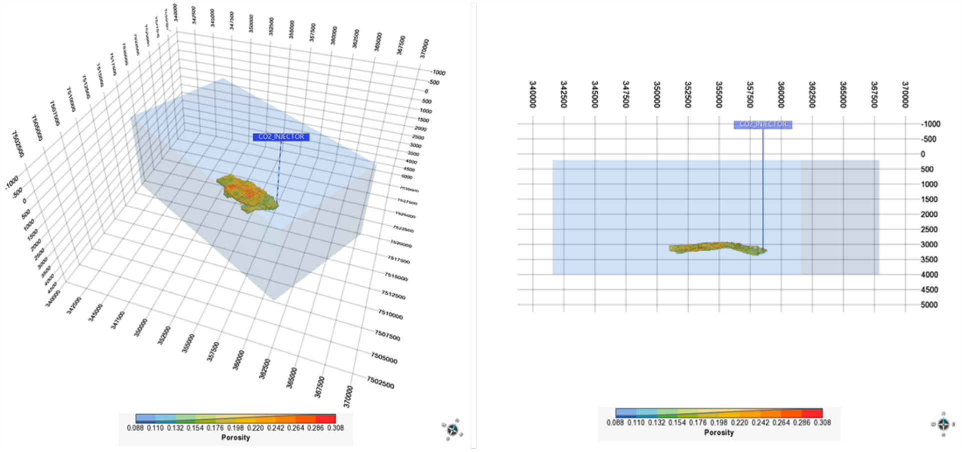


Figure 2.1: 3D view of the geological model showing porosity distribution and injector well position (left). A sliced view highlighting the depth of the reservoir and porosity characteristics (right). Coordinates are in meters.

Table 2.1: Summary of main petrophysical properties used in the geological model.

Property	**Sandstone**	**Shale**
Porosity (ϕ)	0.20 (mean)	0.10
Permeability (k)	300 mD (mean)	0.001 mD
k_v/k_h ratio	0.1	0.1
Pore compressibility	5.8×10^{-7} kPa^{-1}	5×10^{-8} kPa^{-1}
Young's modulus	1 GPa	10 GPa
Poisson's ratio	0.25	0.30
Relative permeability	-	-

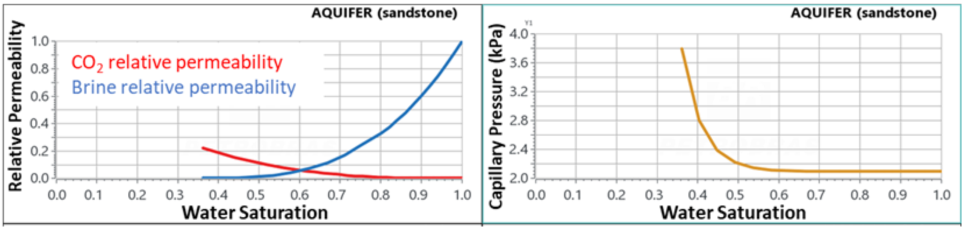


Figure 2.2: Drainage relative permeability (left) and capillary pressure (right, in orange) for the sandstone.

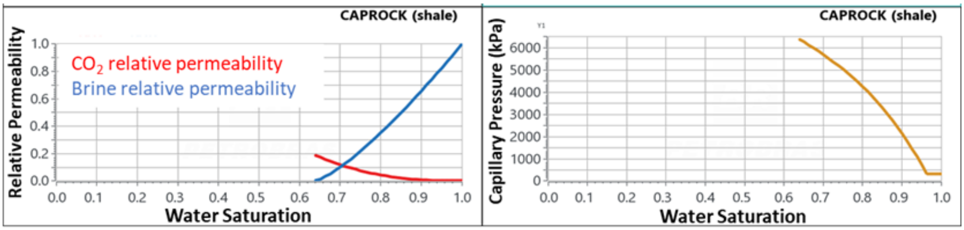


Figure 2.3: Drainage relative permeability (left) and capillary pressure (right, in orange) for the shale caprock.

2.3. Modeling CO$_2$ entrapment for sandstone and shale formations

The sandstone saline aquifer and its shale caprock in the model in this study were numerically simulated using CMG-GEM (Computer Modelling Group Ltd., 2023). CMG-GEM is known for its versatility and robustness in predicting and analyzing various thermodynamic properties and fluid behavior for GCS projects (Delshad et al., 2013; Farajzadeh et al., 2009; Han et al., 2010; Landry et al., 2016; Nghiem et al., 2009; Rackley, 2017). It is based on the discretization of component conservation and energy balance equations in space and time using finite volume and finite difference methods (Balhoff, 2022; Machado, 2023). One of the strengths of the CMG-GEM model lies in its accurate prediction of CO$_2$ phase behavior for wide ranges of pressure and temperature. It can effectively capture the transition of CO$_2$ from gaseous to liquid states, as well as its behavior in supercritical conditions. Overall, the CMG-GEM simulation model enables accurate prediction and analysis of CO$_2$ behavior and the following trapping mechanisms:

- CO$_2$ solubility in brine is modeled in this study by applying the Li and Nghiem model (Li & Nghiem, 1986). This model has been calibrated using published experimental data, and it calculates Henry's constant based on Equation 2.1. The impact of brine salinity on gas solubility in the aqueous phase is accounted for through the salting-out coefficient (Bakker, 2003).

$$\ln H_i = \ln H_i^* + \frac{\overline{v_i} RgT}{p} - \frac{p}{p^*} \tag{2.1}$$

where H_i is Henry's constant at current pressure (p) and temperature (T); H_i^* is Henry's constant at reference pressure (p^*) and temperature (T); $\overline{v_i}$ is partial molar volume at

infinite dilution;
Rg is the universal gas constant; and i denotes the species dissolved in water ($CO_{2(aq)}$ in this work). The H_i^* constant was computed considering the initial conditions of the models. Solubility trapping in brine can be enhanced through physical diffusion, which is mandatory for an accurate representation of the convective mixing to obtain a grid-converged solution. To model this effect, even in a simplified way due to our model's scale, the diffusion coefficient (D_{diff}) for supercritical CO_2 in brine is applied to compute the effective CO_2 diffusion (D_{eff}) considering tortuosity (τ) (Rezk et al., 2022):

$$D_{\text{eff}} = \frac{D_{\text{diff}}}{\tau} \tag{2.2}$$

In shales, where diffusion is a relevant mechanism due to lower permeability, the tortuosity value ranges from approximately 40 to 70 (Busch et al., 2008). An effective diffusion equal to 2.8×10^{-7} cm^2/s (Montegrossi et al., 2022) is used in the simulations. The residual CO_2 trapping due to the relative permeability and capillarity hysteresis with the saturation changes was modeled with the maximum gas trapped (S_{gt}) converted to the Land's constant C (Land, 1968) in the two-phase Carlson's model (Carlson, 1981), according to Equation 2.3:

$$S_{gt} = \frac{S_{g_{max}}}{1 + C S_{g_{max}}} \tag{2.3}$$

where $S_{g_{max}}$ is the maximum gas saturation. Burnside and Naylor (Burnside & Naylor, 2014) obtained a S_{gt} distribution for sandstone, shale, and carbonates based on more than 30 published coreflood data for CO_2 and brine. The mean values for each lithology are presented in Table 2.2.

Table 2.2: Maximum gas trapped due to hysteresis of relative permeabilities and capillary pressure (Land, 1968).

Lithology	**Sandstone**	**Shale**
S_{gt}	0.25	0.35

2.4. GEOMECHANICAL MODELING OF CO_2 INJECTION

As emphasized in the Introduction, the modeling of geomechanics for CO_2 injection in sandstone and shale formations is an aspect of ensuring the safety and efficacy of GCS projects. CMG-GEM's Finite Element geomechanical coupling provides a sophisticated tool for analyzing the interactions between rock formations and the injected CO_2 (Computer Modelling Group Ltd., 2023). This module relies on the fundamental principles of rock mechanics, including stress–strain relationships, deformation, and failure theories. The software employs various constitutive models, such as elasticity, elastoplasticity, and caprock models, to represent the complex behavior of geological materials under different loading conditions. These models are essential for understanding how reservoir rocks react to the injection of CO_2, including potential fracturing and changes in permeability and porosity.

The CMG-GEM geomechanics module leverages three fundamental equations to simulate the mechanical response of rock formations to CO_2 injection. The equilibrium

Equation 2.4 asserts the force balance within the reservoir, indicating that the divergence of the stress tensor ($\boldsymbol{\sigma}$) is balanced by body forces (B). The strain–displacement Equation 2.5 relates the deformation of the reservoir to displacement fields (u), which is essential for understanding how the rock fabric responds to changes. Finally, the stress–strain Equation 2.6 ties the strain ($\boldsymbol{\epsilon}$) within the rock to the induced stress, incorporating the impact of pore pressure (p), temperature change (ΔT), and rock stiffness, embodied in the stiffness tensor ($\mathbf{K}$) (Balhoff, 2022). These core relationships are integral to predicting the deformation of the rock, its fracturing potential, and changes in permeability and porosity, all critical for assessing the effects of CO_2 injection.

$$\nabla \cdot \boldsymbol{\sigma} - B = 0 \tag{2.4}$$

$$\boldsymbol{\epsilon} = \frac{1}{2}(\nabla u + (\nabla u)^T) \tag{2.5}$$

$$\boldsymbol{\sigma} = \mathbf{K} : \boldsymbol{\epsilon} + \alpha_B p + \eta \Delta T \mathbf{I} \tag{2.6}$$

where $\boldsymbol{\sigma}$ represents the stress tensor; B denotes body forces; $\boldsymbol{\epsilon}$ is the strain tensor; u is the displacement vector; $\mathbf{K}$ is the stiffness tensor; α_B is the Biot-Willis coefficient; p stands for pore pressure; η is the thermoelastic coefficient; ΔT represents the temperature change; and $\mathbf{I}$ is the identity tensor.

In this study, linear elastic modeling was adopted in our flow-geomechanics simulations of ground deformation due to CO_2 injection. This conservative approach aligns with our focus on understanding ground deformation. It is worth noting that if plasticity were to occur, it would likely lead to even larger displacements. Therefore, by using linear elasticity, we are taking a cautious approach and providing a baseline for potential deformation. This method strikes a balance between accuracy and computational efficiency, enabling us to realistically represent the mechanical behavior of the reservoir under CO_2 injection stress while managing computational resources effectively.

Our reservoir model's predictive capability for geomechanical responses to CO_2 injection is contingent on an accurate representation of the in situ stress profile. This profile is established by mapping pore pressure and minimum total stress from a depth of −2900 m to −3400 m within the injection well. These mappings are important for evaluating rock deformation potential and failure risk. We assume hydrostatic conditions for the pore pressure gradient, while the minimum stress gradient is aligned with observed values for passive margin regimes, specifically reflecting the low tectonic stress environment of Brazil's offshore Campos Basin (Assumpção, 1998; Bizzo, 2017; Lima et al., 1997). By applying a stress gradient of roughly 22 kPa/m, our model captures the geomechanical nuances of the Campos Basin's passive margin stress regime (Bjørnarå et al., 2010; Lima et al., 2021).

With respect to meshing, it was possible to employ a dual-grid system, which optimizes computational efficiency. Figure 2.4 shows the grid of the model used in this work, with 421,344 elements. In this study, we maintain a consistent grid for both fluid flow and geomechanical simulations across the entire model. This approach differs from typical reservoir geomechanics studies, where flow is often simulated only within the reservoir, while geomechanical calculations are extended to the surrounding formations. By applying this approach, the essential processes, such as the dissipation of pressure beyond the confines of the reservoir, can be captured. This method ensures that no critical physical interactions are overlooked, addressing a gap that is often present in geomechanical studies (Machado et al., 2023a; Shi et al., 2013). This study considers a one-way coupling option for integrating the reservoir simulator with the geomechanical module. The one-way coupling allows for representing the impact of fluid flow on geomechanics without considering the reverse effect. This approach is suitable for studies where the influence of geomechanical changes on fluid dynamics is not a primary concern. On the other hand, the full coupling method allows for a more interactive exchange between fluid flow and geomechanics, with either solving each aspect in turn and updating their states iteratively (Kim et al., 2011) or applying a monolithic approach (Garipov & Hui, 2018). A fully coupled method is more suited for analyses requiring a nuanced understanding of fluid dynamics and rock mechanics (Tran et al., 2002). While the software offers multiple ways to couple these simulations, ranging from fully coupled to one-way interactions, for our specific study, we have chosen to use a one-way coupling approach. This decision aligns with our study's objective, as we are not investigating the feedback of geomechanical responses on fluid dynamics. The chosen method balances computational efficiency and the accuracy needed for our analysis, allowing us to dynamically simulate the reservoir's behavior concerning CO_2 injection without the added complexity of full coupling.

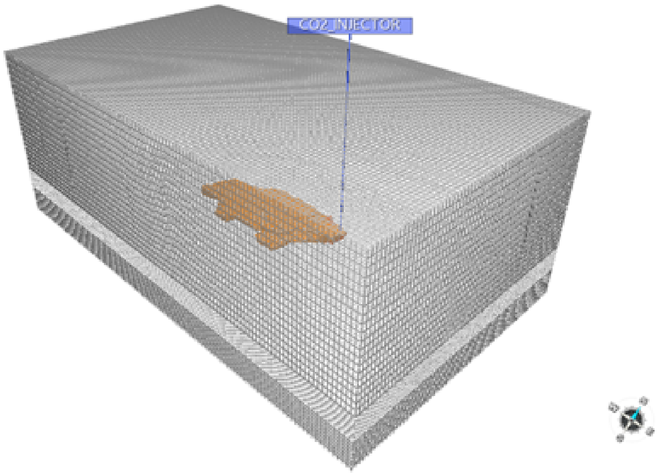


Figure 2.4: UNISIM-I extended model grid for flow and geomechanical simulations.

2.5. METHODOLOGY AND RESULTS

The methodology employed in this study characterizes the diffusion of pressure in porous media resulting from CO_2 injection and its impact on the ground deformation during and after the cessation of the injection. This analysis has been carried out using the UNISIM-I model described in the previous sections, which features a large, single-shale caprock. By utilizing this model, we explore the geomechanical impacts of CO_2 injection, specifically simulating ground uplift, the interplay between rock properties, injection rates, pressure buildup, and CO_2 migration. We also investigate the role of caprock in pressure dissipation, offering insights into optimizing monitoring strategies. Furthermore, a sensitivity analysis is conducted to broaden the scope of our conclusions by considering various levels of mechanical rock properties and different rates of CO_2 injection.

2.5.1. CO_2 INJECTION FLOW RESULTS

In the simulation, the CO_2 was injected at a constant rate of 2200 tons per day over 20 years, followed by 50 years of monitoring, resulting in 70 years of simulation. This extended duration was chosen to simulate realistic long-term injection scenarios, reflecting the commitment to sustained carbon sequestration efforts. The total injected volume culminated in 16 million metric tons of CO_2, as depicted by the cumulative injection curve (Figure 2.5, blue curve) and the average reservoir pressure, showcasing an initial increase followed by a dissipation phase (Figure 2.5, red curve). The pressure increase during the injection period (Figure 2.5) can be interpreted as a clear response of the reservoir to the large volume of CO_2 being stored. The steep rise in pressure suggests that the aquifer has a limited capacity to accommodate the injected CO_2 without significant pressure buildup, which implies that the reservoir volume or the permeability may not be as large as required for the injected volumes. After reaching a peak, 20 years after the start of the injection, the pressure curve shows a gradual decline, signaling the onset of pressure dissipation. This phase transition may denote the beginning of CO_2 redistribution within the reservoir and potential migration into adjacent rocks, suggesting that the initial storage volume is approaching its capacity limit. The phase of pressure decline may also suggest that the pressure is being alleviated not just within the reservoir but also across the surrounding rock formations. This diffusion into the adjacent rocks can occur even in the absence of significant CO_2 migration due to the nature of pressure transmission through the pore fluids, which precedes the physical movement of CO_2. The pressure within a fluid-filled porous medium can propagate more rapidly than the CO_2 itself due to the fluid's compressibility and the interconnectedness of the pore spaces, allowing for a swift response and redistribution of pressure throughout the surrounding rocks.

Figure 2.6 captures the evolution of reservoir pressure via delta pressure visualizations, marking changes from the initial state due to CO_2 injection. Panel (A) is the 3D view at the end of the 20-year injection phase, and it shows a relevant delta pressure gradient not only around the injector, with warmer hues indicating more significant pressure increases, but also within the entire confined reservoir. This confinement and the magnitude of pressure buildup at the injection end, evident from the vibrant coloration in Figure 2.6A, present subsequent long-term dissipation into the surrounding shaly rocks. Figure 2.6B's vertical cross-section near the injector after 20 years reveals the onset of this dissipation, with color transitions highlighting pressure migration beyond the reservoir boundaries. By 70 years, Figure 2.6C illustrates that the pressure has dissipated even further into the surrounding rocks. This extended spread of pressure, despite the absence of considerable CO_2 movement (Figure 2.7), indicates that the diffusion of pressure through the fluid phase outpaces the physical redistribution of CO_2, emphasizing the role of the interconnected pore network in the broader geological environment for managing injected pressures.

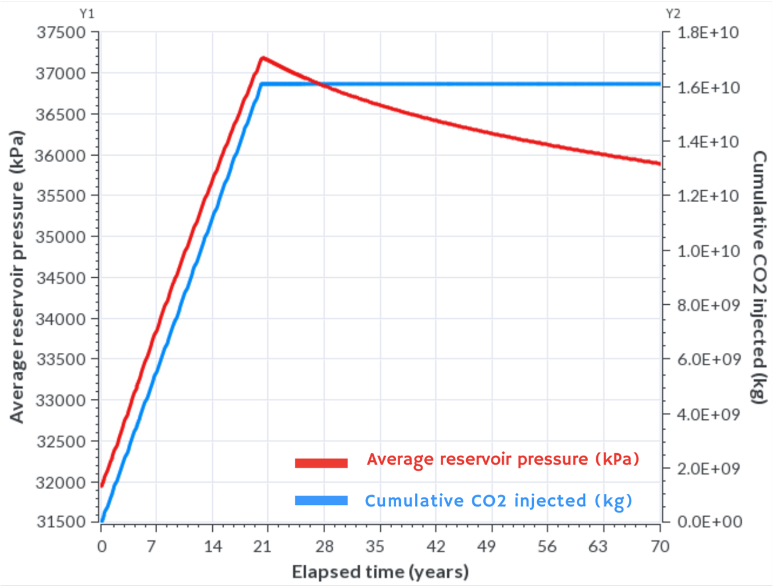


Figure 2.5: Reservoir pressure and cumulative CO_2 injection over 20 years and subsequent 50 years of monitoring period. The blue curve indicates the total CO_2 injected, while the red curve shows the corresponding average pressure in the reservoir.

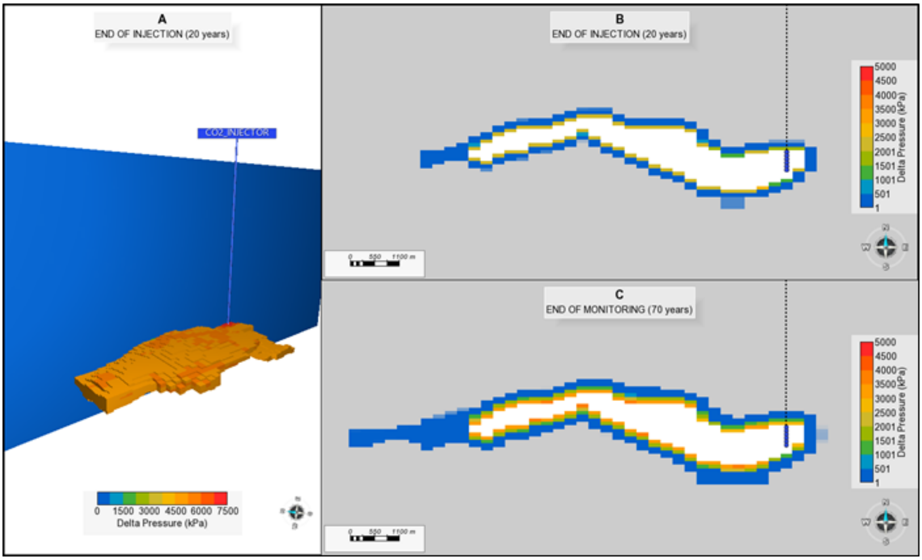


Figure 2.6: Delta pressure (difference between initial and current pressure per gridblock) distribution in the reservoir and surrounding rocks. Panel (A) shows a 3D view of the reservoir at the end of the 20-year CO_2 injection phase, indicating high delta pressure near the injection well. Panels (B,C) depict vertical cross-sections at 20 and 70 years, respectively, highlighting pressure dissipation into the surrounding shaly rocks over time. On panels (A,B), the reservoir is depicted in white.

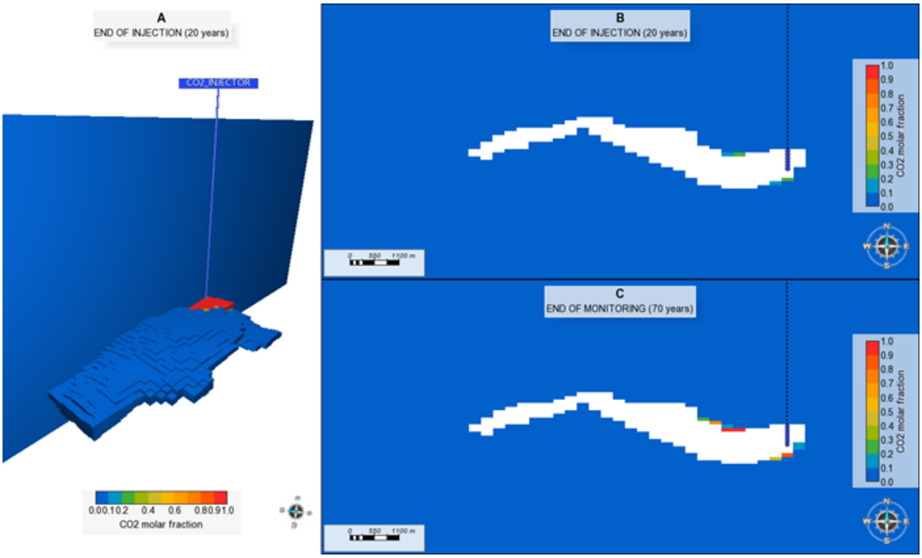


Figure 2.7: CO_2 global molar fraction post-injection and monitoring phases. Panel (A) shows the CO_2 concentration around the injector at the end of the 20-year injection period. Panels (B,C) reveal vertical cross-sections at 20 and 70 years, respectively, with minimal CO_2 migration into the caprock. On panels (A,B), the reservoir is depicted in white.

Figure 2.8 illustrates the CO_2 migration within the reservoir at two times. The left panel of Figure 2.8 reveals the state of the reservoir immediately after 20 years of injection, highlighting the buoyant rise of CO_2 towards the less dense upper regions, a behavior that emphasizes the consistency of our simulation's fluid and rock physics. The right panel demonstrates the situation after 70 years. In this period between 20 and 70 years, the CO_2 within the reservoir has moved upward but not significantly beyond its boundaries, indicating the effectiveness of the caprock's structural integrity and capillary forces in limiting vertical movement. This containment is further evidenced by the minimal lateral dispersion into the surrounding rocks, evidencing the robustness of natural trapping mechanisms and the security of CO_2 storage over an extensive monitoring period. Even using a coarse mesh, the model was able to capture the vertical movement of CO_2, demonstrating the buoyant rise and containment within the geological formation over time. This outcome suggests that even with less detailed spatial resolution, the processes governing CO_2 migration and trapping can be effectively simulated, providing insights into the overall behavior of the CO_2 plume. However, if the primary objective of a study is to investigate the nuanced details of CO_2 plume dynamics, such as precise migration paths or small-scale heterogeneities within the reservoir, a much finer grid should be considered. A refined mesh would enhance the model's ability to capture fine-scale physical processes and interactions, offering a more detailed and accurate representation of CO_2 behavior within the subsurface. This is especially important for projects where the exact delineation of the CO_2 plume and its interaction with the geological environment are critical for safety assessment and regulatory compliance.

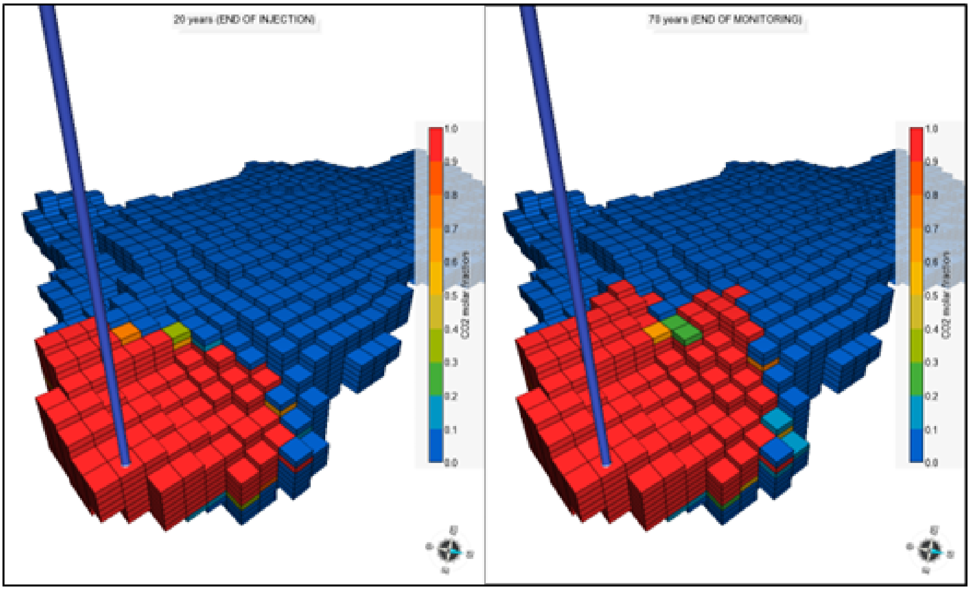


Figure 2.8: CO_2 global molar fraction distribution at the end of injection and post-monitoring periods. The (left) illustrates the reservoir immediately after the 20-year injection phase and the (right) panel reveals the CO_2 distribution after 70 years.

2.5.2. GEOMECHANICAL RESULTS

Having established the flow behavior and CO_2 containment within the reservoir, we now analyze how these subsurface processes manifest as measurable surface deformation. This part of the study addresses questions raised earlier, such as the practicality of using uplift measurements for monitoring purposes and the implications of subsidence patterns over the long term. We also look at how these ground movements relate to the position of the injection well and the dissipation of pressure during the post-injection phase. Different strategies for the strategic placement of monitors are considered, aiming to enhance the detection of vertical deformations.

Figure 2.9 illustrates the temporal evolution of seabed uplift due to subsurface CO_2 injection, captured at four different times: 3, 10, 20, and 70 years. The negative values shown correspond to an upward displacement against the downward-oriented Z-axis, reflecting rock expansion as a result of CO_2 injection. Initially, significant uplift is observed, especially around the 20-year mark, indicating significant rock expansion due to CO_2 injection and the reservoir pressure buildup. Over time, as the injection ceases and the system begins to equilibrate, the 70-year mark reveals a slightly decreased uplift at the reservoir's center. This change is attributed to the dissipation of pressure into the surrounding rock formations, which acts as a pressure-relief mechanism for the reservoir. The uplift is not uniform, with the largest displacement occurring slightly away from the wellhead, indicative of the CO_2 migrating towards the flank of the reservoir where the structural conditions are conducive to upward movement. This suggests a more pronounced rock expansion in these shallower areas of the reservoir, with less overburden stress facilitating uplift. These insights challenge conventional expectations, particularly the intuitive anticipation of the largest displacements occurring at the wellhead location (Tang et al., 2022). They emphasize the importance of a detailed understanding of temporal geomechanical changes for the design and monitoring of CO_2 sequestration sites, highlighting the need for a comprehensive approach to address the implications of these changes on long-term storage integrity.

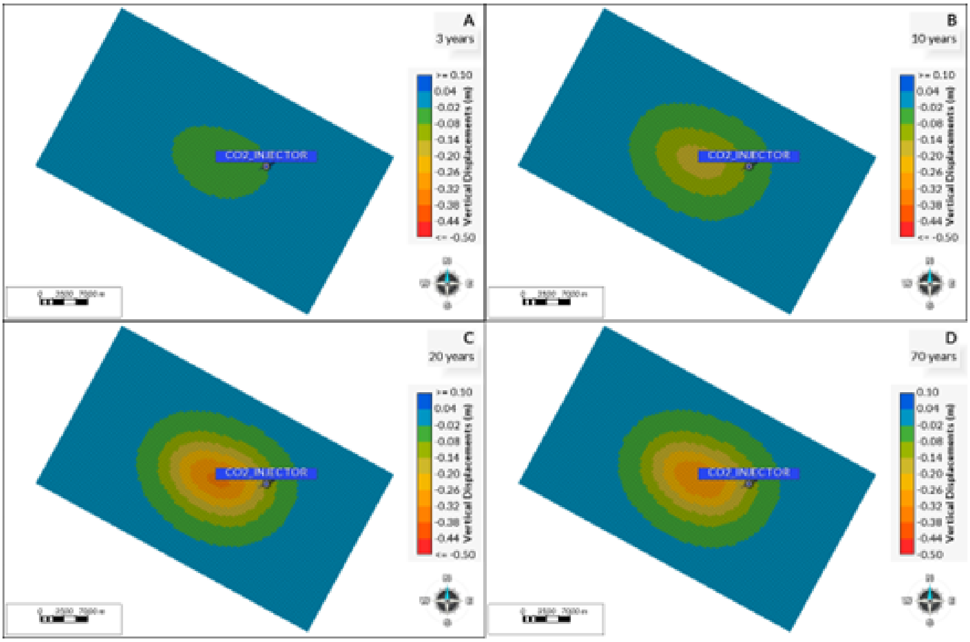


Figure 2.9: Seabed uplift progression over time due to CO_2 injection. Panels (A–D) show the seabed uplift at 3, 10, 20, and 70 years, respectively. The uplift is depicted as negative values of subsidence, signifying upward movement.

The graph in Figure 2.10 showcases vertical displacement at the wellhead (red) and the center of the reservoir (blue) over time. Negative values indicate upward movement. This phenomenon implies that the subsurface system is still dynamically adjusting to the introduced CO_2 long after the injection phase has concluded. The data reveal a persistent ground movement due to pressure dissipation, extending beyond the injection period, which represents significant insight into the behavior of the storage system. It shows that the ground deformation rate is not null after the end of injection and that it can still be used as a source of data to be captured. Thus, their deployment can provide continuous monitoring of the geomechanical state of the CO_2 storage site.

Figure 2.11 illustrates the geomechanical behavior at the end of the 20-year injection period and brings an insight related to the unexpected location of maximum uplift, which, as depicted in the vertical cross-section of the model, is not at the wellhead but rather closer to the center of the reservoir. This area, being shallower, is subject to less overburden stress, making it more responsive to the pressure changes induced by the

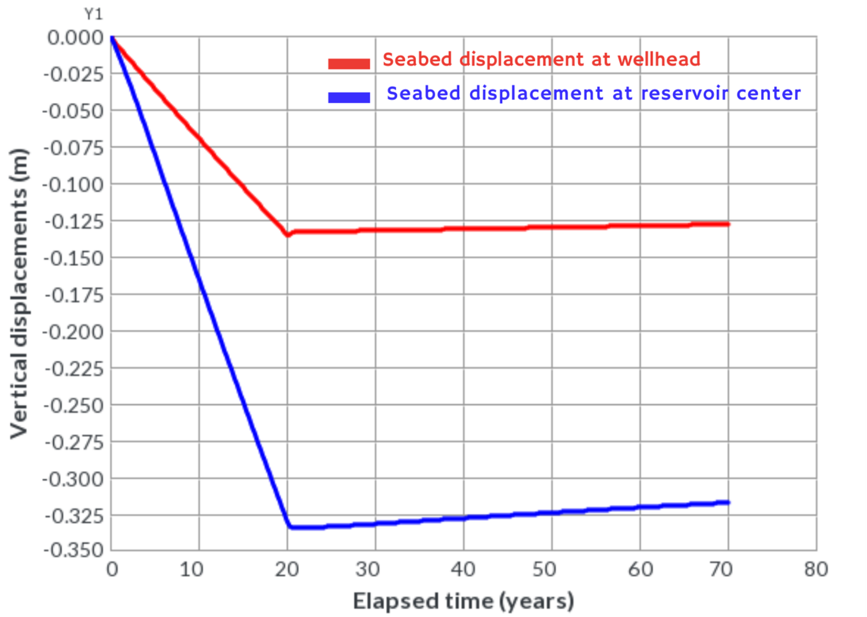


Figure 2.10: Seabed vertical displacement at wellhead (red) and reservoir center (blue) over time. Negative values indicate upward movement.

injection process. The significance of this finding extends beyond mere observation; it highlights the need for precise geomechanical modeling before the implementation of monitoring strategies. While the highest delta pressure remains near the injection point, the most significant uplift occurs in a different location, influenced by the reservoir's structural nuances. This highlights the heterogeneous nature of the reservoir's response to CO_2. Insights like these help with accurately determining the optimal placement of monitoring instruments, such as tiltmeters, within the so-called Area of Review, as described in the Introduction.

To illustrate the dynamic nature of ground displacement associated with CO_2 injection, Figure 2.12 shows the displacement rate (millimeters per year) at various times. It is particularly noteworthy that the highest displacement rate undergoes a significant shift in a relatively short period. Initially, the greatest displacement rate is observed in the vicinity of the CO_2 injector. However, as time progresses, this zone of maximum displacement migrates away from the injector, reaching a point further from the wellhead and above the center of the reservoir in a period of 3 months. This pattern is evident as early as 3 months into the injection process and becomes more pronounced at the 3-year mark. At the injection shutoff, 20 years, the displacement is clearly pronounced, while by the 21st year, there is a significant reduction in displacement velocity, indicating the beginning of stabilization. It is important to point out that after CO_2 injection is stopped (after 20 years), the ground movement near the wellhead drops below 1 mm/year, which might be

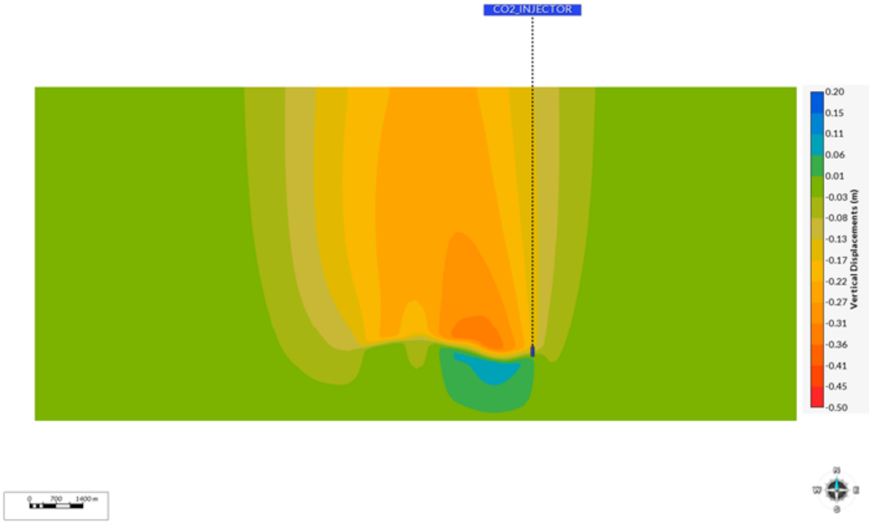


Figure 2.11: Vertical cross-section showing reservoir deformation at 20 years. The color gradient represents the magnitude of uplift, with the CO_2 injector indicated by the dashed line and the point of maximum uplift highlighted in the central region, offset from the injector.

too small for some sensors to detect. Such observations are critical for understanding the temporal and spatial evolution of deformation, thereby aiding in the optimal placement of monitoring equipment. As we conclude this section, we have laid a robust foundation for a simulation framework that not only highlights the insights gained from monitoring ground deformation and pressure dissipation but that also emphasizes the intricacies of CO_2 redistribution within the geological storage. These insights are paramount for the development of effective and secure Carbon Capture and Storage strategies and to further refine our understanding of geomechanical responses to CO_2 storage.

Moving forward, this study will embark on a series of sensitivity analyses. This next phase is designed to evaluate how variations in parameters—such as the mechanical properties of rocks, caprock permeability, and CO_2 injection rates—impact our findings. The sensitivity analyses aim to further refine our understanding of geomechanical responses to CO_2 storage, ensuring that the conclusions drawn are robust and applicable across a range of scenarios. This will enhance the predictive accuracy of our models, contributing to the optimization of CO_2 sequestration practices and the advancement of monitoring technologies.

2.5.3. Sensitivity analysis

The forthcoming phase of this study will center on a sensitivity analysis to elucidate the effects of variations in rock mechanical properties, caprock permeability, and CO_2 injection rates on the geomechanical behavior associated with GCS. Detailed ranges of the parameters under consideration for the uniform distribution of properties are delineated in Table 2.3, and the designated points of interest within the 3D geomechanical model

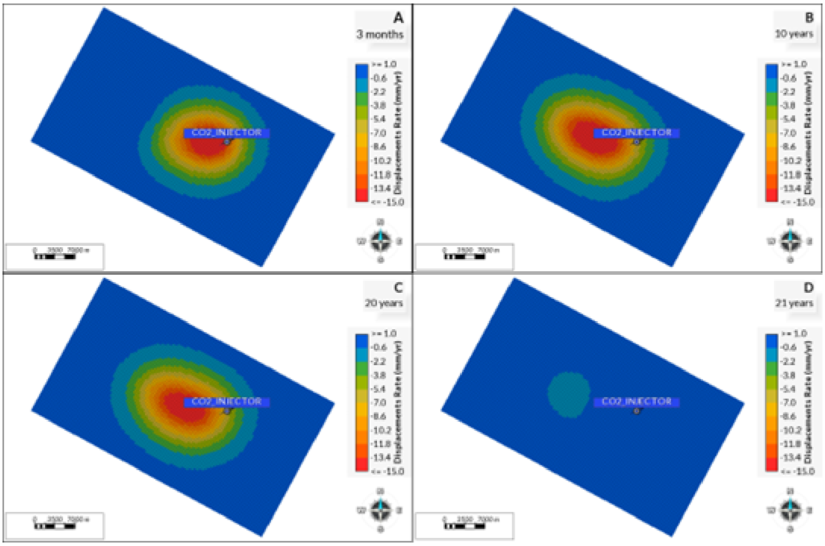


Figure 2.12: Vertical displacement rates (mm/year) at various times. Panels (A–D) show the seabed vertical displacement rates at 3 months, 20 years, and 21 years, respectively.

are highlighted in Figure 2.13. Upon defining the range of parameters for this study, 80 realizations were executed to produce an ensemble of displacement curves, as visualized in Figure 2.14. Panel A corresponds to the path along J = 33, while Panel B represents I = 56. These figures demonstrate a significant dispersion in the displacement outcomes, indicative of the varied responses within the model to changes in CO_2 injection rates. The maximum vertical displacement reaches approximately −0.35 m, while some models barely have it. It is important to note that these displacements occur after the CO_2 injection after 20 years, representing the expected maximum displacement within the modeled timeframe. The spread of the displacement curves emphasizes the importance of considering a wide range of scenarios to adequately capture the complexities and uncertainties inherent in geomechanical behavior during CO_2 storage operations.

Table 2.3: Ranges of the uniform distribution of the parameters used in the sensitivity analysis (Gercek, 2007; Molina et al., 2017).

Property	**Min Value**	**Max Value**
Injection rate (tons/day)	100	3000
Poisson's ratio (sandstone)	0.22	0.39
Poisson's ratio (shale)	0.28	0.48
Young's modulus (sandstone) (GPa)	0.7	34
Young's modulus (shale) (GPa)	6.9	69
Caprock permeability (mD)	10^{-9}	10^{-2}

Upon defining the range of parameters for this study, 80 realizations were executed

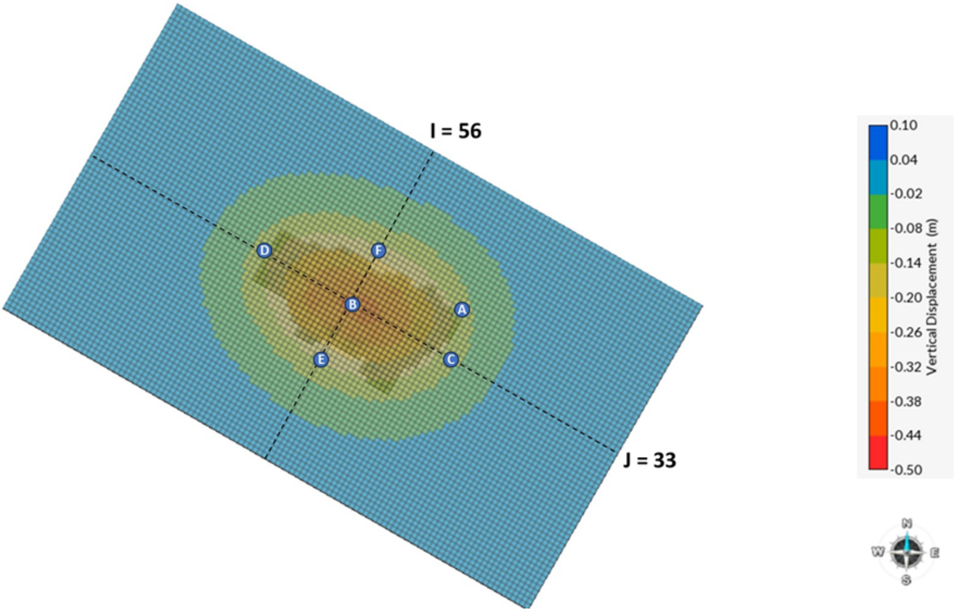


Figure 2.13: Monitoring points of interest for geomechanical sensitivity analysis. The figure displays points A–F, indicating areas of interest against a backdrop depicting vertical displacement. Points A and B represent the wellhead and the central reservoir peak displacements in meters, respectively. The orthogonal I and J grid paths intersect at the reservoir's surface midpoint, situating these points within the model.

to produce an ensemble of displacement curves, as visualized in Figure 2.14. Panel A corresponds to the path along J = 33, while Panel B represents I = 56. These figures demonstrate a significant dispersion in the displacement outcomes, indicative of the varied responses within the model to changes in CO_2 injection rates. The maximum vertical displacement reaches approximately −0.35 m, while some models barely have it. It is important to note that these displacements occur after the CO_2 injection after 20 years, representing the expected maximum displacement within the modeled timeframe. The spread of the displacement curves emphasizes the importance of considering a wide range of scenarios to adequately capture the complexities and uncertainties inherent in geomechanical behavior during CO_2 storage operations.

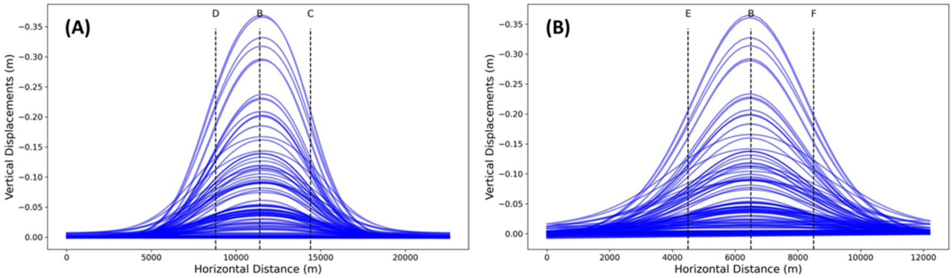


Figure 2.14: Ensemble of vertical displacement curves post-injection. Panel (A) (J = 33) and Panel (B) (I = 56) display the range of vertical displacements resulting from 80 realizations based on varied parameter inputs.

Figure 2.15 presents the sensitivity analysis of vertical displacement curves along path J = 33. Displacement curves are stratified by parameters used in the sensitivity analyses.

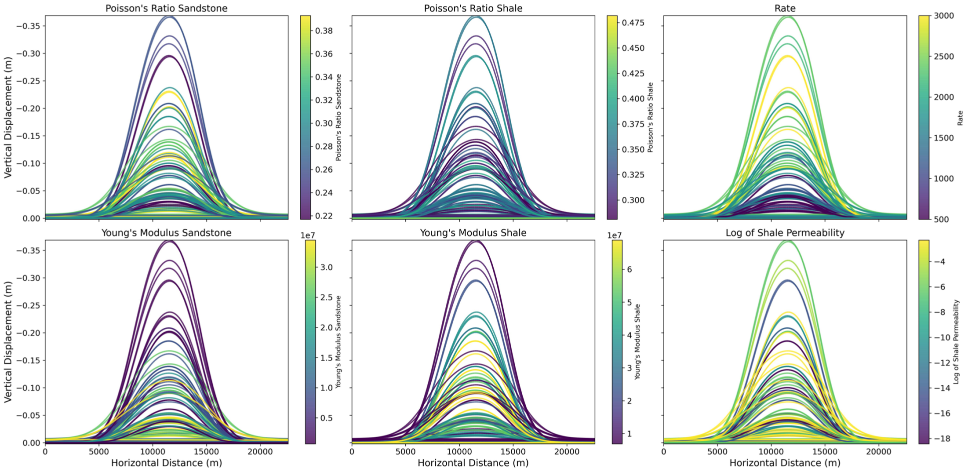


Figure 2.15: Sensitivity analysis of vertical displacement curves along path J = 33. Displacement curves are stratified by parameters used in the sensitivity analyses.

Figure 2.16 integrates the analysis of geomechanical responses' and properties' impacts on vertical displacement. Panel (A) presents box plots of vertical displacements at monitoring points A–F. Panel (B) quantifies the correlation between geomechanical properties and vertical displacement.

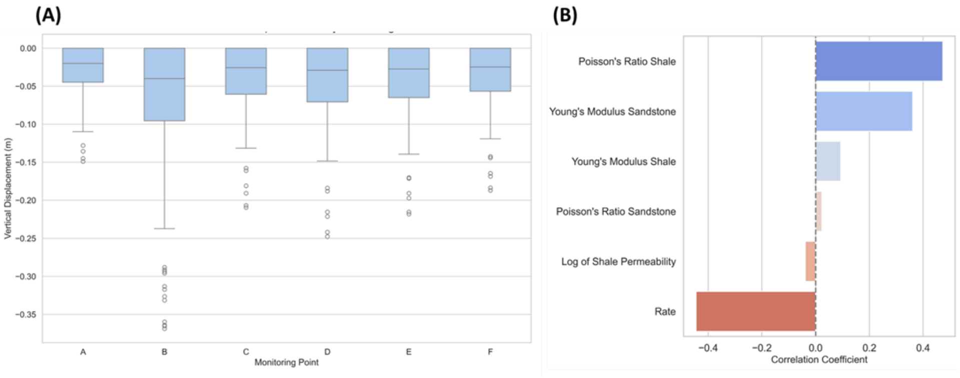


Figure 2.16: Integrated analysis of geomechanical responses' and properties' impacts on vertical displacement. Panel (A) presents box plots of vertical displacements at monitoring points A–F. Panel (B) quantifies the correlation between geomechanical properties and vertical displacement.

Figure 2.17 presents the correlations between geomechanical properties and vertical displacements at individual monitoring points within the reservoir after twenty years of CO_2 injection. Subplots A through F denote these points, indicating localized responses. The correlation coefficients, derived from variables including Young's modulus, Poisson's ratio for sandstone and shale, and CO_2 injection rates, reveal a multifaceted relationship that aligns with the overall behavior patterns identified across the reservoir. This suggests that the geomechanical behavior at single monitoring points reflects the broader, systematic influences of the injection process.

Building on the comprehensive analyses detailed earlier, this study progresses to evaluate the influence of shale permeability on the dissipation of reservoir pressure over time in Figure 2.18. It depicts the interplay between shale permeability, CO_2 injection rates, and their combined effects on reservoir pressure over time. Panel (A) shows the influence of shale permeability on reservoir pressure, and it is observed that except for very low permeability scenarios (10^{-6} mD to 10^{-9} mD), reservoir pressure decreases significantly after the cessation of CO_2 injection. This trend suggests that while shale acts as a semi-permeable barrier, its permeability is crucial in regulating the rate at which pressure normalizes post-injection. For reservoir management, this implies that higher shale permeability can potentially be harnessed to mitigate risks associated with overpressure, provided that the permeability does not compromise the seal's integrity. Conversely, very low permeability could pose challenges in pressure management, possibly leading to an increased risk of induced seismicity or caprock failure. Panel (B) further elaborates on the effect of injection rates on reservoir pressure. It demonstrates that under typical conditions (Hosa et al., 2011), increased injection rates are associated with a proportional rise in average reservoir pressure. However, when juxtaposed with high caprock permeability, the anticipated linear relationship between injection rates and pressure is

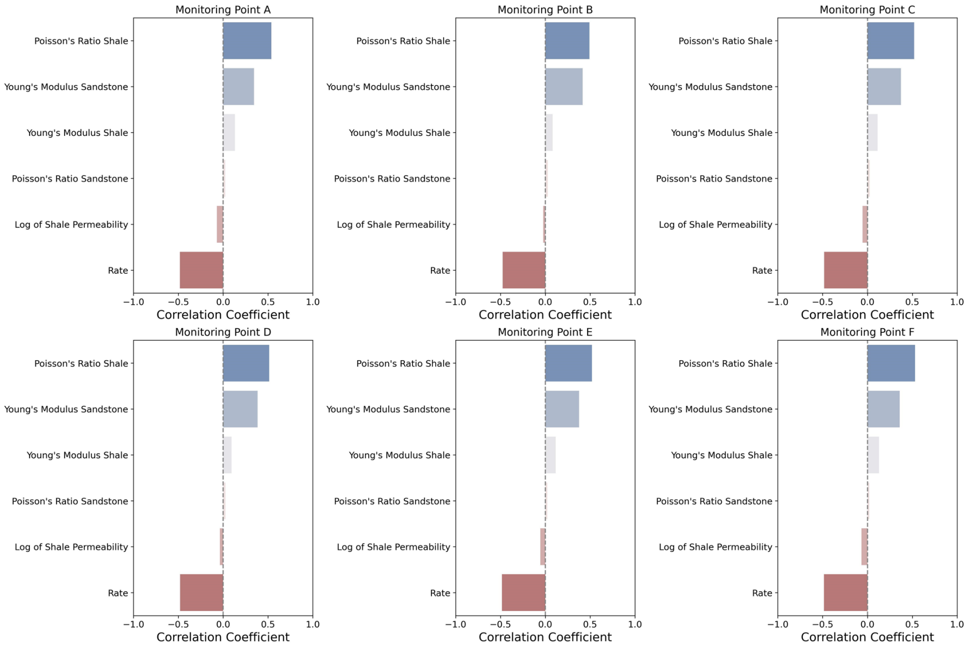


Figure 2.17: Correlation coefficients by monitoring point for geomechanical properties impacting vertical displacement. This figure demonstrates the correlation between geomechanical properties and vertical displacement, disaggregated by monitoring points (A–F).

disrupted. This suggests that at higher permeabilities, the caprock may allow for a more rapid dissipation of pressure, thereby muting the impact of increased injection rates on overall reservoir pressure. From a practical standpoint, Panel (B) suggests that for effective CO_2 storage and containment, both the injection rate and the caprock permeability must be considered. Injection strategies should be tailored to balance the rate of CO_2 being injected with the permeability characteristics of the caprock to maintain desired pressure levels without compromising seal integrity or storage capacity.

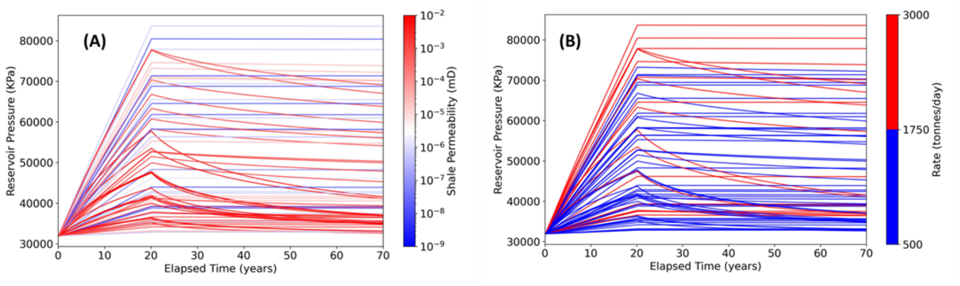


Figure 2.18: Effect of shale permeability and injection rates on reservoir pressure over time. Panel (A) displays the influence of shale permeability on reservoir pressure, and Panel (B) shows the impact of varying CO_2 injection rates on reservoir pressure across different caprock permeabilities.

Addressing the sensitivity in the vertical displacement rates is important because it is a threshold of detectability for several monitoring tools (Bouqueta et al., 2022). Figure 2.19 presents the vertical displacement rates at monitoring points A (wellhead) and B, with a focus on the detectability of such movements. The temporal scope of this figure spans 70 years, comprising both the active injection period and subsequent monitoring phases. A threshold delineates the detectable range of vertical displacements, demarcated as rates exceeding 1 mm/year in absolute value, with the shaded area representing the undetectable range (between −1 and 1 mm/year). The vertical dashed line at 20 years marks the cessation of CO_2 injection, serving as a temporal delimiter for the analysis of displacement behaviors. At monitoring point B, the vertical displacement rates are predominantly higher than at monitoring point A (wellhead). Monitoring point B consistently demonstrates higher vertical displacement rates than point A (wellhead), reaffirming its suitability as a strategic location for enhanced detectability in geomechanical monitoring. This analysis shows the advantage of optimizing the monitoring location, which, for the case study and parameters considered, would be monitoring point B, as it offers a higher probability of yielding detectable displacements across a broader spectrum of realizations. Post-injection, a decline in the vertical displacement rate is evident, with most realizations converging towards undetectable rates, falling within the ±1 mm/year range. This reduction in rate suggests geomechanical stabilization within the storage formation, indicative of pressure redistribution and equilibrium attainment in the subsurface. However, some realizations still present detectable displacement rates for several years of monitoring point B. Interestingly, certain realizations reveal an inversion in displacement trends, accentuating the reservoir's heterogeneity and the caprock's ability to assimilate and reallocate stress over time. This behavior is likely indicative of pressure diffusion effects within the caprock in specific scenarios. It is also probably an effect of the pressure diffusion on the caprock.

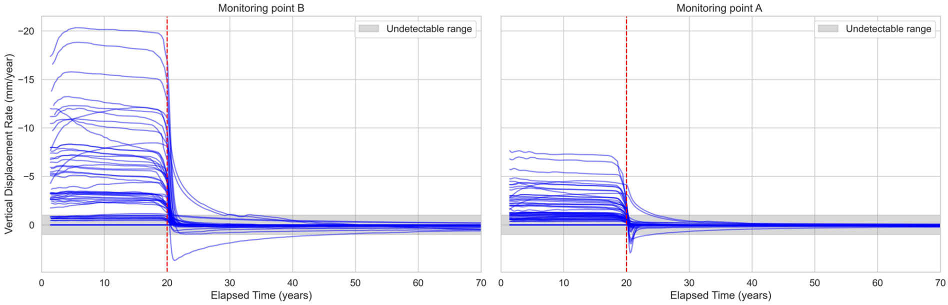


Figure 2.19: Vertical displacement rates at monitoring points A (wellhead) and B. The undetectable range of vertical displacements is shaded, with a vertical dashed line indicating the end of CO_2 injection at 20 years.

Figure 2.20 presents the vertical displacement rates at monitoring points A and B, correlating the injection rates with the resulting vertical displacement rates. The expectation that higher injection rates would result in increased displacement is generally supported; however, the data do not show a simple one-to-one correspondence. The highest displacement rate is not associated with the peak injection rate, suggesting that displacement is influenced by a complex set of reservoir characteristics beyond just the injection rate. At monitoring point B, a range of injection rates, including some of the lower values, still result in displacement rates above the detection threshold. This demonstrates the sensitivity of this point and makes it a reliable location for monitoring detectable geomechanical changes among the other locations considered. In contrast, at point A, the wellhead, most displacement rates fall below the detection threshold, especially at lower injection rates, indicating less sensitivity to changes in CO_2 injection volume at this location.

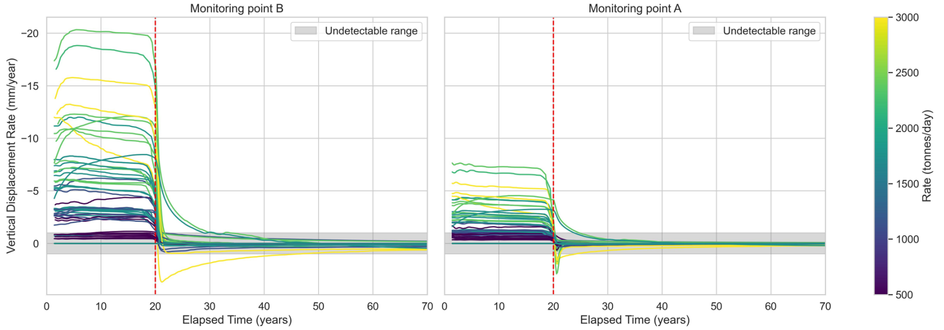


Figure 2.20: Vertical displacement rates at monitoring points A and B correlated with CO_2 injection rates. The undetectable range of vertical displacements is shaded, with a vertical dashed line indicating the end of CO_2 injection at 20 years.

The sensitivity analysis provides insights into the geomechanical impacts of CO_2 storage operations. By examining a spectrum of parameters, this study has revealed significant variability in geomechanical responses, which are important for enhancing the predictive accuracy of CO_2 sequestration models.

2.6. DISCUSSION

The findings presented in this work underscore the need for comprehensive monitoring and robust calibration of models to effectively manage the geomechanical risks associated with CO_2 injection. The interdependencies between rock mechanical properties and operational parameters revealed through this analysis serve as a foundation for future research aimed at optimizing CO_2 storage strategies and ensuring the integrity of geological storage sites.

Expanding upon previous work on the positioning of a CO_2 injection well for GCS (Kuk et al., 2020; Shchipanov et al., 2022), our study discusses the impact of this placement of the well in a flank of the reservoir on the vertical displacements caused by the injection as a subsequent impact on the monitoring strategy. Adding to the discourse on geomechanical risks associated with GCS initiated by (Lyu & Voskov, 2023; Spiteri et al., 2005; Xu et al., 2014), our research emphasizes the importance of conducting sensitivity analyses on geomechanical modeling for the strategic selection of monitoring locations. This approach builds upon the insights from (Nagel, 2001; Song et al., 2023; Wouters et al., 2021), who advocated for the need for accurate characterization of geomechanical properties and injection rates. Our study not only corroborates these findings but also advances the methodology by demonstrating how such analyses can optimize monitoring strategies, thereby enhancing the predictive accuracy and safety of GCS operations.

Furthermore, we discuss the importance of explicitly modeling the flow within the caprock, a critical element often not addressed in coupled geomechanics studies that could help evaluate the true impact of geomechanics on the caprock. This highlights the impact of incorporating flow modeling in caprock analysis to ensure a comprehensive understanding of CO_2 storage safety and efficacy. Moreover, our research sheds light on the interplay between injection rates, rock properties, and pressure buildup, a complex relationship that has been explored to a lesser extent in the literature. By providing a detailed analysis of how these factors influence CO_2 migration and ground deformation, our study offers a novel perspective that challenges the conventional selection of wellhead monitoring sites, as typically suggested by previous studies, like (Bouqueta et al., 2022). We reveal the limitations of traditional monitoring approaches in capturing the nuanced geomechanical responses to CO_2 injection, suggesting a reevaluation of monitoring strategies to better manage the associated risks.

It is important to consider the limitations of this study to ensure a nuanced understanding of our findings and to accurately direct future avenues of research. One consideration is the foundation of our analysis on simulations derived from a singular reservoir model. While providing insights into the specific geological configuration examined, this approach may not encompass the diversity found in various geological settings where GCS is applicable. The results, while enlightening for the studied scenario, may not be broadly generalizable across all possible configurations of GCS. For reservoirs in which permeability is greatly affected by the changes in stress during injection, like naturally fractured rocks, the methodology could benefit from the integration of consistent coupling mechanisms (March et al., 2018). The inclusion of chemical reactions in our modeling efforts could offer a deeper understanding of how these processes influence the geomechanical properties of subsurface formations, potentially unveiling new insights into the long-term stability and integrity of CO_2 storage sites.

Further investigation using fine-scale models is recommended to delve deeper into the mechanisms of enhanced CO_2 dissolution. The use of a finer grid in this model can provide a more detailed and accurate representation of the dissolution and diffusion processes. By conducting such investigations, a more comprehensive understanding of CO_2 entrapment in the caprock can be gained, thus enhancing the overall analysis and the findings of this study. While our simulations provide insights into GCS, the lack of field verification is a significant limitation. Theoretical models, no matter how detailed, need empirical data to validate their predictions and enhance their credibility. Future efforts should focus on integrating field data from similar GCS projects into the theoretical models and compare their predictions with actual conditions. This integration would allow for the calibration of our model based on real-world outcomes, thus enhancing its predictive power and reliability.

Although our model is based on simulations derived from a singular reservoir model, it offers a methodological framework that could be adapted for broader use. We recommend further investigation using field data from multiple geological settings to test the adaptability of our approach. Incorporating multi-scale data and chemical reactions into our models could deepen our understanding of subsurface dynamics, thus aiding in the design of more effective CO_2 storage solutions. In addition to mechanical properties and CO_2 injection rates, our focus in this study, we recognize the potential influence of considering a broader range of parameters in the sensitivity analyses, such as permeability, porosity, relative permeabilities, and fluid properties, and their impacts on the outcomes of CO_2 injection scenarios. These parameters play important roles in the dynamics of GCS, but they were not extensively analyzed due to their vast variability and complex interaction effects. Including a comprehensive sensitivity analysis incorporating these hydrodynamic and fluid properties would have expanded the scope and increased the complexity of our study significantly. Therefore, while our focused approach provided essential insights into the geomechanical impacts of CO_2 injection, future studies could benefit from an expanded sensitivity analysis that includes these additional variables.

Addressing these limitations not only highlights areas ripe for further research but also emphasizes the need for ongoing development and validation of geomechanical models. Such endeavors are paramount for advancing our collective understanding of GCS's safety and efficacy, ultimately aiding in the broader acceptance and implementation of GCS technologies as viable solutions for carbon management across a spectrum of geological scenarios.

2.7. CONCLUSIONS

This study demonstrates the technical feasibility of using a comprehensive numerical simulation framework that combines fluid flow and geomechanics to model ground deformation caused by CO_2 injection for storage. Although the study employed a specific geological model, the conducted sensitivity analysis enables the extension and generalization of the findings, offering important insights into geomechanical considerations. These findings serve as information for guiding future research and experimentation in this area. The main conclusions derived from this study are outlined below.

1. Pressure diffusion into adjacent rocks can occur even without significant CO_2 migration due to the nature of pressure transmission through pore fluids. This precedes the physical movement of CO_2.

2. The presence of small amounts of CO_2 detectable in the surrounding rocks, mainly through diffusion, does not pose a leakage concern in this model, assuming the absence of fractures or other flow localization features.

3. The limited movement of CO_2 into the surrounding rocks demonstrates the combined effect of the caprock's low permeability and the high capillary entry pressure, which serve as a robust containment mechanism ensuring the long-term integrity of the geological storage.

4. Further investigation using fine-scale models is recommended to delve deeper into the mechanisms of CO_2 entrapment due to diffusion. The use of finer gridblocks in this model can provide a more detailed and accurate representation of the dissolution and diffusion processes.

5. The study findings indicate that there is persistent ground movement resulting from pressure dissipation, which continues even after the injection period. This validates the use of high-precision instruments, such as floor tiltmeters, for monitoring CCS operations, as they are capable of accurately measuring the displacement rates associated with these activities.

6. The sensitivity analysis highlights the importance of comprehensive monitoring and robust calibration of these models to effectively manage geomechanical risks associated with CO_2 injection. The findings suggest that rocks with higher Poisson ratios and lower Young's modulus values experience increased vertical displacements. Additionally, higher CO_2 injection rates and reduced shale permeabilities are associated with greater displacement magnitudes. This interplay between rock mechanical properties and operational parameters provides a basis for future research aimed at optimizing CO_2 storage strategies and ensuring the integrity of geological storage sites.

DATA AVAILABILITY STATEMENT

The data used in this study are provided within the article, and the UNISIM-I geological model is openly available at https://www.unisim.cepetro.unicamp.br/benchmarks/en/ (accessed on 12 December 2023).

3

INTEGRATING GEOMECHANICAL PROXY MODELS WITH DATA ASSIMILATION FOR ENERGY TRANSITION APPLICATIONS

This chapter develops a framework that integrates physics-based geomechanical proxy models with ensemble smoother data assimilation (ESMDA) to quantify subsurface uncertainties in CO_2 *storage. The methodology combines a computationally efficient nucleus-of-strain proxy model with the open-DARTS simulator to assimilate surface deformation measurements, addressing the spatial limitations of traditional well data. Testing on both conceptual and realistic datasets demonstrates uncertainty reduction: 53.1% for a 2D conceptual case and 18.5% for the complex 3D Brugge field. Entropy-based metrics quantify these improvements, validating the approach's effectiveness for constraining permeability distributions and* CO_2 *plume predictions through surface monitoring data.*

Parts of this chapter have been published in Ilshat Saifullin, Gabriel Serrão Seabra, Anne Pluymakers, Femke C. Vossepoel, Denis Voskov, Integrating Geomechanical Proxy Models with Data Assimilation for Energy Transition Applications, *Geomechanics for Energy and the Environment* **40** (2024) 100618 (Saifullin et al., 2024). Changes have been applied to make the text and figures consistent with the thesis.

3.1. INTRODUCTION

The need to manage carbon emissions calls for efficient and safe utilization of subsurface resources. The strategies needed to achieve such complexity are outlined with the scenarios by international organizations such as the International Energy Agency (IEA), pointing to the needed marked difference in sources of energy generation to achieve the set levels of greenhouse gas reductions. As IEA (2022) point out in their "World Energy Outlook 2022" report, Geological Carbon Storage (GCS) has been singled out as one of the core enabling technologies with the substantive potential for CO_2 emissions reduction in a fossil fuel-constrained scenario. Exploiting subsurface resources for CO_2 storage involves uncertainties in properties such as permeability and porosity, which impact project feasibility, operational efficiency, and environmental safety. These uncertainties necessitate advanced methods for risk management (March et al., 2018; Witter et al., 2019).

The accuracy requirements for subsurface characterization vary by project, based on local economic, regulatory, and geological factors. For CO_2 sequestration projects, monitoring CO_2 concentrations near sensitive areas like faults or legacy wells is required to meet regulatory standards and assess leakage risks. Additionally, pressure and temperature variations near existing faults must be estimated for induced seismicity risk assessment. CO_2 storage operations often lack comprehensive monitoring systems, and integrating new measurement techniques can supplement production data, improving both risk mitigation and performance. As we discuss in this work, surface displacement data can provide insights into reservoir behavior and help mitigate project risks (Ringrose et al., 2013; Shi et al., 2019). Modern instruments like tiltmeters enable high-precision measurements of these displacements at sub-millimeter scale (Davis et al., 2000). However, incorporating such data requires advanced geomechanical modeling, which poses challenges due to the complexity of rock-fluid interactions, limited calibration data, and the computational demands of fully-coupled simulations.

To address these challenges, researchers have developed surrogate geomechanical models, including neural networks (Tang et al., 2022) and physics-based proxies combining analytical and numerical methods. van Wees et al. (2019) and Paullo Muñoz and Roehl (2017) introduced methods using Green's function and the nucleus of strain concept to model stress responses and predict compaction and subsidence. These approaches, validated against finite element models, have been applied to real-world challenges like assessing induced seismicity in the Groningen gas field (Candela et al., 2022b).

We have enhanced these models by incorporating thermoelastic responses, developing a framework for efficient displacement evaluation, and adapting to non-rectangular grids (e.g., Corner Point Grid). By integrating our model with the open-DARTS simulator (Denis Voskov et al., 2025; Voskov et al., 2024), we aim to refine physics-based geomechanical proxy models for CO_2 storage applications. These computationally efficient proxies enable advanced data assimilation techniques that would otherwise be prohibitively expensive, addressing the challenges of subsurface modeling and uncertainty quantification in CO_2 storage applications.

Data-assimilation methods, including ensemble-based strategies like EnKF and ES-MDA (Emerick & Reynolds, 2013b), and gradient-based methods such as RML (Tian et al., 2024) and 4DVar (Courtier et al., 1994; Evensen et al., 2022; Oliver et al., 2008; Tarantola, 2005), integrate observational data with model realizations to refine numerical model accuracy. ESMDA has been successfully applied in various subsurface applications (Candela et al., 2022a; Gazzola et al., 2021; Tang et al., 2022) and its sampling performance has been investigated in simplified settings (Emerick & Reynolds, 2013a). This method's iterative nature, coupled with the use of an inflated covariance matrix, allows for more accurate and robust updates to the model, even with the sparse and diverse data characteristic of subsurface modeling challenges (Emerick & Reynolds, 2013a).

We incorporate ESMDA to reduce uncertainty in reservoir permeability, enhancing understanding of the variables in CO_2 storage projects. While vertical displacements may be subject to uncertainties in rock geomechanical properties, the comparative behavior of models remains important for estimating thermoporoelastic responses. Our framework is tested on a 2D conceptual dataset with Gaussian distributed permeability, followed by application to the more complex Brugge Dataset (Peters et al., 2010). While there is no standardized metric for uncertainty reduction in CO_2 storage applications, our approach demonstrates improvements in subsurface characterization. Using entropy as a measure of uncertainty, we achieved reductions of up to 53.1% for CO_2 storage applications. These improvements enhance decision-making processes and risk assessment in subsurface resource management initiatives. This sequence evaluates the model's performance from simplified contexts to real-world simulations, demonstrating its applicability for decision-making in CO_2 storage projects. In summary, the main contributions of this work include the development and application of a physics-based geomechanical proxy model, which is integrated with ESMDA.

This study also enhances the monitoring and management of CO_2 injection through the data assimilation of surface displacement measurements into subsurface models. This method has been effectively applied to both a conceptual 2D dataset and a realistic scenario inspired by the Brugge dataset. This paper covers the methodology, including the geomechanical proxy model development and ESMDA framework, followed by test cases for CO_2 injection models. Results are presented for both conceptual and Brugge dataset grids, with a final discussion synthesizing the findings and their implications for CO_2 storage applications.

3.2. METHODOLOGY

3.2.1. HYDRODYNAMIC MODEL

For the hydrodynamic simulation part of our methodology, we used open Delft Advanced Research Terra Simulator (open-DARTS). open-DARTS is an open-source reservoir simulation software with a focus on energy transition applications. Within this framework, we modeled CO_2 injection scenario employing compositional thermal physics, described in Section 3.3. Source code and examples are available at the open-DARTS repository (open-DARTS Development Team, 2024). In our study we leverage the Operator-Based Linearization (OBL) interpolation caching feature of open-DARTS simulator. An OBL approach exploited in open-DARTS replaces the computation of fluid and rock properties at a particular state (pressure, temperature and composition) with the construction of parameter space and approximation of PDE operators in it using multi-linear interpolation (Voskov, 2017). The advantage of the OBL approach is that once the operator's space has been constructed and dynamically filled with the computed operators, it can be further reused, thereby diminishing the time required for Jacobian evaluations. Moreover, this feature facilitates the recycling of computation points from previous simulations when conducting subsequent model simulations within the same ensemble, and across all iterations of the ESMDA. This approach reduces the computational cost associated with running extensive model simulations. The source/sink is implemented in open-DARTS as an additional term in the PDEs, which is non-zero for the reservoir blocks with well perforations. The inflow/outflow is proportional to so called well index, computed by the commonly used Peaceman's formula and the pressure difference between the perforated reservoir block and the bottom hole pressure defined at the well. This approach does not require mesh refinement, which would increase computational time. Instead, we add a source/sink term in the PDEs for the reservoir blocks containing well perforations. The well index, which determines the pressure change in the cell with a well perforation for a rate-controlled well, is computed using Peaceman's formula. This index depends on the well radius, grid cell size, and permeability. Notice that for the transient analysis, the refined mesh should be implemented around the well (Li et al., 2001). One of the recently implemented features of open-DARTS is a fully-coupled geomechanical modeling using Finite Volume discretization Method with a multi-point flux and stresses approximations (Novikov, 2024; Novikov et al., 2022). Although the coupled solver allows us to neglect some assumptions we have made in this study and resolves the geomechanical state precisely, in order to preserve performance, we used a proxy geomechanical model described below. The mathematical formulation and the solution approach are described in the Appendix.

3.2.2. GEOMECHANICAL PROXY MODEL

In this Section, we describe a developed numerical methodology of thermoporoelastic rock response approximation based on the nucleus of strain approach. This model, referred to as a geomechanical proxy model, calculates the poroelastic and thermoelastic displacements at the surface for the requested points at the surface using pressure and temperature changes from a hydrodynamic simulator. One of the advantages of this approach is that evaluation points do not have to be at cell centers or nodes of the original grid as in fully-coupled geomechanical simulators. The nucleus of strain concept allows the computation of displacements at any location, even outside the reservoir grid. Thus, the model can provide the values of the vertical displacements at the surface without constructing a computational grid up to the surface. The inputs of the proxy model are: cells geometry, pressure and temperature changes at each cell and geomechanical properties (Poisson ratio, Young modulus, and thermal expansion coefficient). Note that this approach has the limitation that geomechanical rock properties should be homogeneous (also in the surroundings of the reservoir) as it has been derived from an analytical solution. The poroelastic displacement potential at evaluation point formulated by Geertsma (1973) is

$$\Phi_{\text{disp}} = -\frac{C_m V \Delta p}{4\pi R}, \tag{3.1}$$

where R is the distance from the evaluation point to the nucleus, V is the volume of the nucleus, and Δp is the pressure change and C_m - uniaxial compaction coefficient [1/MPa] defined as

$$C_m = \frac{(1+\nu)(1-2\nu)}{E(1-\nu)}. \tag{3.2}$$

The gradient potential gives the solution for the displacements (Tempone et al., 2010):

$$\begin{pmatrix} u_x^p \\ u_y^p \\ u_z^p \end{pmatrix} = \begin{pmatrix} \frac{\partial \Phi_{\text{disp}}}{\partial x} \\ \frac{\partial \Phi_{\text{disp}}}{\partial y} \\ \frac{\partial \Phi_{\text{disp}}}{\partial z} \end{pmatrix} \tag{3.3}$$

Assuming there are N grid cells (nucleus of strain) with a pressure change $\Delta p_j = p_{initial} - p_{current}$ [MPa], and evaluation points with index i having coordinates (x_i, y_i, z_i), the analytic formulas for the poroelastic displacements for an infinite domain are derived using Geertsma (1973).

This nucleus-of-strain concept is extended to a half-infinite space using so called image nucleus (Segall, 1992) leading to additional terms which will be indicated further. The image nucleus is schematically illustrated in Figure 3.1 and has been derived using a condition that the stresses acting through the free surface (the horizontal plane $z = 0$) are nullified (Tempone et al., 2010).

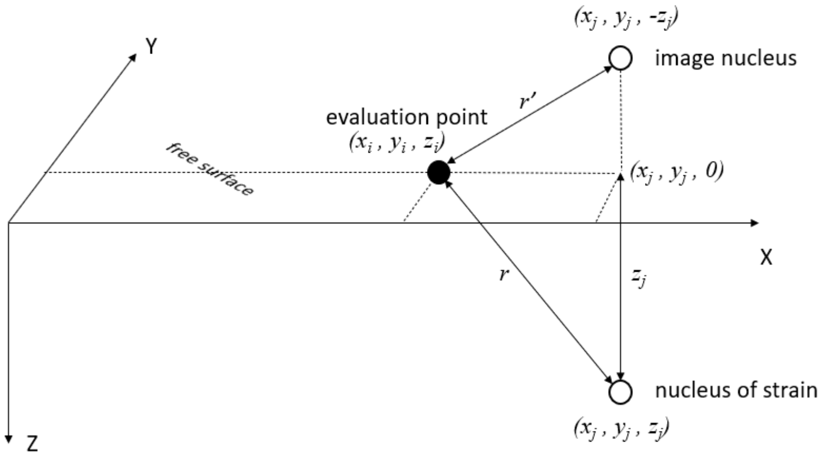


Figure 3.1: The image nucleus illustration for the semi-infinite reservoir.

The poroelastic displacements at point i are evaluated as a superposition of displacement terms that represent the influence of all nuclei of strain, which can be represented by the grid cells:

$$\check{\mathrm{N}} \begin{pmatrix} u_x^p \\ u_y^p \\ u_z^p \end{pmatrix}_i = -\frac{C_m}{4\pi} \sum_{j=1}^{N} \Delta p_j \begin{pmatrix} G_{\text{disp},x} \\ G_{\text{disp},y} \\ G_{\text{disp},z} \end{pmatrix}_{ij}, \tag{3.4}$$

with the geometrical part

$$\begin{pmatrix} G_{\text{disp},x} \\ G_{\text{disp},y} \\ G_{\text{disp},z} \end{pmatrix}_{ij} = \int_{V_j} \begin{pmatrix} \frac{\partial}{\partial x} 1/r \\ \frac{\partial}{\partial y} 1/r \\ \frac{\partial}{\partial z} 1/r \end{pmatrix}_{ij} dv_j + C_\nu \int_{V_j} \begin{pmatrix} \frac{\partial}{\partial x} 1/r' \\ \frac{\partial}{\partial y} 1/r' \\ -\frac{\partial}{\partial z} 1/r' \end{pmatrix}_{ij} dv_j + 2z_j \int_{V_j} \begin{pmatrix} \frac{\partial^2}{\partial x \partial z} 1/r' \\ \frac{\partial^2}{\partial y \partial z} 1/r' \\ \frac{\partial^2}{\partial z^2} 1/r' \end{pmatrix}_{ij} dv_j, \tag{3.5}$$

where the first term represents the infinite-space solution using Equation 3.1 and Equation 3.3, whereas the next two terms are the semi-infinite space corrections (Barbosa et al., 2022), r - the distance vector between the evaluation point i and j-th nucleus of strain, i.e. the j-th prism center, z_j - the depth of j-th prism, dv_j - infinitesimal element of volume located in j-th nucleus of strain, r' - the distance vector between the evaluation point i and j-th image nucleus of strain, $C_\nu = 3 - 4\nu$.

Similarly to equation 3.4, the thermoelastic displacements at point i can be computed:

$$\begin{pmatrix} u_x^T \\ u_y^T \\ u_z^T \end{pmatrix}_i = -\frac{\alpha_{\text{therm}}}{4\pi} \sum_{j=1}^{N} \Delta T_j \begin{pmatrix} G_{\text{disp},x} \\ G_{\text{disp},y} \\ G_{\text{disp},z} \end{pmatrix}_{ij}, \tag{3.6}$$

where ΔT_j [°C] is the temperature change and α_{therm} [1/°C] is a linear thermal expansion coefficient.

Finally, thermoporoelastic displacements are the sums of poroelastic and thermoelastic displacements:

$$u = u^p + u^T. \tag{3.7}$$

Barbosa et al. (2022) and Paullo Muñoz and Roehl (2017) present validations of the described geomechanical proxy model method for the poroelasticity problem, comparing the solution in elastic homogeneous cylindrical reservoir with Geertsma's method for the uniform pressure depletion. Paullo Muñoz and Roehl (2017) also validates the approach against Finite Element Method solution for the non-uniform pressure depletion and the arbitrary shaped reservoir.

In this formulation, the poroelastic and thermoelastic displacements have a common geometric term, which has been utilized in the implemented code.

It is worth emphasizing that this model has intrinsic mechanical boundary conditions, such that the displacements at boundaries are expected to be non-zero and the resulting vertical displacements are expected to be asymmetric on the top and bottom boundaries of the reservoir grid.

The flow chart of the proxy model is shown in Figure 3.2, where the inputs are in blue boxes, and the outputs are in red boxes.

The displacement evaluation is implemented in C++ programming language and parallelized using OpenMP. Parallelization is straightforward, as the evaluation of the displacement at different locations is independent. A Python interface is created using the pybind11 open-source library, and the geomechanical proxy model is integrated into the open-DARTS hydrodynamic model and ESMDA using Python.

In our implementation, the proxy model is linked with the hydrodynamic simulator at runtime with displacement computation at desired time steps. However, it can also be applied as a post-processing tool.

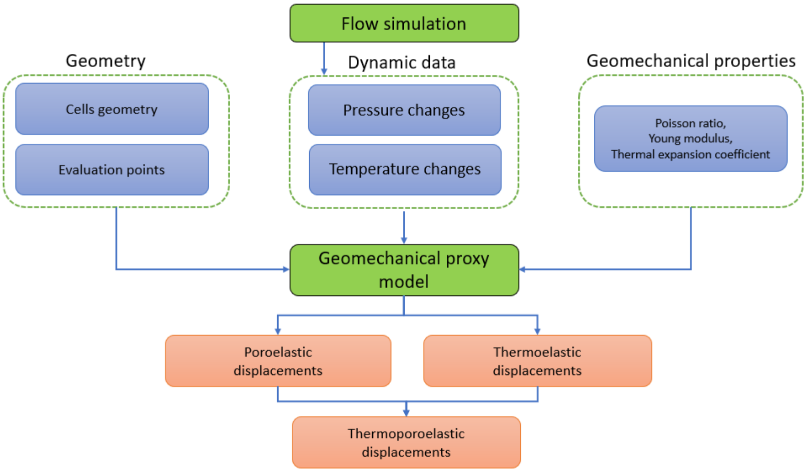


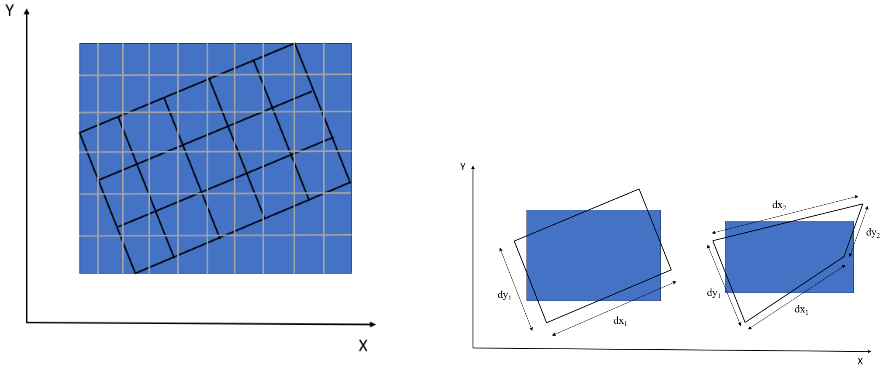
Figure 3.2: Scheme of the geomechanical proxy model of displacements evaluation.

An analytic expression for the volumetric integrals in the Equation 3.5 has been derived by Nagy et al. (2000) for the grid aligned to XYZ-axes and consisting of cells of a rectangular prism shape.

In real-world applications, grids usually consist of complex cell geometries and they are specifically oriented. For such cases, we implemented and tested two approaches of the proxy model application. In the first approach, we construct an auxiliary structured grid, which fulfills the requirements and interpolates the pressure changes Δp and the temperature changes ΔT to it. The auxiliary grid is then used in the geomechanical proxy model. This is illustrated schematically in Figure 3.3a, where the original grid is shown with black lines and the geomechanical grid with white lines. The disadvantage of the interpolation approach is that a structured grid of rectangular prisms should be refined enough to approximate the original mesh.

Another way of different cell shape and orientation treatment is the construction of a rectangular prism with averaged cell sizes of the original cell and oriented to the XYZ-axes. The idea for a 2D case is shown in Figure 3.3b which extension to 3D is straightforward. The cells with dashed black borders are cells of the original grid and the blue cells are the rectangular prisms with geometrically averaged sizes which can be used in the geomechanical proxy model. The case on the left side represents a rectangular prism, which violates the axes' orientation condition. The case on the right side shows a cell of the original grid which violates both shape and orientation conditions.

The presented proxy method assumes the homogeneous and isotropic rock geomechanical properties and linear poroelastic medium. Since the method explicitly evaluates the displacements from the pressure and temperature changes, it ignores the geomechanical effects such as porothermoelastic diffusion. Thus, it does not capture the time-dependent geomechanical response, e.g. the consolidation effect. We also did not consider the stress-dependent behavior of the rock geomechanical properties in this study. Nevertheless, the proxy method can be a good alternative for preliminary



(a) Original grid for flow simulation and auxiliary grid for the geomechanical proxy model (interpolation)

(b) Original grid cell and averaged rectangular prism for geomechanical proxy method

Figure 3.3: Two approaches of a proxy model application to CPG grid: (a) interpolation, (b) averaging

evaluations in complex problems, including heterogeneous reservoirs using averaged homogeneous rock geomechanical properties.

3.2.3. PRIOR DISTRIBUTION OF RESERVOIR PROPERTIES FOR DATA ASSIMILATION

We generate prior distributions for the logarithm of reservoir permeability, a common practice in uncertainty quantification studies (Emerick & Reynolds, 2013b; Remy et al., 2009; Tang et al., 2022). The spatial field for the logarithm of permeability, $\log k(x,y)$, is represented as a Gaussian distribution:

$$\log k(x,y) \sim \mathcal{N}(\mu, \mathbf{C}_{zz}),$$

where μ represents the mean of the logarithm of permeability, and $\mathbf{C}_{zz}$ is the covariance matrix derived from the specified spatial correlation function. The correlation between the values at two points (x_0, y_0) and (x_1, y_1) on the grid is modeled using a modified Gaspari-Cohn function, providing a smooth correlation (Gaspari & Cohn, 1999):

$$C(x_0,y_0,x_1,y_1) = \begin{cases} -\frac{h^5}{4} + \frac{h^4}{2} + \frac{5h^3}{8} - \frac{5h^2}{3} + 1 & \text{for } h < 1, \\ \frac{h^5}{12} - \frac{h^4}{2} + \frac{5h^3}{8} + \frac{5h^2}{3} - 5h + 4 - \frac{2}{3h} & \text{for } 1 \le h < 2, \\ 0 & \text{for } h \ge 2. \end{cases}$$

In this equation, h is the normalized distance scaled by correlation lengths L_x and L_y, and rotated by angle θ_{rot} to align with principal axes of anisotropy:

$$h = \sqrt{\left(\frac{\Delta x'}{L_x}\right)^2 + \left(\frac{\Delta y'}{L_y}\right)^2},$$

$$\Delta x' = \cos(\theta_{\text{rot}})\Delta x - \sin(\theta_{\text{rot}})\Delta y, \quad \Delta y' = \sin(\theta_{\text{rot}})\Delta x + \cos(\theta_{\text{rot}})\Delta y.$$

This covariance structure was discretized over the grid to form the matrix $\mathbf{C}_{zz}$, and the Cholesky decomposition of $\mathbf{C}_{zz}$ was used to simulate spatially correlated random fields that form the ensemble of log-permeability fields:

$$m_i = \mu + L_{\text{chol}} r_i,$$

where L_{chol} is the Cholesky factor of $\mathbf{C}_{zz}$, and r_i is a vector of standard normal random variables.

The implementation involves generating a grid-based model domain, computing the covariance matrix using the Gaspari-Cohn correlation function, and using Cholesky decomposition to generate multiple realizations of the log-permeability field. This approach ensures that the simulated fields respect specified statistical properties, aligning with geostatistical practices as illustrated in SGEMS and GSLIB user guides (Remy et al., 2009; Ziegel, 1995).

3.2.4. DATA ASSIMILATION FRAMEWORK WITH ESMDA

The ESMDA technique is applied to assimilate observations into open-DARTS reservoir model realizations, using the physics-based geomechanical proxy model to account for vertical displacements computation. This method is an excellent choice for this problem, as it offers flexibility by requiring minimal alterations to the existing forward model code and avoiding the computation of adjoints gradients (Evensen et al., 2022). ESMDA performs data assimilation through iterative assimilation of the same observations, adjusting the influence of the observations at every assimilation step to enhance model updates. This adaptation allows taking into account non-linearities in the forward model through its step-wise approach (Emerick & Reynolds, 2013a, 2013b), mainly vertical displacements measured at the surface.

We can describe the analysis for each ensemble member's model parameter $\boldsymbol{z}_j^a$, in our case, the reservoir model permeabilities vector $\boldsymbol{k}_j^a$ as follows:

$$\boldsymbol{z}_j^a = \boldsymbol{z}_j^f + \mathbf{C}_{zd}^f \left(\mathbf{C}_{dd}^f + \alpha_i \mathbf{C}_{dd}\right)^{-1} \left(\boldsymbol{d}_j - G(\boldsymbol{z}_j)^f\right), \quad \text{for } j = 1, 2, \ldots, N_e. \tag{3.8}$$

Here N_e is the total number of ensemble members, $\boldsymbol{z}_j^a$ and $\boldsymbol{z}_j^f$ represent the analyzed and forecasted permeabilities of the j^{th} ensemble member, respectively, $\mathbf{C}_{zd}^f$ and $\mathbf{C}_{dd}^f$ are the cross- and auto-covariance matrices of the model permeabilities and vertical displacements observations in the forecast step, α is a scaling factor, and $\boldsymbol{d}_j$ and $G(\boldsymbol{z}_j)^f$ are the perturbed and forecasted observations for the j^{th} ensemble member. Besides, $G()$ represents the forward model. For more details, see Equation 3.9:

$$\boldsymbol{d}_j = \boldsymbol{d}_{\text{obs}} + \sqrt{\alpha}\mathbf{C}_{dd}^{1/2} r. \tag{3.9}$$

In this equation, $\boldsymbol{d}_{\text{obs}}$ is the vertical displacements observed data, $\mathbf{C}_{dd}$ is the measurement error covariance, and r is sampled from a standard normal distribution with zero mean and an identity matrix as the covariance. The factor $\sqrt{\alpha}$ scales the perturbations. In this study, we employed 4 ESMDA iterations, with $\alpha = 4$.

The ESMDA workflow, summarized below, iteratively employs these equations to update each ensemble member.

1. **Initialization**: Initialize the ensemble with randomized reservoir permeability parameters, collect measured vertical displacement data $\boldsymbol{d}_{\text{obs}}$, and define the covariance matrix for measurement errors $\mathbf{C}_{dd}$.
2. **Iteration Parameter Determination**: Establish the number of assimilation cycles N_a and select an inflation factor α for the iterations, spanning $i = 1, \ldots, N_a$.
3. **Iterative Update Cycle**: Execute the following steps in each iteration for every ensemble member:
    (a) **Prediction Phase**: Calculate the expected vertical displacements $G(\boldsymbol{z}_j)^f$ based on the current estimates of reservoir permeabilities.
        - Compute pressure and temperature fields using open-DARTS, incorporating the updated permeability fields.
        - Calculate vertical displacements employing the geomechanical proxy, which integrates the pressure and temperature effects from open-DARTS to predict deformation.
    (b) **Observation Perturbation**: Formulate $\boldsymbol{d}_j$ by implementing the perturbed observation technique to consider observational uncertainty.
    (c) **Parameter Update**: Adjust the reservoir permeability parameters $\boldsymbol{z}_j^a$ via the update equation to enhance alignment between the predicted and observed vertical displacements.
4. **Posterior Assessment**: Upon completing the iterative process, assess the adjusted ensemble of permeabilities $\boldsymbol{z}_j$ using open-DARTS for the final posterior computation of the pressure and temperature field and the geomechanical proxy for the vertical displacements field.

Information theory provides a framework for quantifying uncertainty in model predictions and has been applied in various geoscience applications (Kim & Vossepoel, 2024), including reservoir surveillance optimization (Le & Reynolds, 2013). Shannon entropy measures the uncertainty of a random variable based on the probability distribution of its possible outcomes (Shannon, 1948). In our approach, the reduction in Shannon entropy of x_{CO_2} serves as a measure of the uncertainty reduction achieved by assimilating both geomechanical and BHP data. For a discrete random variable X with possible outcomes $\{x_1, x_2, \ldots, x_n\}$ and associated probabilities $p(x_i)$, the entropy is defined as:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \ln p(x_i). \tag{3.10}$$

In this study, the random variable X represents the CO_2 molar fraction x_{CO_2}, characterizing the CO_2 plume within the reservoir. The posterior probability distribution $p(x_{\text{CO}_2})$—the cell-wise CO_2 molar fraction computed by the flow simulator-is estimated from an ensemble of model realizations computed with ESMDA. To calculate entropy, the range of x_{CO_2} is discretized into n bins, with $p(x_{\text{CO}_2,i})$ representing the probability of the ith bin. Alternative non-parametric density estimators, most like kernel-density or

k-nearest-neighbour methods could also be considered, but they demand bandwidth or neighbourhood tuning and, at our ensemble size of 100, provide no discernible advantage; we therefore adopt the fixed-bin histogram for its transparency, computational economy, and consistency with previous studies (Kim & Vossepoel, 2024; Le & Reynolds, 2013)

3.3. TEST CASES

We used two grids: a conceptual model with a structured grid and a Brugge grid upscaled to one layer. The model setups are designed to test the ESMDA and geomechanical proxy integrated framework. The conceptual model allows for controlling the known variables in a more predictable environment, whereas the Brugge dataset model introduces complexities that challenge the method's capability to handle real-world unpredictability and heterogeneity.

The initial pressure and temperature are defined by a gradient of 100 bar/km and a thermal gradient of 30 °C/km from 1 bar and 20 °C at the surface, respectively.

The flow boundary conditions are defined by assigning a large effective volume to the lateral boundary cells, thereby simulating an infinite reservoir, while enforcing no-flow restrictions at the top and bottom boundaries. The impact of overburden and underburden layers in the hydrodynamic model is out of scope for this study.

This leads to the assumption that the pressure and the temperature remain constant above and below the reservoir. The assumption of no-flow boundary conditions for the reservoir top and bottom is reasonable for CO_2 injection applications, as its leakage from the reservoir is commonly avoided and reservoir with sufficient cap rock integrity is preferable to be chosen. Since the geomechanical proxy model uses only the pressure changes in the reservoir part, imposing drained boundary conditions at the reservoir top and bottom can impact the geomechanical response, the reservoir grid should be extended vertically up to non-permeable barriers to properly evaluate the displacements.

The presence of the pressure changes in the lateral reservoir surrounding is neglected in our study since one can extend the reservoir laterally as well until the impact to the displacements in the interest area will be negligible due to the large distance and low pressure changes there.

The simulation time is 30 years, with 10 years of injection followed by 20 years of monitoring. In the Data assimilation section, the simulation runs only for the first 10 years to imitate the field development-in-progress study. However, application to the whole time-series might improve the data assimilation results. The well injection rates are listed in Table 3.1.

The GCS model operates within the open-DARTS compositional formulation, which utilizes equilibrium flash calculations to establish phase distributions (Wapperom et al., 2023). Comprising CO_2 and H_2O components, the system is initially at water saturation with CO_2 as the injection component in the liquid phase. The simulation consists of 10 years of injection proceeded with 20 years of monitoring.

A conceptual reservoir model with 2D structured grid utilized in this study is a simplified representation to emulate GCS scenarios for our experiments (Figure 3.4). The reservoir comprises a 51x51x1 structured grid with right prism-shaped cells and uniform properties. For the GCS case, the injection well is situated in the center of the grid to explore the radial flow and pressure propagation characteristics typical of single-well injection systems. The rock properties, maintained uniform across the model, are specified within the accompanying table 3.1. We will plot hydrodynamic model variables along the monitoring points line at reservoir depth and geomechanical variables (vertical displacements) at the surface (depth=0).

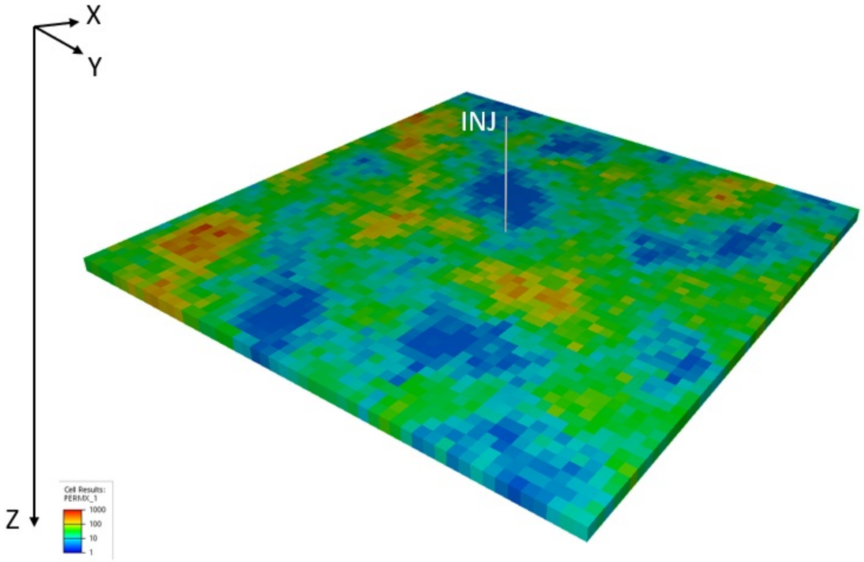


Figure 3.4: 2D Conceptual Model setup showing the grid layout (top view), well positions, and monitoring points for the GCS case.

The Brugge dataset offers a more challenging and realistic scenario for our data assimilation and subsurface modeling approaches. Unlike the conceptual model, the Brugge dataset model includes complex geological features and heterogeneity that are common in actual reservoirs. This model is constructed using a Corner Point Geometry (CPG) grid, which is commonly used to capture the geological layers and faults within the reservoirs. The geometry of the Brugge dataset has been upscaled to one layer and permeability from the first layer of the original model is used.

The rock properties, well rate and geometry summary are specified in table 3.1. The grid contains 6672 cells with 4950 active cells. The grid exaggerated 5 times by Z-axis, the permeability field used for the single run and monitoring points distributed along the wells locations as displayed in the Figure 3.5. The GCS model setup has the injection well situated at I-index=100 and J-index=31.

According to the application to the generic grid discussed in Section 3.2.2, we used the interpolation in order to apply the geomechanical proxy method for Brugge dataset geometry. Two different resolutions are used for the construction of a structured grid: with 5 and 10 refinement factors of the original grid.

The results showed that the basic method of a cell approximation with a right prism

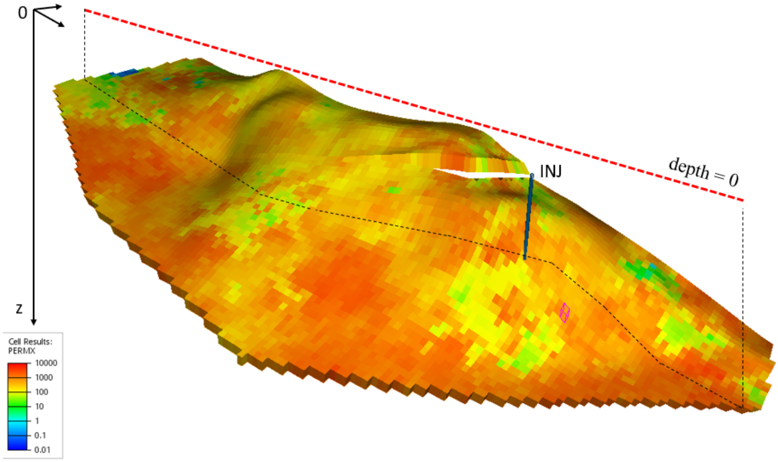


Figure 3.5: Brugge dataset model setup illustrating CPG grid, well locations, and monitoring points along J-index=31 for the GCS case.

and averaged sizes provides a solution for the subsidence/uplift similar to the interpolation with a refinement factor of 10. Meanwhile, the implementation with interpolation significantly increases the number of prisms which affects the performance of geomechanical evaluation. Thus, we used a simple averaging approach described in the Section 3.2.2, since data assimilation requires a fast displacement evaluation.

Table 3.1: Model setup

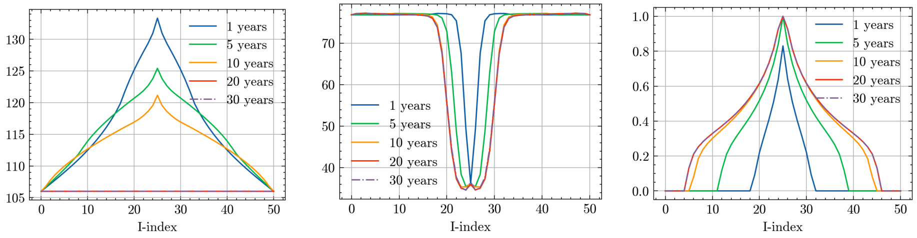
Property	**Conceptual case**	**Brugge case**
Grid size (XY), km	4x4	17x9
Thickness, m	100	58.2-224.7
Grid dimensions	51x51x1	139x48x1
Reservoir top depth, km	1000	1.545-1.945
Injection rate, $Mt/year$	2.5	10
Injection temperature, °C	11.5	36.85
Porosity	0.2	0.18 (mean)
Permeability, mD	100	540 (mean)
Heat capacity, $kJ/m^3/K$	2200	2200
Heat conductivity, $kJ/m/day/K$	181.44	181.44
Poisson ratio	0.25	0.25
Young modulus, MPa	3300	3300
Linear thermal expansion coefficient, 1/°C	1.3e-5	1.3e-5

3.4. RESULTS

This section presents an integrated analysis of results from hydrodynamic simulations combined with the application of a geomechanical proxy model, focusing on monitoring points at various times evaluated at the surface for the test cases described earlier. It is important to note that as shown in the Figure 3.5, for this study, the Z-axis points downwards, so negative values of the vertical displacements at the surface correspond to uplift, and positive values correspond to subsidence.

For the conceptual model, hydrodynamic model results are illustrated in Figure 3.6, where the pressure increase in the reservoir leads to the uplift. The conceptual case grid is 2D, so there is no gravity or gas migration after injection stops at 10 years. The pressure change decreases over time as the CO_2 plume propagates, as CO_2 has lower viscosity than water. These results in a decrease of uplift from a peak of 20 mm (at 1 year) to 15 mm (at 10 years) are shown in Figure 3.7. In the meantime, thermoelastic subsidence is observed due to the rock thermal contraction as the injection gas has a lower temperature with a maximum value of 4 mm (at 10 years). Thus, thermoporoelastic displacement has a cap-cutted parabola shape.

After stopping the injection at 10 years, the pressure change vanishes, which leads to the dissipation of the poroelastic effect. However, the thermoelastic effect stays leading to the subsidence of 1.8 mm in maximum above the injection well.

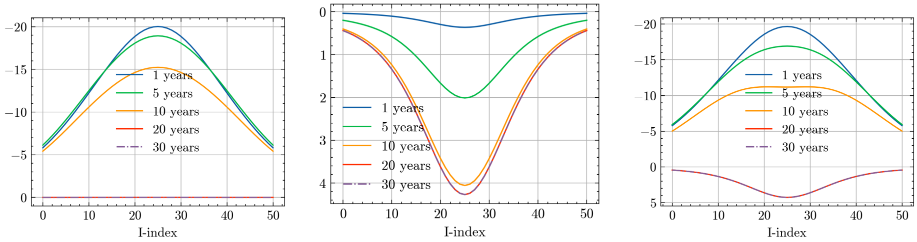


(a) Pressure, bars.

(b) Temperature, °C.

(c) CO_2 saturation

Figure 3.6: Model results for the conceptual case. 1D plots at the monitoring points at different times



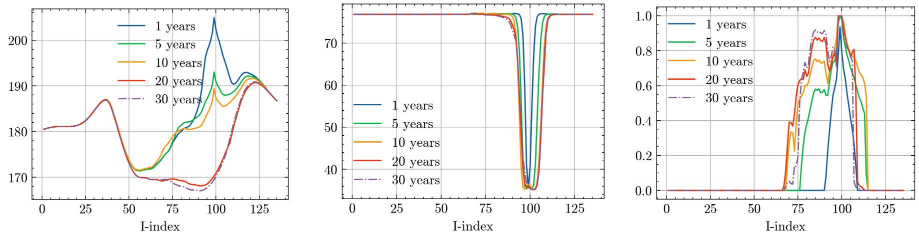
(a) Poroelastic displacements, mm.

(b) Thermoelastic displacements, mm.

(c) Thermoporoelastic displacements, mm.

Figure 3.7: Model results for the conceptual case. 1D plots of the displacements at the monitoring points at different times

For the Brugge dataset, the hydrodynamic model simulation results are shown in Figure 3.8, displaying the pressure and temperature distributions and the CO_2 saturation at different times. The corresponding elastic displacements at the surface are illustrated in Figure 3.9. The overall behavior is similar to the conceptual case results. However, the results slightly differ due to the heterogeneity and gravity effects. The vertical displacement amplitude at the surface for the Brugge dataset is lower than the conceptual case due to the deeper location of the Brugge dataset grid.

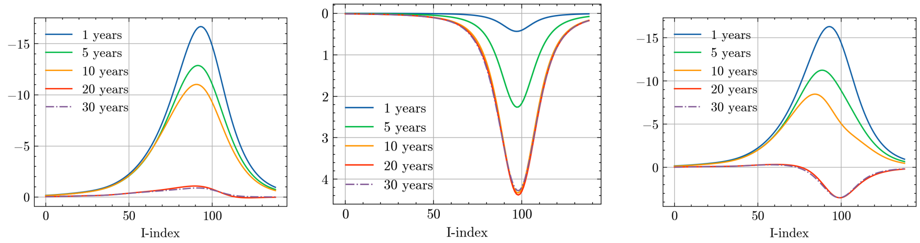


(a) Pressure, bars.

(b) Temperature, °C.

(c) CO_2 saturation

Figure 3.8: Model results for the Brugge dataset. 1D plots at the monitoring points at different times



(a) Poroelastic displacements, mm.

(b) Thermoelastic displacements, mm.

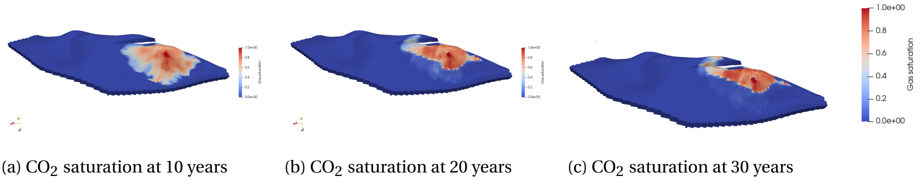
(c) Thermoporoelastic displacements, mm.

Figure 3.9: Model results for the Brugge dataset. 1D plots of the displacements at the monitoring points at different times

Apart from the conceptual case, the Brugge grid cells have different depths, and gas migration can be observed due to the gravity effect after stopping the injection at 10 years, which is shown in Figure 3.10. A part of the fault is impermeable due to juxtaposition, which adjusts the gas movement direction. Gas migration leads to the shape change of the cold plume, as illustrated in Figure 3.10.

The displacements at the surface indicate an uplift at the injection stage, as the poroelastic response is higher than the thermoelastic response working in the opposite direction.

Since we did not aim for detailed CO_2 modeling, we used grids with one layer only. However, CO_2 injection modeling often requires a high-fidelity multilayered 3D grid to properly represent the gas migration, given the fingering effect and gas dissolution in the water (Lyu & Voskov, 2023; Wapperom et al., 2023).



(a) CO_2 saturation at 10 years (b) CO_2 saturation at 20 years (c) CO_2 saturation at 30 years

Figure 3.10: Model results for the Brugge dataset. CO_2 saturation at different times.

One must notice that the combined thermoelastic and poroelastic effects are highly impacted by the parameters applied in this study, and the final results can change if different parameters are used.

3.5. DATA ASSIMILATION RESULTS

This section outlines results for the application of ESMDA in a geomechanical proxy model for the conceptual model and Brugge dataset cases. These evaluations reveal the method's efficacy in both synthetic environments and actual field applications when applied to CO_2 geological storage. We generated a prior ensemble for each case as described in Section 3.2.3, designating one realization as the reference 'true' model. Synthetic observations of vertical displacements were sampled from this reference at the final time step along a transect intersecting the well trajectories (Figure 3.4). These data, despite inherent noise, are used to adjust the model's permeability distribution, enhancing predictions of CO_2 plume distributions.

The prior ensemble consists of 100 permeability fields, with an additional realization serving as the reference model. For the conceptual model, we adjusted the ensemble's permeability distribution, setting the mean log-permeability to 3 mD, with bounds from 1 mD to 8 mD (Figure 3.11). This adjustment resulted in more varied pressure propagation and uplift patterns, allowing for a clearer demonstration of the data-assimilation process in CO_2 storage.

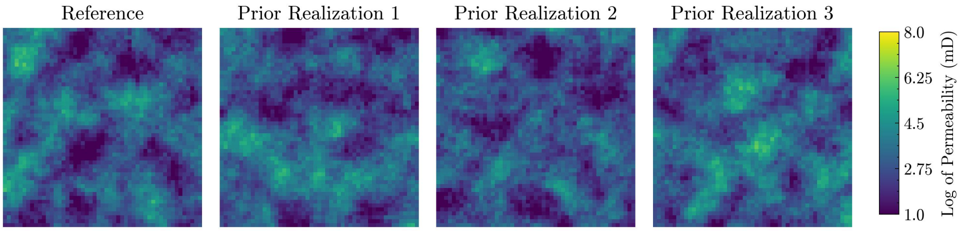


Figure 3.11: Reference model realization and three realizations of the prior permeability distribution for the conceptual dataset.

The posterior realizations exhibit uplift curves characteristic of CO_2 injection, with pronounced uplift centered around the injection site (Figure 3.12). This pattern reflects the pressure diffusion from the injection well and aligns well with the observed reservoir response.

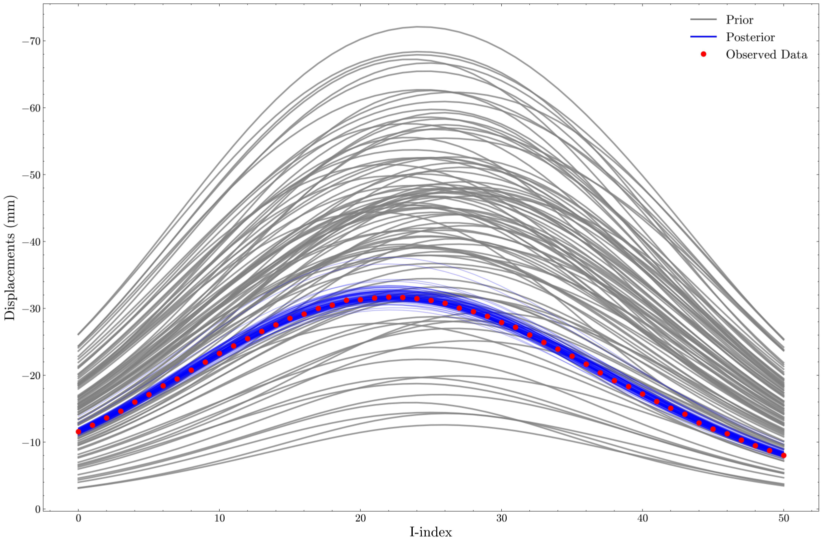


Figure 3.12: ESMDA results for the CO_2 storage scenario, illustrating the vertical displacement realizations. The observed data are shown as red dots, with prior realizations in gray and posterior realizations in blue.

The difference plots show the shifts in permeability close to the CO_2 injection site (Figure 3.13), indicating that data assimilation strongly impacts areas experiencing large pore pressure changes. This can be attributed to the more pronounced pressure variations in this area, leading to larger displacements at the surface in the vertical projection of the injector well.

Reviewing the histograms and CDFs across the realizations (Figure 3.14), the posterior distributions show adjustments from the prior distributions, indicating the impact of data assimilation on the permeability field.

To quantitatively evaluate the impact of data assimilation on the probability distribution of CO_2 molar fraction within the reservoir, we constructed probability maps considering a threshold of CO_2 molar fraction equal to 0.1. Figure 3.15 shows the change in the probability distribution post-data assimilation, indicating how the observations inform on CO_2 molar fraction.

Following the approach described in Section 3.2.4, we quantified the model's uncertainty using the entropy of CO_2 molar fraction distribution probabilities. Table 3.2 shows the entropy reduction from 725.2 $nats$ to 340.1 $nats$, representing a 53.1% decrease in relative uncertainty after data assimilation. This large reduction is noteworthy because it demonstrates the ability to constrain the CO_2 plume prediction, even in the face of non-unique permeability distributions. This improvement in predictive capability is important for risk assessment and decision-making in CO_2 storage projects, where accurate plume tracking is more critical than determining the exact permeability field.

For the Brugge dataset, the difference plots (Figure 3.16) highlight the most significant adjustments to permeability occurring in the vicinity of the injection site, similar to the 2D case. However, the spatial distribution of these changes is more complex, reflecting

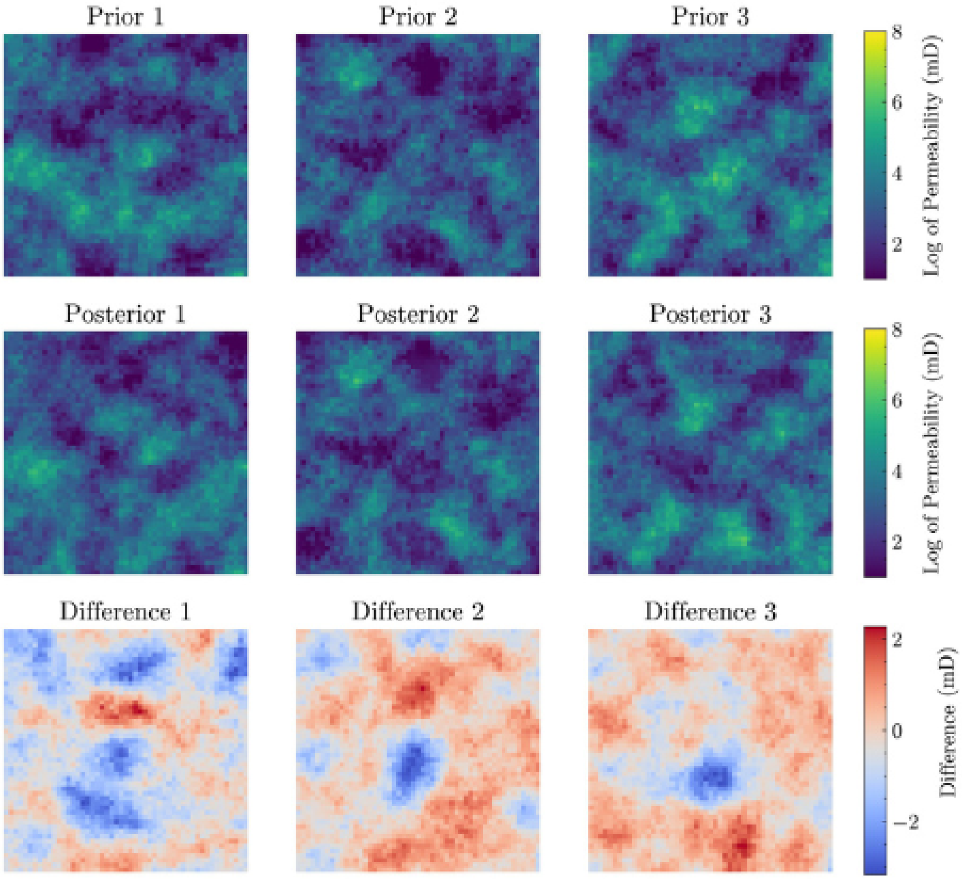


Figure 3.13: Comparison of prior, posterior, and the difference in permeability fields. The difference plots emphasize the localization and magnitude of permeability adjustments around the injection site made by the ESMDA.

	Entropy (*nats*)	Uncertainty Reduction (*nats*)	Relative Uncertainty Reduction
Prior	725.2	–	–
Posterior	340.1	385.1	53.1%

Table 3.2: Quantification of the uncertainty reduction in terms of entropy for the CO_2 molar fraction distribution before and after data assimilation.

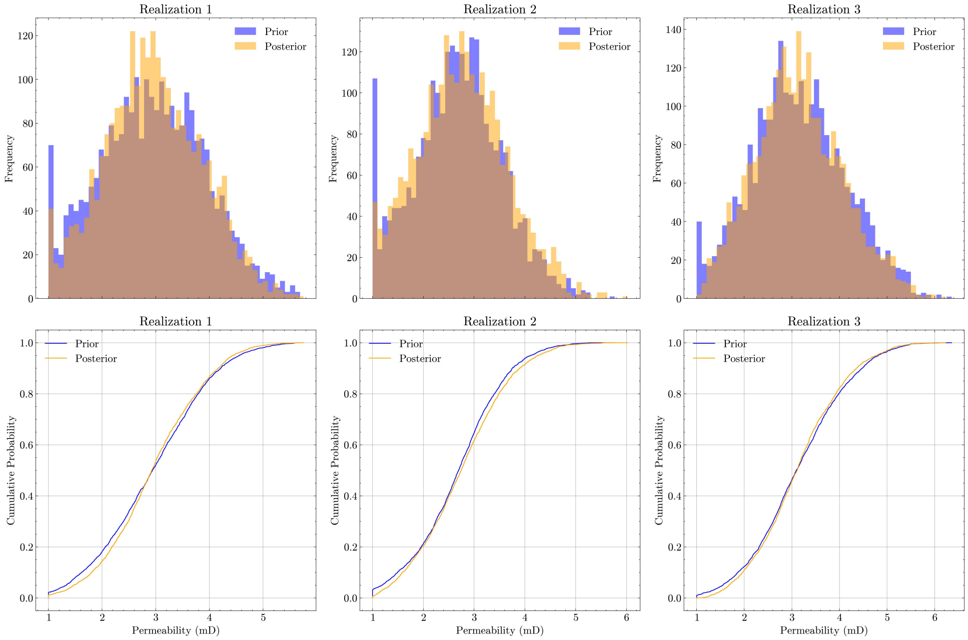


Figure 3.14: Histograms and cumulative distribution functions displaying the permeability values before (blue) and after (yellow) the use of ESMDA. The brown area is the overlap region of the distributions.

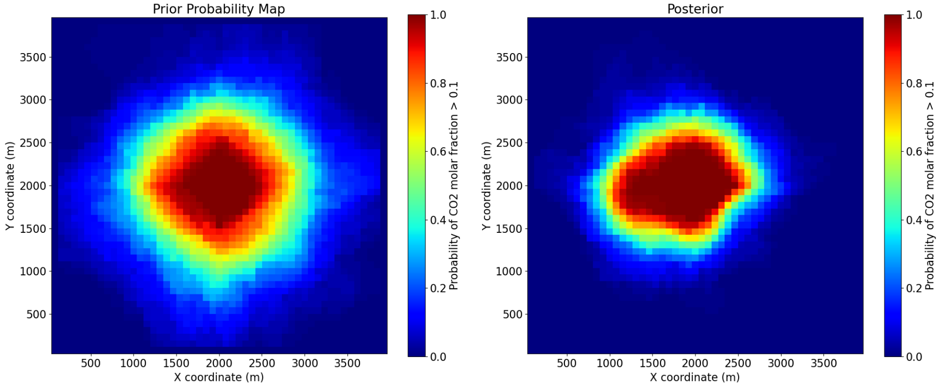


Figure 3.15: Probability maps of CO_2 molar fraction exceeding 0.1 for the prior (left) and posterior (right) models. The post-data-assimilation map shows a distinct change in the probability distribution.

the heterogeneity and structural complexity of the Brugge dataset. This emphasizes the need for detailed, high-resolution models to capture the nuances of CO_2 injection and its impact on the subsurface in a realistic case.

The histograms and CDFs (Figure 3.17) exhibit a slight narrowing of the posterior permeability distribution compared to the prior. This suggests that while the ESMDA was able to refine the permeability realizations to better match the simulated vertical displacements with the observed ones, the inherent complexity of the Brugge dataset limits the degree of adjustment possible.

The probability maps for CO_2 molar fraction exceeding 0.1 in the Brugge dataset (Figure 3.18) show a pronounced change in the probability distribution post-data assimilation. This indicates a significant impact of the data assimilation process on the CO_2 plume prediction in the complex reservoir setting.

The entropy reduction analysis for the Brugge dataset (Table 3.3) shows a decrease from 354.0 nats to 288.5 nats, corresponding to an 18.5% reduction in CO_2 molar fraction uncertainty post-data assimilation. While this improvement is less dramatic than in the conceptual case, it still represents an enhancement in our ability to predict CO_2 plume behavior in a complex reservoir setting.

	Entropy ($nats$)	Uncertainty Reduction ($nats$)	Relative Uncertainty Reduction
Prior	354.0	–	–
Posterior	288.5	65.1	18.5%

Table 3.3: Entropy reduction for the Brugge dataset, indicating enhanced model reliability and predictive accuracy.

3.6. DISCUSSION

In this study, we integrate vertical displacement observations with open-DARTS reservoir simulations and a physics-based geomechanical proxy model to enhance subsurface model predictions for CO_2 storage applications. While different permeability fields may generate similar vertical displacement responses, leading to non-unique permeability estimates, our ESMDA method effectively estimates likely permeability fields and their probability distributions. More importantly, this approach successfully constrains the posterior states of the reservoir, reducing uncertainty in critical variables such as CO_2 plume extent, even when the exact permeability field cannot be uniquely determined.

The integration of our geomechanical proxy model with ESMDA enhances subsurface modeling for CO_2 storage by leveraging often underutilized surface displacement measurements. This approach improves interpretation of complex subsurface behaviors in response to CO_2 injection activities. However, we acknowledge that the assumption of homogeneous geomechanical properties may not fully reflect real-world formation heterogeneity, potentially affecting the accuracy of displacement and stress distributions.

Our analysis demonstrates improvements in uncertainty reduction, with the conceptual case showing a 53.1% reduction in CO_2 plume uncertainty and the more complex Brugge case achieving an 18.5% reduction. The difference in improvement levels between the two cases highlights the challenges of working with realistic reservoir complexities.

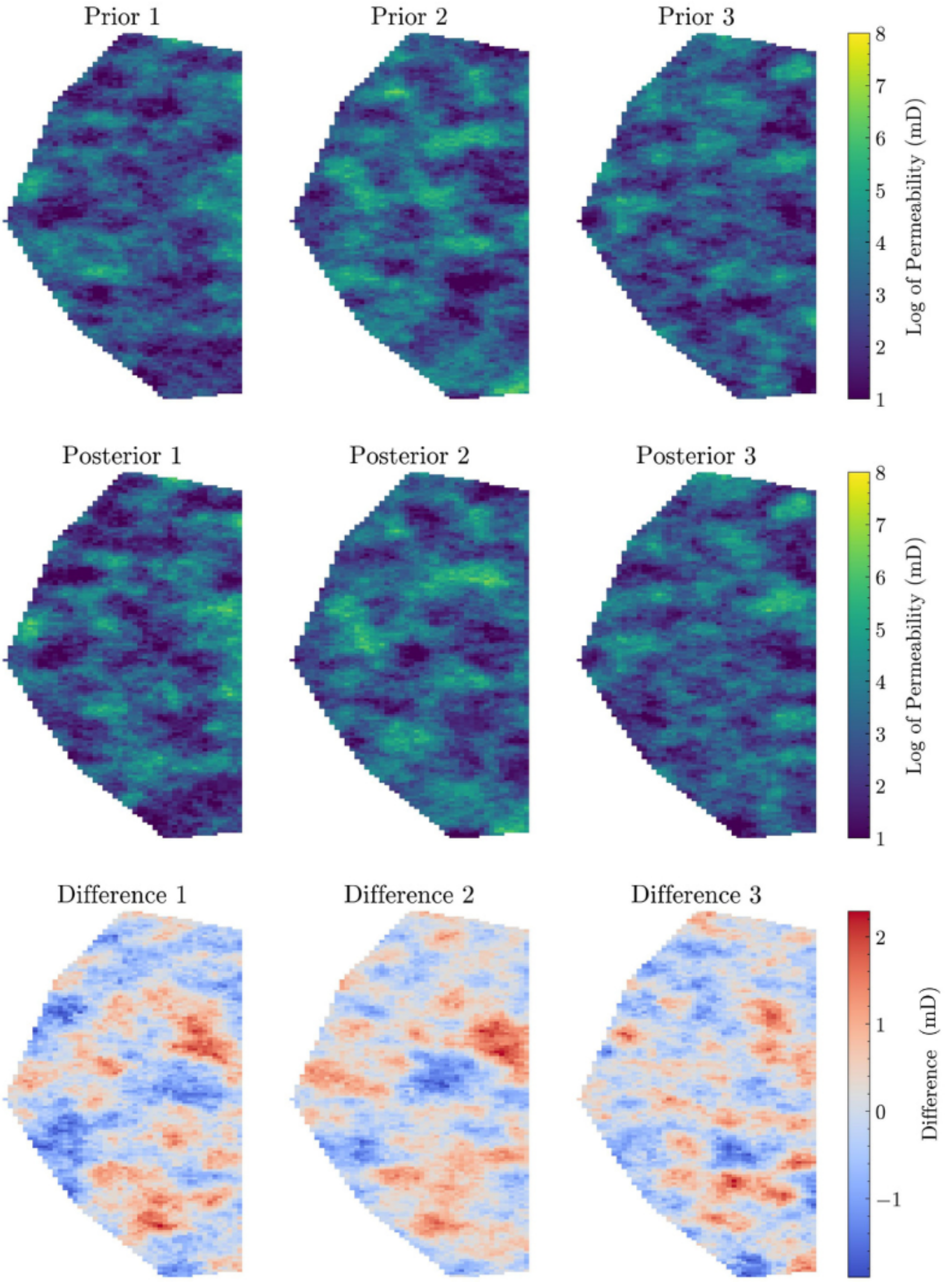


Figure 3.16: Comparison of prior, posterior, and the difference in permeability fields for the Brugge dataset. The difference plots showcase the localized adjustments made to the permeability field around the injection site.

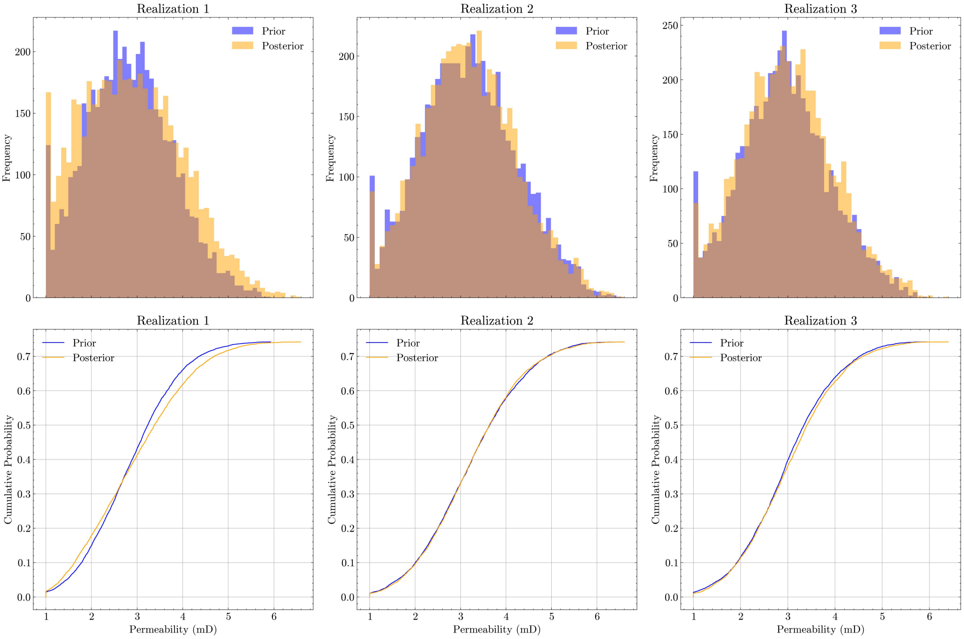


Figure 3.17: Histograms and cumulative distribution functions displaying the permeability values before (blue) and after (yellow) the ESMDA for the Brugge dataset. The brown area is the overlap region of the distributions.

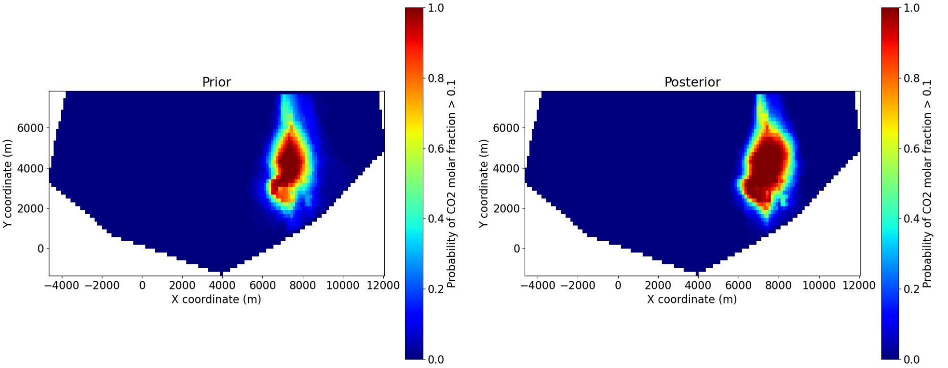


Figure 3.18: Probability maps of CO_2 molar fraction exceeding 0.1 for the prior (left) and posterior (right) models for the Brugge dataset.

The more modest improvement in the Brugge case can be attributed to:

- Complex geological structures and heterogeneities that create more intricate flow patterns
- Deeper reservoir placement affecting the magnitude of surface displacements
- The presence of faults influencing CO_2 migration patterns
- More complex pressure and temperature distributions due to varying reservoir properties

Incorporating 2D surface displacement measurements and using displacement data from multiple time steps could enhance the accuracy of subsurface models by providing more comprehensive temporal and spatial constraints. While our current study focused on 1D displacement data at the final time step to demonstrate the effectiveness of our methodology, future research could explore these extensions as part of a broader sensitivity analysis. Such analyses—including the impact of data sparsity, topographical density, and the timing of data collection—are valuable but beyond the scope of the present work. Addressing these aspects in future studies would help optimize data collection strategies and further validate the robustness of our approach in various geological settings.

Our study established a Data Assimilation framework that has several opportunities for future research. These include comparing the effectiveness of various data types, quantifying uncertainty in geomechanical parameters, conducting more extensive validation of the proxy model, and exploring other data assimilation techniques beyond ESMDA. Future studies should also focus on incorporating heterogeneous geomechanical properties, extending the framework to different geological settings, and integrating diverse data types such as seismic data to further validate its robustness and expand its applicability.

It is important to note that the use of vertical displacement data for assimilation in CO_2 storage applications is not common practice. Traditional approaches often rely on well data or seismic surveys, overlooking the potential of surface deformation measurements. Our study demonstrates the value of incorporating these often-underutilized observations, offering a novel approach to constrain subsurface models and reduce uncertainties in parameters such as permeability distributions and CO_2 plume extents. This innovative use of vertical displacement data opens new avenues for improving subsurface characterization in CO_2 storage projects.

3.7. CONCLUSION

This study presents the development and application of a physics-based geomechanical proxy model for approximate thermoporoelasticity solutions, designed to integrate vertical displacement observations within a data assimilation framework. Our approach employs ESMDA to quantify uncertainties in CO_2 plume distribution for Geological Carbon Storage (GCS) projects. The model's flexibility allows for runtime application or post-processing in hydrodynamic simulations, with displacement evaluations at desired time steps and compatibility with various simulators. Its adaptability to non-rectangular grids enhances its suitability for realistic geological settings. The integration with open-DARTS'

OBL caching feature significantly improved computational efficiency, demonstrating the potential for more complex and extensive simulations in CO_2 storage applications. Our ESMDA implementation demonstrated efficacy in uncertainty quantification across both conceptual and realistic scenarios. In the conceptual model, the method achieved a 53.1% reduction in CO_2 plume uncertainty, benefiting from larger vertical displacements and ensemble variance. The application to the Brugge dataset further validated the method's effectiveness in complex geological settings, showing an 18.5% reduction in CO_2 molar fraction uncertainty. These results highlight the method's robustness in handling realistic reservoir complexities. The comparative analysis between the conceptual and Brugge cases highlights both the potential and challenges of our approach. While the degree of improvement varied based on geological complexity and reservoir depth, both cases benefited from ESMDA application. Our study pioneers the use of vertical displacement data for assimilation in CO_2 storage applications, a practice not commonly employed in the field. This innovative approach demonstrates the potential of integrating geomechanical proxy models with ESMDA to enhance subsurface modeling accuracy and reliability, leveraging often overlooked surface deformation data. By doing so, we offer a valuable tool for advancing CO_2 storage projects through improved reservoir characterization and uncertainty quantification, potentially changing how the industry approaches data assimilation in carbon storage projects.

CODE AND DATA AVAILABILITY

The developed code is open source and available at the gitlab repository https://gitlab.com/open-darts/subsidence-esmda together with input files of the numerical models. The open-DARTS simulator is available at https://gitlab.com/open-darts/open-darts. The dageo Python package for data assimilation in geosciences is available at https://github.com/tuda-geo/dageo.

4

GEOMECHANICAL PROXY MODEL AND DATA ASSIMILATION FOR EVALUATING MONITORING STRATEGIES IN GEOLOGICAL CARBON SEQUESTRATION

This chapter presents an entropy-based framework to quantitatively evaluate monitoring strategies for geological carbon sequestration. The methodology integrates CO_2 *flow simulations, geomechanical proxy models, and ensemble-based data assimilation to assess how spatial sensor configurations, temporal sampling frequencies, and data types affect uncertainty reduction. Testing on both conceptual and realistic reservoir models reveals that temporal measurement frequency can provide information gain comparable to or exceeding that of spatially dense sensor networks. The framework demonstrates that strategic temporal sampling with limited sensors can achieve similar uncertainty reduction to extensive spatial arrays, offering quantitative guidance for cost-effective monitoring design in carbon storage projects.*

Parts of this chapter will be published in Gabriel Serrão Seabra, Ilshat Saifullin, Denis Voskov, Femke C. Vossepoel, Geomechanical Proxy Model and Data Assimilation for Evaluating Monitoring Strategies in Geological Carbon Sequestration, in *Subsurface Data Assimilation: Theory and Applications* (X. Luo, Ed.), book chapter – in preparation, expected 2025. Changes have been applied to make the text and figures consistent with the thesis.

4.1. INTRODUCTION

For GCS projects to succeed and gain public acceptance, it's important that we can reliably monitor the injected CO_2, ensuring it stays securely stored without causing environmental harm (Ringrose et al., 2013). This involves monitoring the CO_2 plume to avoid potential leakage, assessing the caprock integrity, and complying to regulatory frameworks (Jenkins, 2015). However, GCS projects are complex, primarily due to important uncertainties in subsurface properties such as the permeability field of the reservoir (March et al., 2018; Witter et al., 2019). These pose challenges to project feasibility, operational efficiency, and environmental integrity. Knowing underground properties helps predict how the CO_2 will spread and how pressures will change, important for managing risks and operations.

Traditionally, monitoring strategies for GCS have relied on seismic surveys, well logging, and pressure measurements (Ringrose, 2020). For certain fields, there is also geomechanical data available, particularly surface deformations resulting from subsurface activities, which can provide insights into reservoir behavior and caprock integrity (Bohloli et al., 2022; Glegola et al., 2012; Ringrose et al., 2013; Rutqvist, 2012; Seabra et al., 2024a). When geomechanical data are not used, we miss important information about underground conditions. For offshore environments, recent reviews summarize subsidence monitoring techniques (Thomas et al., 2024). While this data is occasionally incorporated into monitoring programs, there remains a gap in quantitative frameworks for evaluating the effectiveness of different monitoring configurations and data types (Davis et al., 2023; Seabra et al., 2024a). The spatial arrangement and temporal frequency of measurements influence the information gained about subsurface processes, yet monitoring design is often determined determined through ad hoc decisions rather than through systematic, quantitative analysis (Bouqueta et al., 2022; Seabra et al., 2024a). This gap shows the need for more methodologies that help quantitatively assess the result of different monitoring strategies.

Effective monitoring of surface deformation is important for GCS projects and require instruments with sufficient precision. Advancements in geodetic and geophysical monitoring technologies have improved our ability to measure surface displacements at increasingly finer scales, providing reliable data for subsurface characterization (Im et al., 2017; Rucci et al., 2012). Table 4.1 summarizes the precision capabilities of monitoring instruments commonly employed in geomechanical monitoring programs, highlighting their ability to detect sub-millimeter displacements relevant for CO_2 storage monitoring.

These technologies provide complementary capabilities that, when deployed in integrated monitoring systems, can effectively characterize surface deformations associated with subsurface pressure changes during CO_2 injection operations. Researchers have demonstrated that when multiple monitoring techniques are combined, they provide robust characterization of reservoir behavior with detection thresholds approaching sub-millimeter scale (Im et al., 2017). The sensitivity of modern instruments enables the detection of millimeter-scale displacements, and incorporating those additional observations can provide useful information when evaluating caprock integrity. (Raziperchikolaee et al., 2021).

Integrating geomechanics studies with reservoir modeling faces obstacles, including the difficulties of modeling rock mechanics-fluid flow interactions and the substantial computational resources required for traditional coupled numerical simulations. To

Table 4.1: Precision capabilities of modern geomechanical monitoring instruments.

Instrument & Technique	**Typical Resolution/Precision**	**Detectable Deformation**	**Applications**
Tiltmeter (surface or borehole)	~1 nanoradian sensitivity; <5 nanoradians for ultra-precision sensors; Long-term stability ~≤1 microradian/year (Henbest et al., 1978)	Angles ~1 nanoradian ≈ 0.001 mm over 1 km baseline; Can detect <0.1 mm of differential ground movement over tens to hundreds of meters	Highest sensitivity to minute changes; Detecting ground tilt changes; Monitoring subsidence/uplift in GCS projects (Ketelaar et al., 2020)
InSAR (multi-image)	Time-series (20+ scenes): <1 mm/year velocity precision; With corner reflectors: ~0.1–0.5 mm single-measurement noise in ideal conditions (Rucci et al., 2012)	~1 mm of cumulative motion over months to years; Reliably detects 2 mm/yr subsidence patterns; Sub-mm/year trends possible with stable targets (Vasco et al., 2013)	Wide-area coverage (hundreds of km^2); Dense point measurements; Mapping surface deformation related to CO_2 injection
Continuous GPS/GNSS	Single station daily: ~3 mm horizontal, ~6–10 mm vertical precision; Long-term: ~1 mm/yr horizontal, ~1–2 mm/yr vertical velocity (Wang, 2013)	~1–2 mm of annual vertical motion relative to stable reference; 3D vector of motion recorded	High-frequency data captures transient events; Continuous monitoring; Often collocated with InSAR reflectors for cross-calibration in reservoir monitoring (Raziperchikolaee et al., 2021)

overcome limitations related to the high computation demand, researchers have developed surrogate geomechanical models, such as neural networks (Tang et al., 2022) and physics-based proxies that combine analytical and numerical approaches. Significant progress has been made by van Wees et al. (2019) and Paullo Muñoz and Roehl (2017), who introduced methodologies utilizing Green‘s function and the nucleus of strain concept to model stress responses and forecast compaction and subsidence. These innovative approaches have been validated against finite element models and applied to real-world scenarios, such as evaluating induced seismicity in the Groningen gas field (Candela et al., 2022b). These models enable the application of data assimilation techniques that would otherwise be prohibitive with fully-coupled simulations.

Data assimilation methods are important tools to integrate observations with model realizations to quantify the uncertainty of reservoir models. Ensemble-based strategies like EnKF and ESMDA (Emerick & Reynolds, 2013b), and gradient-based methods such as RML (Tian et al., 2024) and 4DVar (Courtier et al., 1994; Oliver et al., 2008; Tarantola, 2005) have shown promise in various subsurface applications. In particular, ESMDA has been successfully applied across several practical data assimilation problems in subsurface modeling (Emerick & Reynolds, 2013a; Evensen et al., 2022, 2024) and its sampling performance has been investigated (Emerick & Reynolds, 2013a); it has also been used to perform information gain studies using information theory (Le & Reynolds, 2013).

Information theory can provide a quantitative mathematical framework to evaluate the effectiveness of different monitoring strategies quantitatively. Shannon entropy (Shannon, 1948) quantifies the uncertainty in a random variable, making it particularly suitable for measuring the information content in probabilistic models. In the context of GCS monitoring, entropy reduction between prior and posterior ensemble distributions provides a quantitative measure of information gain from different monitoring configurations. This approach has been applied in various geoscience applications (Boso & Tartakovsky, 2018; Kim & Vossepoel, 2024) and in reservoir surveillance optimization (Le & Reynolds, 2013), offering a way to quantify monitoring effectiveness beyond traditional metrics like parameter estimation error or prediction accuracy.

In our previous study (Saifullin et al., 2024), we introduced an integrated framework combining a physics-based geomechanical proxy model, ESMDA, and open-DARTS (Denis Voskov et al., 2025; Voskov et al., 2024), demonstrating its capability for data assimilation in CO_2 storage and geothermal applications. The current work extends this framework by evaluating various monitoring strategies through quantitative metrics based on entropy reduction. Specifically, we use Shannon entropy to measure the uncertainty reduction in the characterization of the CO_2 plume, providing an objective and quantifiable means to compare different monitoring configurations.

We acknowledge that several factors could influence the effectiveness of monitoring strategies; however, it is important to clarify the aspects that are beyond the scope of this study. We do not apply optimization techniques to determine the optimal monitoring design, as our focus is on evaluating predefined configurations rather than searching for an optimal solution. We also do not investigate the impact of different data assimilation methods, even though alternative approaches could affect the results. Likewise, we do not consider variations in CO_2 injection strategies, despite their potential influence on monitoring outcomes. The objective remains to establish a systematic and quantifiable framework for assessing monitoring strategies, integrating geomechanical proxy models, ESMDA, and open-DARTS within an entropy-based evaluation approach.

To demonstrate this methodology, we apply our framework to two distinct test cases: a 2D conceptual model with log-Gaussian-distributed permeability and the more complex, heterogeneous Brugge reservoir model. Through these case studies, our objectives are to:

- Propose a systematic framework leveraging entropy reduction metrics to quantitatively evaluate monitoring strategies.
- Assess the relative importance of spatial sensor density versus temporal measurement frequency for reducing subsurface uncertainty.
- Quantify the benefits of integrating surface displacement and Bottom Hole Pressure (BHP) data to enhance subsurface characterization.

This study contributes to a comprehensive framework for evaluating monitoring strategies and supporting effective decision-making for CO_2 containment.

4.2. METHODOLOGY

This chapter evaluates different monitoring strategies by applying the integrated simulation and data assimilation framework established in Chapter 3. The core methodology remains consistent, but it is augmented with an information-theoretic approach to quantitatively measure uncertainty reduction. The numerical model, including input files and the subsidence-ESMDA module, is accessible via https://gitlab.com/open-darts/subsidence-esmda.

4.2.1. HYDRODYNAMIC AND GEOMECHANICAL MODELING

For the Hydrodynamic and Geomechanic modeling, this study employs the computational framework detailed in Chapter 3, where subsurface fluid flow is modeled using the open-DARTS compositional thermal simulator. The resulting pressure and temperature changes are used to compute surface displacements via the computationally efficient geomechanical proxy model, which is based on the nucleus of strain approach. The detailed governing equations and implementation of these models can be found in Sections 3.2.1 and 3.2.2 of Chapter 3.

4.2.2. DATA ASSIMILATION FRAMEWORK WITH ESMDA

We apply ESMDA in a similar way as in Chapter 3 to assimilate data and update the ensemble of reservoir models; however, the method is extended by the usage of distance based localization. To mitigate the issue of ensemble collapse due to spurious long-range correlations, particularly when assimilating BHP data, we implement localization techniques (Evensen et al., 2022). Localization restricts the influence of observations to nearby model parameters. We construct a localization matrix **L** that modifies the update equation, following common practice and insights from sampling-performance studies

(Emerick & Reynolds, 2013a):

$$\boldsymbol{z}_j^a = \boldsymbol{z}_j^f + \mathbf{L} \circ \left[\mathbf{C}_{zd}^f \left(\mathbf{C}_{dd}^f + \alpha_i \mathbf{C}_{dd} \right)^{-1} \left(\boldsymbol{d}_j - \boldsymbol{G}(\boldsymbol{z}_j^f) \right) \right], \tag{4.1}$$

where ∘ denotes the Schur (element-wise) product. The localization matrix **L** is designed using the same Gaspari-Cohn correlation function used to model spatial correlations for the permeability distributions, now centered at the injection well for BHP data, ensuring that the influence diminishes with distance. For the synthetic 2-D case the taper uses the same modified Gaspari–Cohn kernel and a horizontal correlation length of 10.

Localization is applied only to the BHP observations in the synthetic 2-D example; surface-displacement data and all Brugge simulations remain un-localized because no ensemble collapse was observed. The localization weights are computed with the same modified Gaspari–Cohn taper used for the prior covariance. Using the GC taper keeps a consistent distance metric, provides compact support to suppress spurious long-range updates, and is already implemented in our open-source library `dageo`, removing the need for extra code. Other taper functions were not investigated here, as the GC kernel fulfilled the numerical and practical requirements of this study.

4.2.3. ENTROPY COMPUTATION

To quantify the uncertainty reduction achieved through data assimilation, we employ Shannon entropy following the methodology detailed in Chapter 3. The entropy computation is applied to the CO_2 molar fraction x_{CO_2} to assess how the integration of geomechanical and BHP data reduces uncertainty in CO_2 plume characterization. The relative entropy reduction, measured in nats, provides a quantitative measure of the information gain achieved through the data assimilation process.

4.3. TEST CASES AND MONITORING STRATEGY DESIGN

4.3.1. TEST CASES

To evaluate the effectiveness of different monitoring strategies, this study employs the same two test cases established in Chapter 3: a 2D conceptual model and the more complex Brugge field dataset. These cases provide a basis for assessment in both controlled and realistic geological scenarios. A complete description of the model geometries, reservoir properties, and simulation parameters can be found in Section 3.3 and Table 3.1. The reference permeability fields used to generate synthetic observations are shown in Figure 3.4 and Figure 3.5.

4.3.2. MONITORING STRATEGY AND EXPERIMENTAL DESIGN

We designed an experimental framework that varies three monitoring aspects: spatial configuration, temporal frequency, and data types to evaluate and quantify the entropy reduction of different monitoring approaches for GCS projects. This design evaluates monitoring configurations that can be applied in various realistic scenarios, including onshore and offshore deployments of tiltmeters, for example. Onshore installations can be constrained by surface infrastructure such as buildings, whereas offshore deployments may face limitations due to existing installations and operational challenges. Different arrangements were tested to account for these factors.

Observation errors are assumed independent and Gaussian. For each point n the variance is $\sigma_n^2 = (\epsilon_n d_n)^2$ with $\epsilon_{\mathrm{disp}} = \epsilon_{\mathrm{BHP}} = 0.01$ (1 %). These figures match modern tilt-/InSAR systems and high-resolution downhole gauges and may be adjusted in future site-specific applications

SPATIAL CONFIGURATIONS

We investigated three spatial arrangements representing realistic deployments of monitoring instruments such as tiltmeters or InSAR reference points. These configurations are illustrated in Figure 4.1, and represent progressively more comprehensive monitoring strategies:

- **Linear Arrays**: Sensors positioned at the surface along a single transect crossing the injection well, representing a monitoring approach where instruments are deployed along the main axis of the reservoir. This configuration simulates scenarios where tiltmeters or GPS stations might be installed along accessible corridors such as roads or property boundaries. For the 2D conceptual model, this array contains 51 measurement points, while for the Brugge model, it comprises 139 points.

- **Low-Density Grids**: A network of 36 measurement points arranged in a regular 6×6 equally-spaced grid covering the area of interest. Tiltmeters or InSAR monitoring points are placed at key locations, with artificial corner reflectors enhancing InSAR reliability in areas with poor natural radar reflectivity, such as vegetated or water-covered regions.

- **High-Density Grids**: A denser network of 121 measurement points configured in a regular 11×11 equally-spaced grid. This arrangement mimics intensive monitoring programs where numerous tiltmeters, GPS stations, or dense InSAR reference points are deployed to capture detailed surface deformation patterns, which might be required in environmentally sensitive areas or under strict regulatory frameworks.

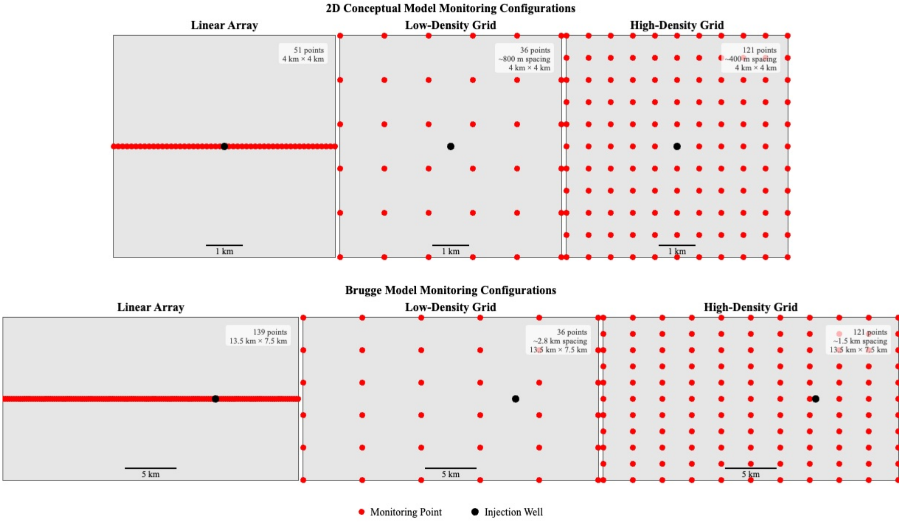
TEMPORAL SAMPLING

For each spatial configuration, we tested two temporal sampling approaches within the simulation period of 10 years:

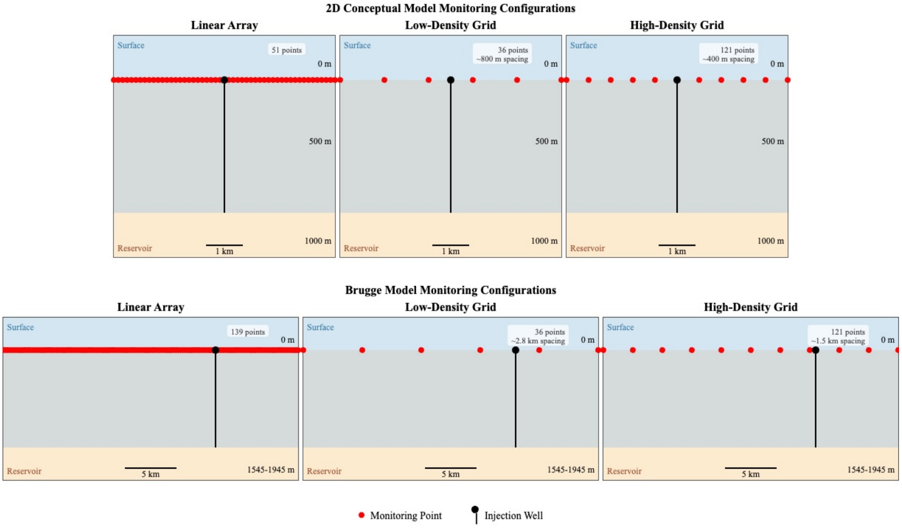
- **Final-Time Measurement**: A single set of measurements taken at the end of the simulation period (year 10), representing a simplified monitoring approach with minimal temporal resolution.

- **Multiple-Time Measurements**: Data collected at years 1, 5, and 10, capturing early, mid-term, and end-of-injection CO_2 plume behavior. This approach enables tracking the evolution of subsurface changes throughout the injection period.

BHP is logged at every simulation time step and is assimilated at this full native resolution. The ’final-time’ and ’multiple-time’ scenarios defined previously, therefore, apply only to the surface-displacement observations.



(a) Top view of monitoring configurations



(b) Side view of monitoring configurations

Figure 4.1: Monitoring configurations for both test cases. (a) Top view showing the spatial arrangement of monitoring points (red) and injection wells (black) for both 2D conceptual model (top row) and Brugge model (bottom row). Linear arrays contain 51 points (2D) and 139 points (Brugge), low-density grids contain 36 points, and high-density grids contain 121 points. (b) Side view illustrating the vertical relationship between surface monitoring points and the subsurface reservoir, with injection wells extending from the surface to the reservoir.

DATA TYPES

We evaluated three data collection strategies:

- **Surface Displacement Only**: Vertical displacement measurements at the surface, typically collected using tiltmeters, GPS stations, or InSAR. These measurements provide spatially distributed information about subsurface pressure and temperature changes.
- **Bottom Hole Pressure (BHP) Only**: Pressure measurements collected at the injection well at all time steps, providing direct but spatially limited information about reservoir conditions.
- **Combined Measurements**: Integration of both surface displacement and BHP data, representing a multi-physics monitoring approach that leverages complementary data sources.

The relationship between surface monitoring points and the subsurface reservoir is illustrated in Figure 4.1, showing how surface measurements relate to the underlying reservoirs.

Table 4.2 presents our experimental design, with monitoring configurations tested across both models. This approach allows us to quantitatively compare different monitoring strategies and assess the relative contribution of spatial coverage, temporal frequency, and data types to uncertainty reduction in CO_2 plume characterization.

Table 4.2: Summary of Experimental Configurations and Data Points

Spatial Configuration	Measurement Type	2D Model		Brugge Model	
		ID	Points	ID	Points
Linear array	Final-time	1A	51	2A	139
	Multiple-time	1B	153	2B	417
	Multiple-time + BHP	1C*	193	2C	457
Low-density grid	Final-time	1E	36	2E	36
	Multiple-time	1F	108	2F	108
	Multiple-time + BHP	1G*	148	2G	148
High-density grid	Final-time	1H	121	2H	121
	Multiple-time	1I	363	2I	363
	Multiple-time + BHP	1J*	403	2J	403
BHP only	All time steps	1D*	40	2D	40

* Configurations with BHP data for the 2D model were implemented with localization.

The total number of data points varies from 36 (low-density grid, single time) to 457 (linear array with BHP for Brugge model), allowing us to systematically evaluate different monitoring configurations. This experimental setup helps us see how each monitoring strategy changes our understanding of underground CO_2 plumes. Measuring the information gained from each monitoring setup lets engineers compare options and choose more effective plans. For clarity and computational economy we use a single reference realization for each test case.

4.3.3. Prior ensemble generation

Following the approach described in Section 3.2.3, we generated ensembles of permeability fields. For each test case, we created 100 prior permeability realizations, with an additional realization designated as the reference “true” model used to generate synthetic observations. For the 2D conceptual model, we defined the reference model with a uniform log-permeability of log(6) mD. The prior ensemble comprises spatial variations from log(3) mD to log(9) mD in the mean permeability. These variations were spatially correlated using the modified Gaspari-Cohn function detailed in Section 3.2.3. Figure 4.2 illustrates the reference permeability field alongside three sample realizations, demonstrating the variability in permeability distributions.

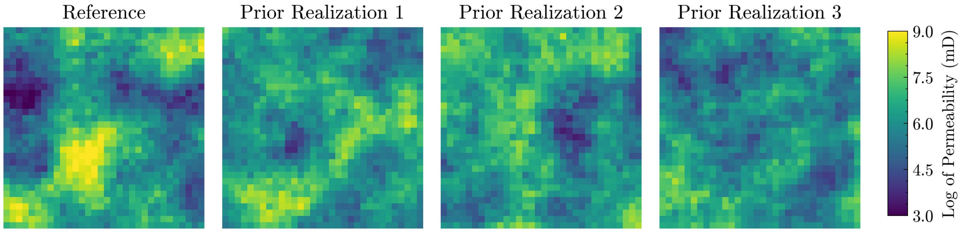


Figure 4.2: Reference model and three sample realizations of the prior permeability distribution for the 2D conceptual model.

For the Brugge model, we used geostatistical characteristics observed in previous studies of this benchmark dataset (Peters et al., 2010; Saifullin et al., 2024). The mean permeability was set to 540 mD, while maintaining the realistic heterogeneity patterns characteristic of this field. We applied the same ensemble generation technique as for the 2D model. Figure 4.3 displays the reference permeability field alongside three prior realizations.

By quantifying the reduction in uncertainty from this initial state, as measured through Shannon entropy reduction, described in Section 4.2.3, we can evaluate how effectively each monitoring configuration constrains our knowledge of the subsurface in terms of the CO_2 plume distance from the well. Specifically, we calculate the relative entropy reduction between prior and posterior ensemble distributions, which represents the percentage of initial uncertainty eliminated through data assimilation. This quantitative approach enables the comparison of monitoring strategies based on their information content.

Figure 4.6 shows the reference displacement fields from CO_2 injection in both models. The 2D model exhibits a circular pattern centered at the injection well, while the Brugge model shows an asymmetric pattern reflecting its complex geological structure. Figure 4.4 presents the CO_2 molar fraction distribution at the end of the 10-year injection period for both reference models. In both cases, the CO_2 plumes show asymmetric patterns due to the heterogeneity of the reservoir models. The plumes have higher concentrations near the injection wells, with maximum values of approximately 0.024 for the 2D model and 0.025 for the Brugge model.

Figure 4.5 shows the evolution of average reservoir pressure throughout the 10-year CO_2 injection period for both models. The pressure increase is a primary driver of the observed surface displacements, with both models showing characteristic pressure buildup

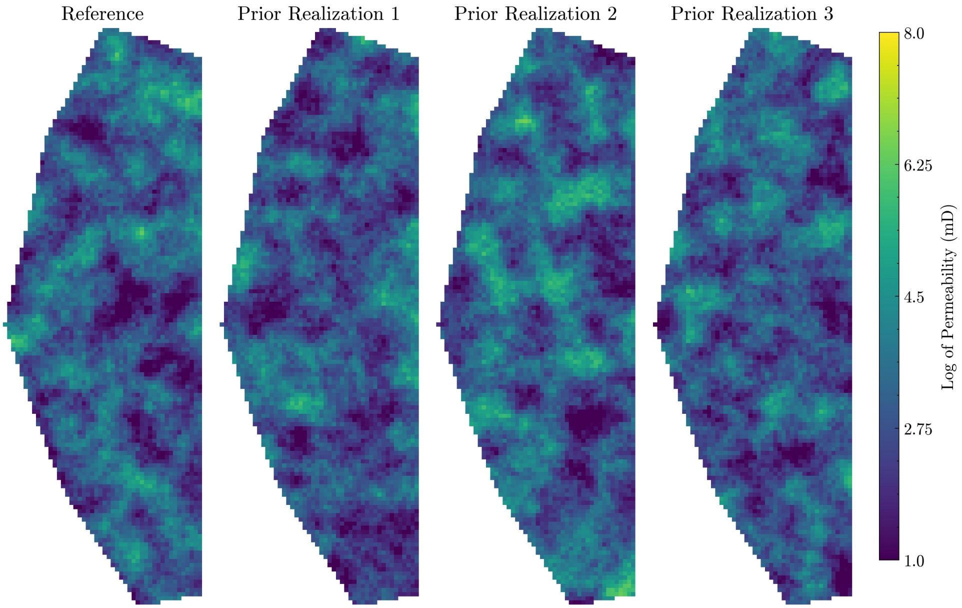


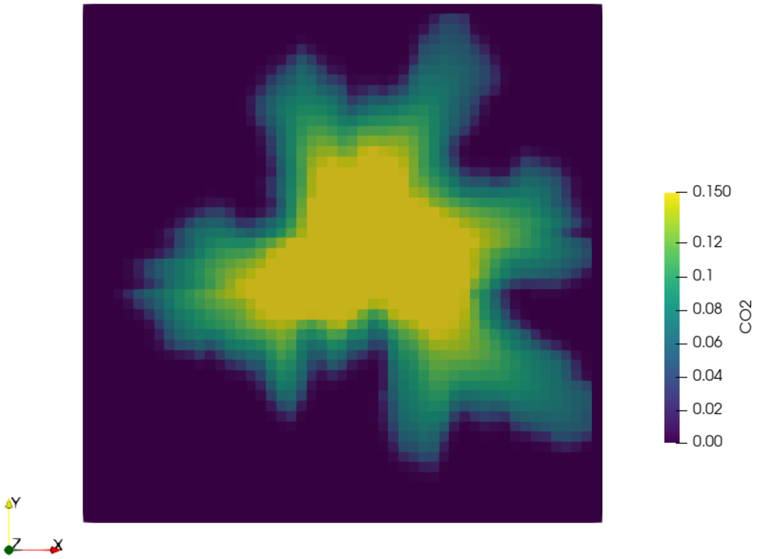
Figure 4.3: Reference model and three sample realizations of the prior permeability distribution for the Brugge model.

patterns. The 2D conceptual model exhibits a relatively steep initial pressure increase followed by a more gradual rise as injection continues, reaching a maximum of approximately 123.3 bars around year 7-8, after which it begins to decline despite continued CO_2 injection. This decline occurs due to the open boundary conditions allowing pressure relief as the pressure front reaches the model boundaries. In contrast, the Brugge model shows a continual pressure increase throughout the simulation period, reaching approximately 155 bars by the end of the injection period. This difference occurs because the Brugge model is larger than the conceptual model, and during the 10-year simulation period, the pressure effects from injection remain predominantly within the reservoir area without substantial interaction with the model boundaries. Figure 4.6 shows the reference displacement fields from CO_2 injection in both models. The 2D model exhibits a circular pattern centered at the injection well, while the Brugge model shows an asymmetric pattern reflecting its complex geological structure.

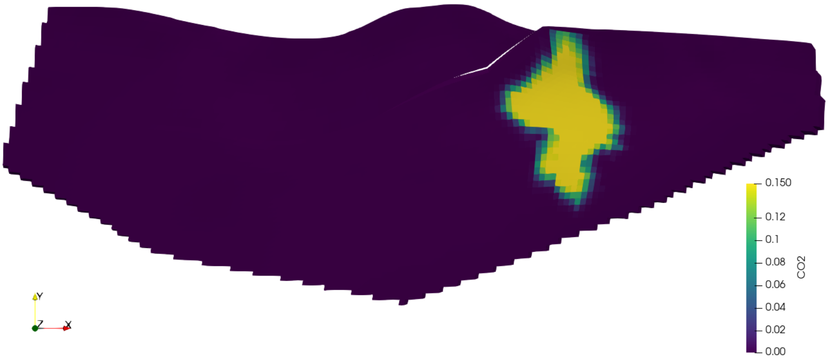
All monitoring strategies are therefore evaluated against the same ’ground-truth’ data, so the relative ranking of information gain reflects the monitoring design alone. Extending the analysis to a suite of alternative truths is straightforward but outside the scope of the present study.

4.4. MONITORING STRATEGY DATA ASSIMILATION

In this section, we present the findings from our evaluation of monitoring strategies. Due to the large number of experimental configurations tested (different combinations

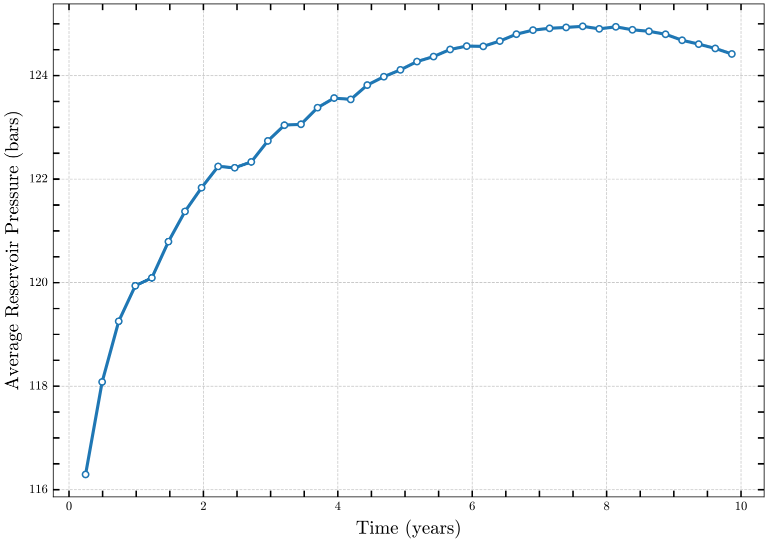


(a) 2D conceptual model CO_2 plume

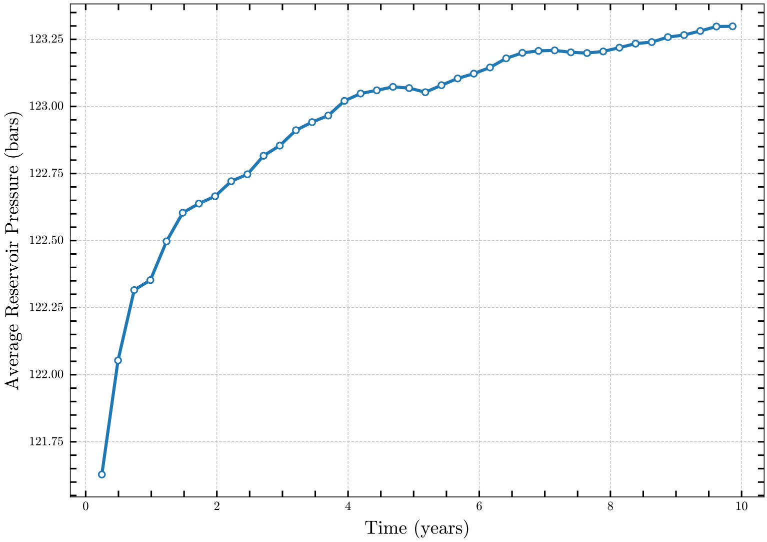


(b) Brugge field model CO_2 plume

Figure 4.4: CO_2 molar fraction distribution at the end of the 10-year injection period for reference models.

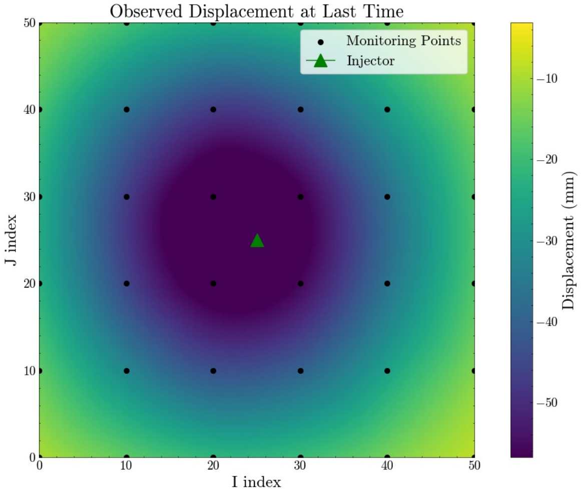


(a) 2D conceptual model

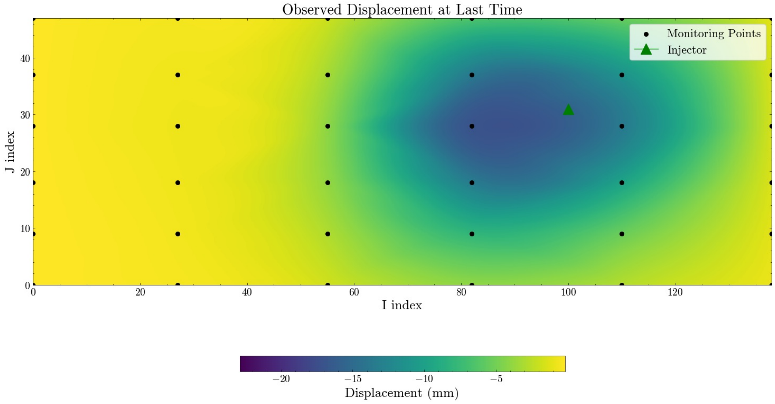


(b) Brugge field model

Figure 4.5: Average reservoir pressure evolution during the 10-year CO_2 injection period for both models. The pressure buildup directly affects the surface displacement response observed at the surface.



(a) 2D conceptual model



(b) Brugge field model

Figure 4.6: Displacements simulated in the reference models at the final time step.

of spatial arrangements, temporal sampling schemes, and data types), we showcase representative results rather than exhaustive plots for all cases. Selected visualizations illustrate the most significant patterns and outcomes, while comprehensive performance metrics are provided through normalized RMSE comparisons (Figure 4.16) and entropy reduction values (Table 4.4).

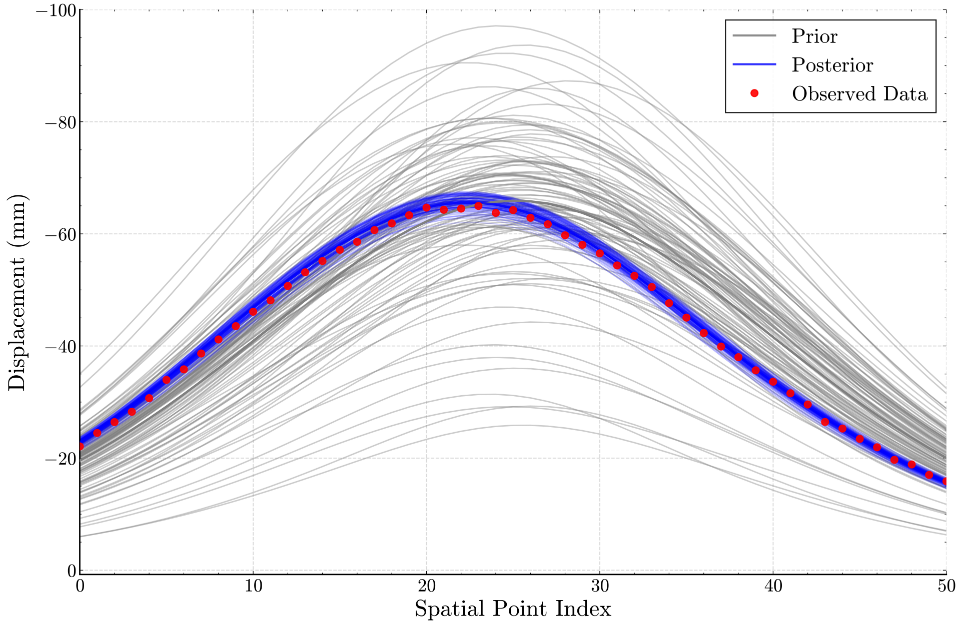
4.4.1. DATA ASSIMILATION OF LINEAR ARRAYS OF SENSORS

We first examined the configuration consisting of a linear array of surface displacement measurements at the last time step (Configuration 1A for the 2D model and 2A for the Brugge model). Figure 4.7 demonstrates how ESMDA updates the ensemble of permeability realizations to better match observed displacement data. For the 2D conceptual model, the posterior ensemble shows convergence toward observed values, while maintaining a spread that reflects the remaining uncertainty and the Brugge model exhibits similar behavior.

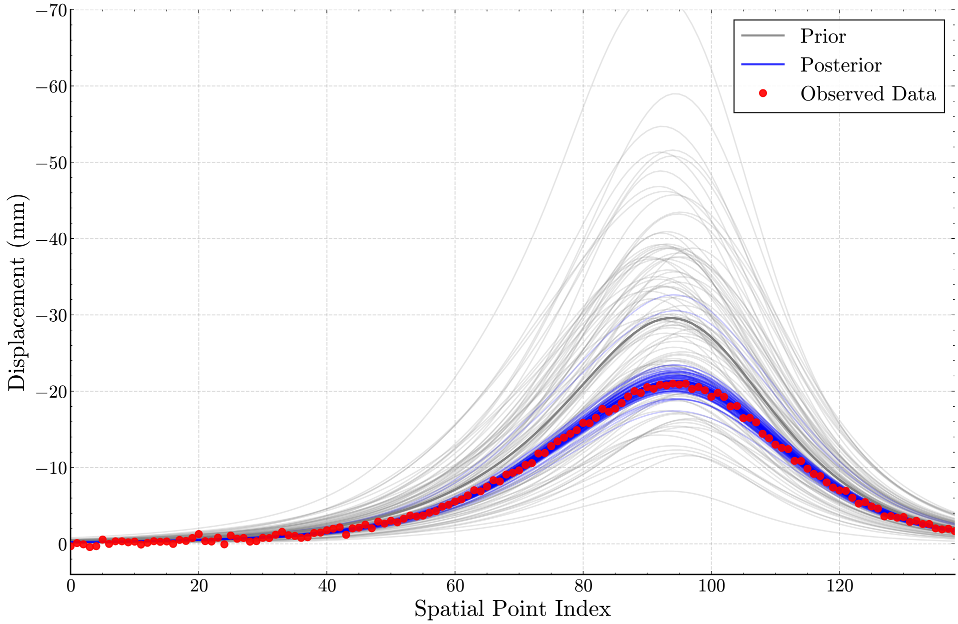
The impact of the linear array monitoring configuration on CO_2 plume characterization is visualized through probability maps in Figure 4.8. In an effective data assimilation process, we expect to see two characteristics in these probability maps: (1) a reduction in uncertainty evidenced by more focused and defined plume boundaries in the posterior compared to the prior, and (2) preservation of reasonable uncertainty reflecting the inherent limitations of indirect surface measurements in constraining subsurface properties. An ideal outcome balances uncertainty reduction with realistic preservation of variability, avoiding both excessive confidence (overly sharp probability transitions suggesting unrealistic certainty) and insufficient constraint (minimal difference between prior and posterior). To create these maps, we analyzed the CO_2 molar fraction distribution across all ensemble members, calculating for each grid cell the probability of exceeding a threshold value by counting the fraction of ensemble members that exceed this threshold at that location.

For the 2D conceptual model (Fig. 4.8a), the posterior probability map shows more concentration in plume extent compared to the prior, with the highest probability region better defined around the injection well. It demonstrates meaningful information gain from the assimilated data. Simultaneously, the probability distribution maintains a gradual transition from high to low values rather than collapsing to binary certainty, appropriately reflecting the remaining uncertainty inherent when using surface deformation to constrain deep subsurface properties. The Brugge model (Fig. 4.8b) shows a more modest refinement of the probability distribution, with the posterior maintaining much of the spatial extent seen in the prior. This less pronounced improvement appropriately reflects the greater challenge in constraining the CO_2 plume in this more complex reservoir. As evident in Figure 4.3, the Brugge model covers a significantly larger area (17×9 km versus 4×4 km for the 2D model) and features complex geological structures including variable reservoir depth (1545-1945 m) and an irregular grid configuration that follows the natural topography of the reservoir. These physical complexities create a more challenging inverse problem that requires more comprehensive monitoring to effectively constrain it.

The entropy reduction analysis for these simple configurations reveals differences in how effectively this monitoring approach constrains model uncertainty. For the 2D

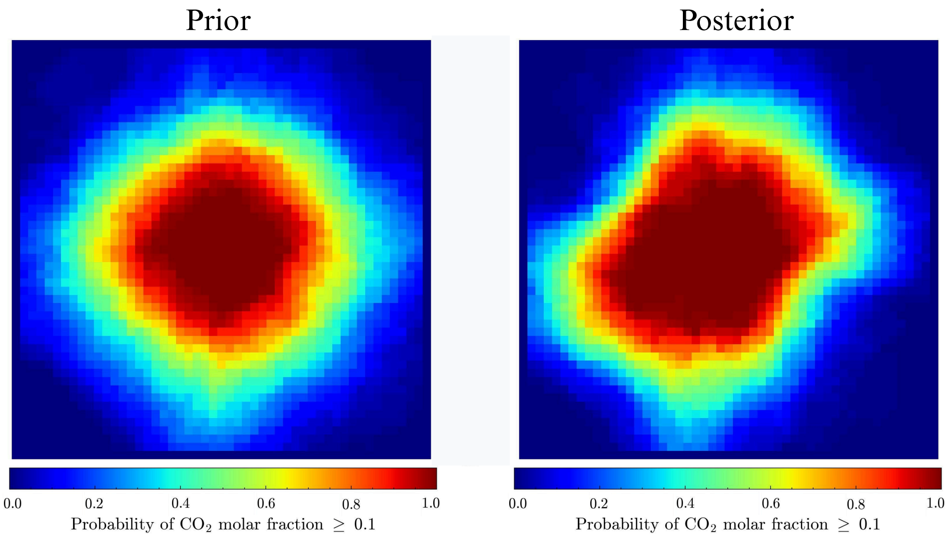


(a) 2D conceptual model (Configuration 1A).

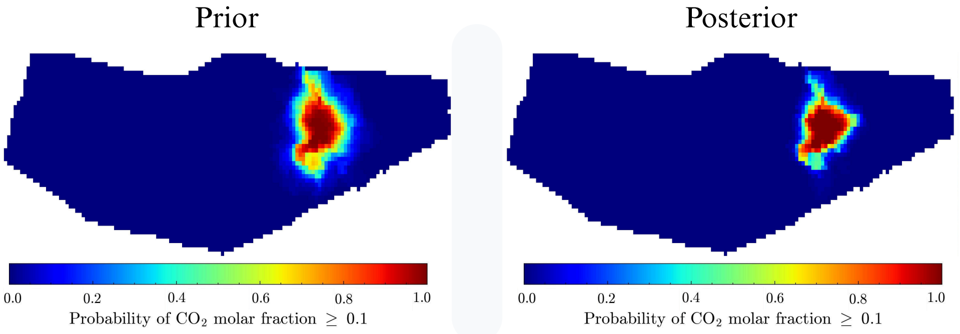


(b) Brugge model (Configuration 2A).

Figure 4.7: Data assimilation results showing surface displacement at the final time step for both models with prior ensemble (gray), posterior ensemble (blue), and observed data (red dots).



(a) 2D conceptual model (Configuration 1A).



(b) Brugge model (Configuration 2A).

Figure 4.8: Probability maps of CO_2 molar fraction exceeding threshold value for linear array configurations. Left panels show prior probabilities; right panels show posterior probabilities.

conceptual model, the relative entropy reduction indicates that 50% of the prior entropy in CO_2 plume characterization was eliminated through data assimilation. The Brugge model achieved a relative entropy reduction of only 0.20, demonstrating how the effectiveness of basic monitoring strategies varies with geological complexity. The lower entropy reduction in the Brugge model correlates with the more limited changes observed in its probability maps, confirming that more complex geological structures can require more comprehensive monitoring systems to achieve comparable uncertainty reduction.

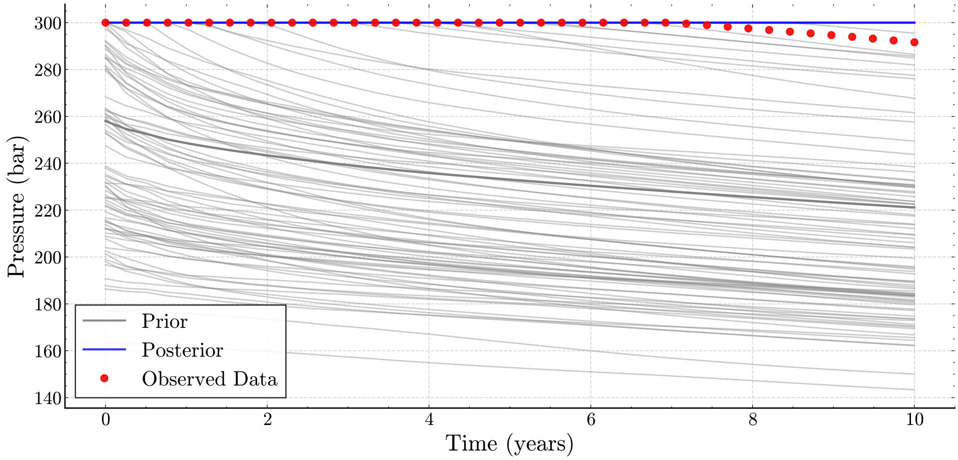
4.4.2. BOTTOM HOLE PRESSURE DATA ASSIMILATION

We next evaluate the information content of Bottom Hole Pressure (BHP) measurements alone (Configuration 1D for the 2D model and 2D for the Brugge model). BHP data provide direct information about pressure conditions at the injection point but lack the spatial coverage of surface monitoring systems. For the 2D conceptual model, localization is necessary when assimilating BHP data to prevent ensemble collapse and preserve uncertainty in the posterior ensemble. In contrast, the Brugge model doesn't require the addition of this artificial constraint as ensemble collapse was not observed. Since the Brugge model maintained an appropriate variance structure without experiencing ensemble collapse, we did not apply localization to avoid introducing unnecessary artificial constraints in the data assimilation process.

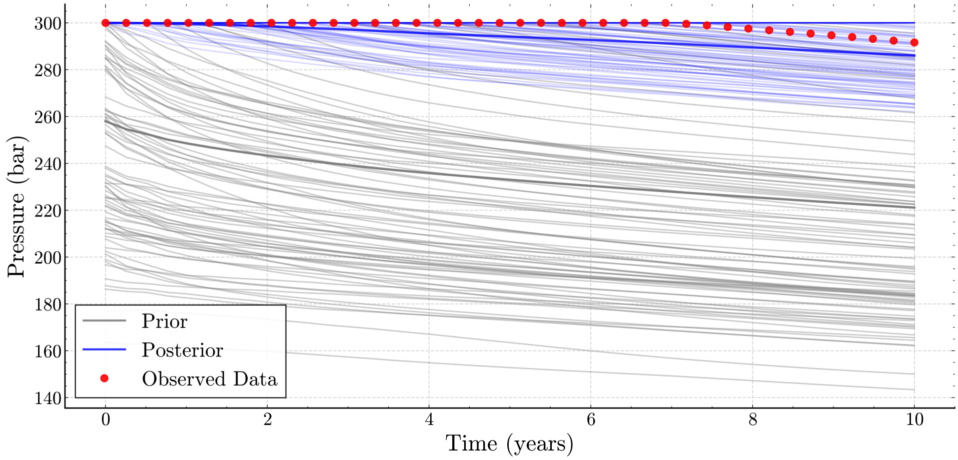
An expected feature worth highlighting in the BHP profiles is the decreasing trend despite continuous CO_2 injection. While the average reservoir pressure increases during injection, as seen previously, the BHP at the injection well shows a decline over time. This behavior occurs because as CO_2 displaces water near the wellbore, the relative permeability of CO_2 increases. This improved mobility allows CO_2 to enter the formation more easily, requiring less pressure at the injection point to maintain the same injection rate. This phenomenon highlights the importance of accurately modeling multiphase flow effects in CO_2 storage simulations.

Figure 4.9 shows the assimilation results for BHP data in both models, including the effect of localization on the 2D conceptual model. Without localization (Fig. 4.9a), the 2D model ensemble experiences extreme variance reduction. With localization applied (Fig. 4.9b), the posterior ensemble maintains appropriate diversity while still improving the match to observations. The Brugge model (Fig. 4.9c) matches the observed BHP data without requiring localization.

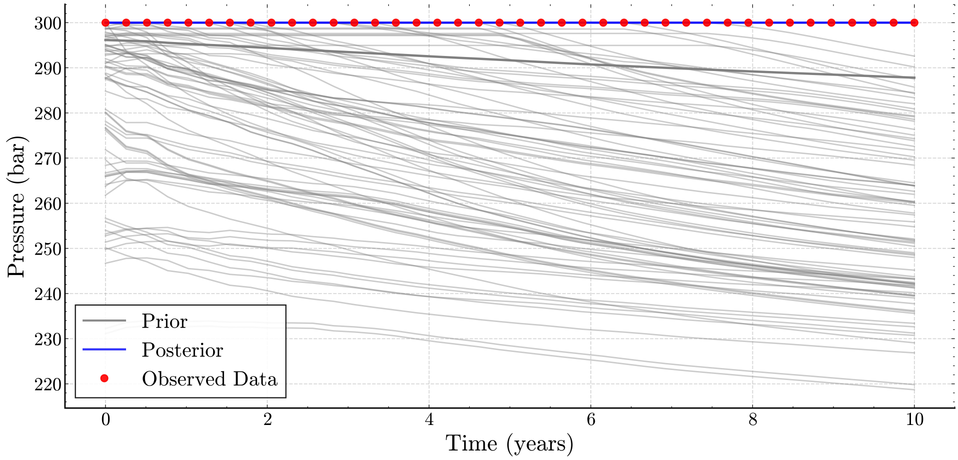
The impact of localization on the characterization of CO_2 plumes becomes apparent when examining the probability maps of CO_2 molar fraction exceeding the 10% threshold value. Figure 4.10 presents these maps for different BHP-only data assimilation scenarios. For the 2D model without localization (Fig. 4.10a), the posterior probability map collapses completely, showing a degenerate pattern with unrealistic concentration of probability in a few isolated cells. This severe ensemble collapse renders the uncertainty quantification meaningless and highlights why localization was necessary for this model. With localization applied to the 2D model (Fig. 4.10b), the posterior ensemble maintains an appropriate spread of uncertainty while still providing meaningful constraints on the CO_2 plume. The Brugge model (Fig. 4.10c) does not exhibit this collapse behavior even without localization, naturally preserving a reasonable uncertainty structure in the posterior due to its inherent geological heterogeneity. These probability maps validate our approach of



(a) 2D model without localization.



(b) 2D model with localization.



(c) Brugge model without localization.

Figure 4.9: Assimilation results for BHP-only data with prior ensemble (gray), posterior ensemble (blue), and observed data (red dots).

applying localization selectively based on model characteristics, using it when there is a suspicion of ensemble degeneracy.

The entropy reduction analysis for BHP-only data assimilation reveals limited effectiveness compared to surface deformation measurements. For the 2D conceptual model, BHP data achieved a relative entropy reduction of 0.25, significantly lower than the 0.50 achieved with linear array surface measurements. The Brugge model showed an even smaller relative entropy reduction of 0.10. The limited effectiveness comes from the way BHP measurements are taken using specific points close to the wells, which provide accurate constraints near the well but lack information about spatial variations across the reservoir.

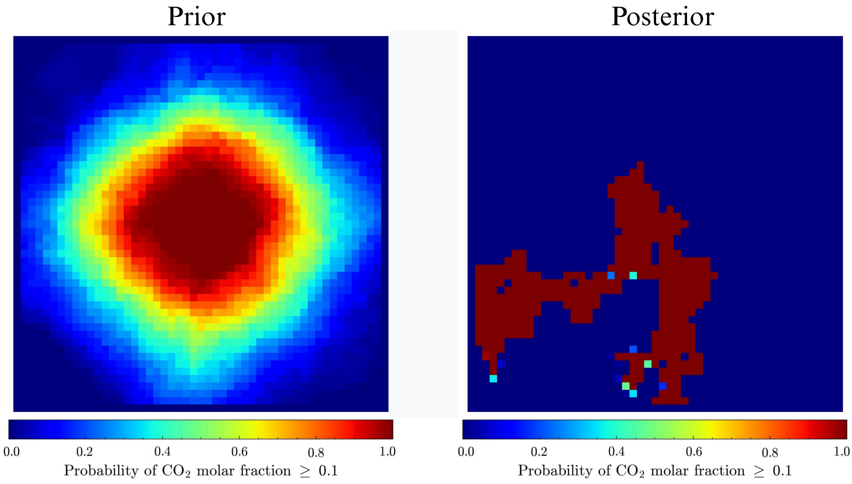
4.4.3. EFFECT OF SPATIAL CONFIGURATION

Having established the baseline performance of linear arrays and BHP-only measurements, we next examine how different spatial configurations affect data assimilation performance. We compare linear arrays (already presented in Figure 4.7), low-density grids (36 points), and high-density grids (121 points) using single-time measurements at the final time step (Configurations 1A/2A, 1E/2E, and 1H/2H). Figure 4.11 shows the displacement differences analysis and observations for the 2D conceptual model with different spatial monitoring configurations. The bubble plots illustrate the spatial distribution of errors between their ensemble predictions and observations, with bubble size corresponding to the magnitude of the error. For each configuration, the comparison between prior and posterior ensembles demonstrates how assimilation reduces uncertainty across the monitoring grid.

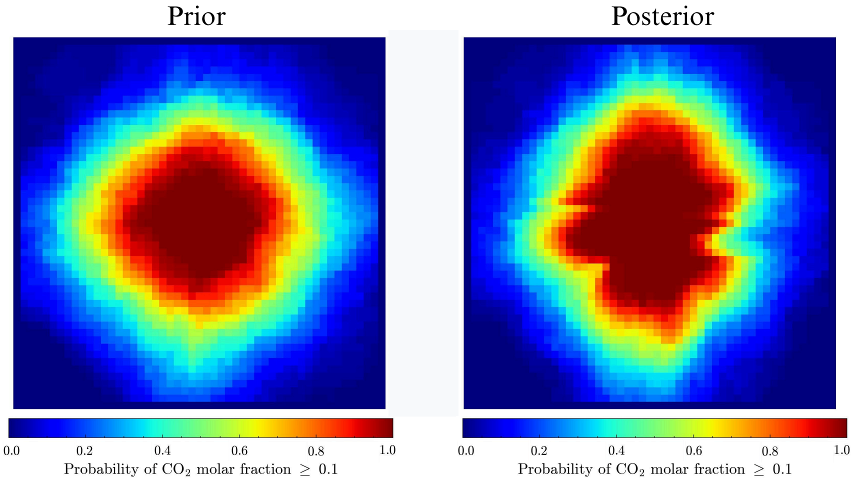
The low-density grid (Configuration 1E) achieved a relative entropy reduction of 0.64, which is higher than the 0.50 reduction obtained with the linear array (Configuration 1A) despite having fewer measurement points (36 versus 51). This demonstrates that the strategic spatial distribution of monitoring points can be more important than simply maximizing their number. The high-density grid (Configuration 1H) achieved the highest entropy reduction among single-time configurations at 0.68. For the 2D conceptual model, we observed increasingly smaller improvements with denser configurations. For the Brugge model, we observe different patterns in monitoring effectiveness due to its elongated reservoir geometry and well position. Figure 4.12 shows the error distribution for different spatial configurations in the Brugge model.

The low-density grid (Configuration 2E) achieved a relative entropy reduction of only 0.15, which is less than the linear array (Configuration 2A) at 0.20 despite covering a broader area. This counter-intuitive result stems from the elongated structure of the Brugge field (17 km × 9 km). Similarly, the high-density grid (Configuration 2H) achieved an entropy reduction of 0.18, which still falls short of the linear array despite using almost the same number of measurement points. This further reinforces the importance of aligning monitoring configurations with the particular case and the reservoir's geometry.

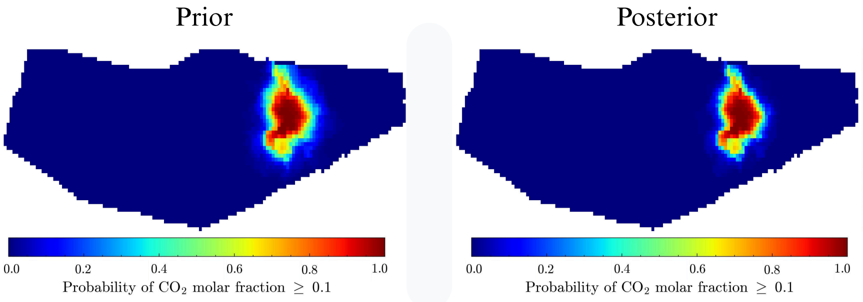
These observations highlight how reservoir characteristics influence monitoring effectiveness in different geological settings. The bubble plots clearly illustrate the spatial distribution of uncertainty reduction after data assimilation, with the posterior ensembles showing smaller error bubbles compared to the prior ensemble members across all configurations. When examining the amplitude of entropy reduction, we observe



(a) 2D model without localization

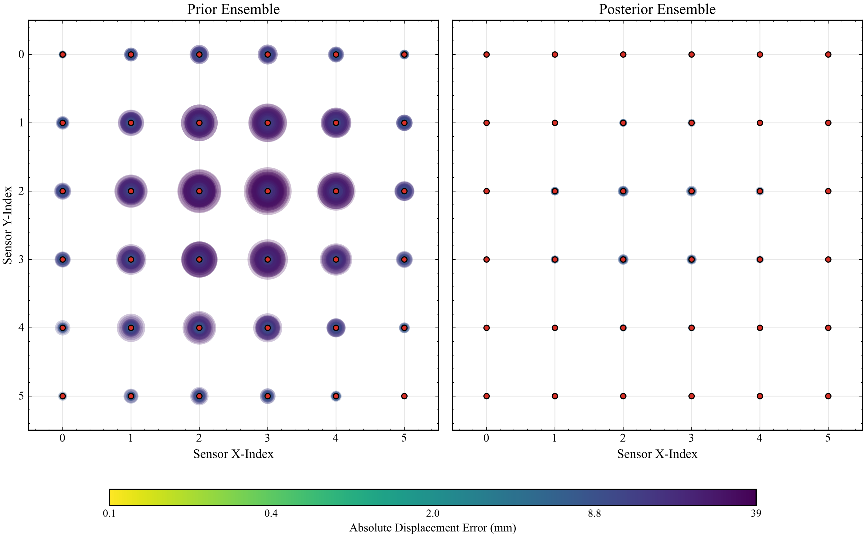


(b) 2D model with localization

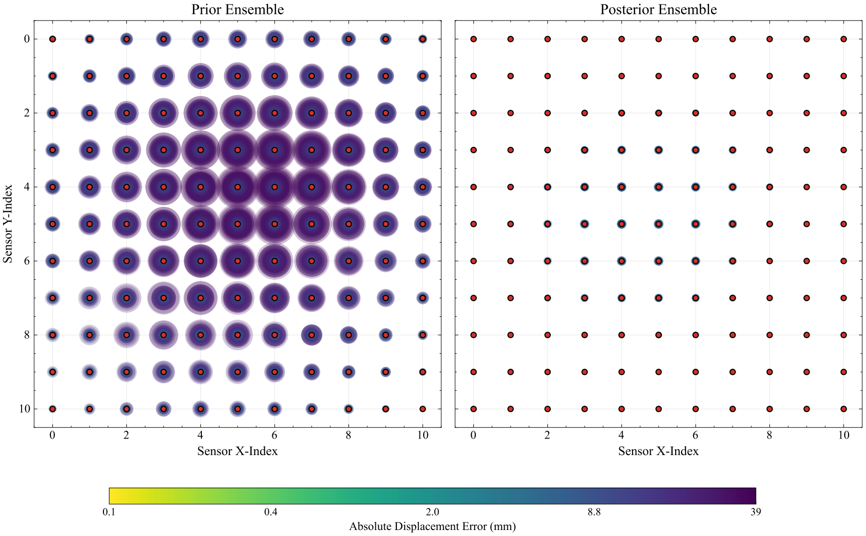


(c) Brugge model without localization

Figure 4.10: Probability maps of CO_2 molar fraction exceeding threshold value. Left panels show prior probabilities; right panels show posterior probabilities.

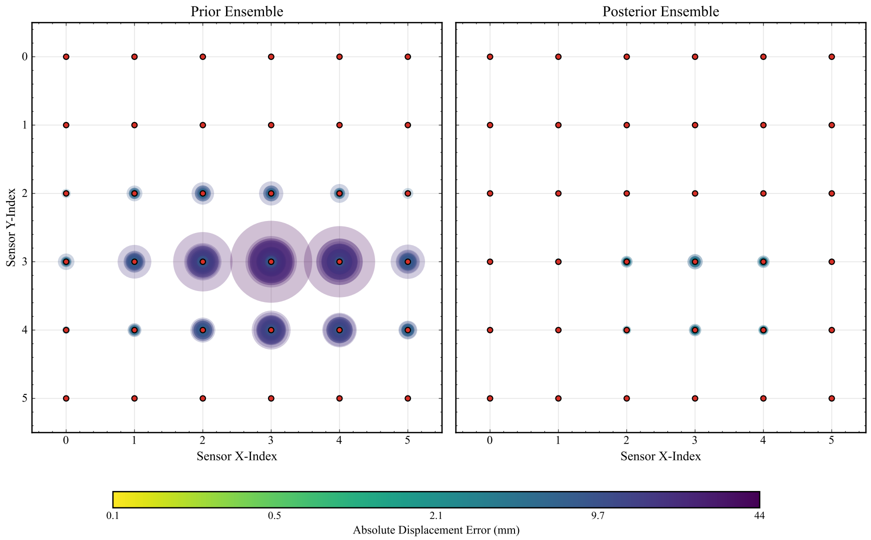


(a) Low-density grid (36 points)

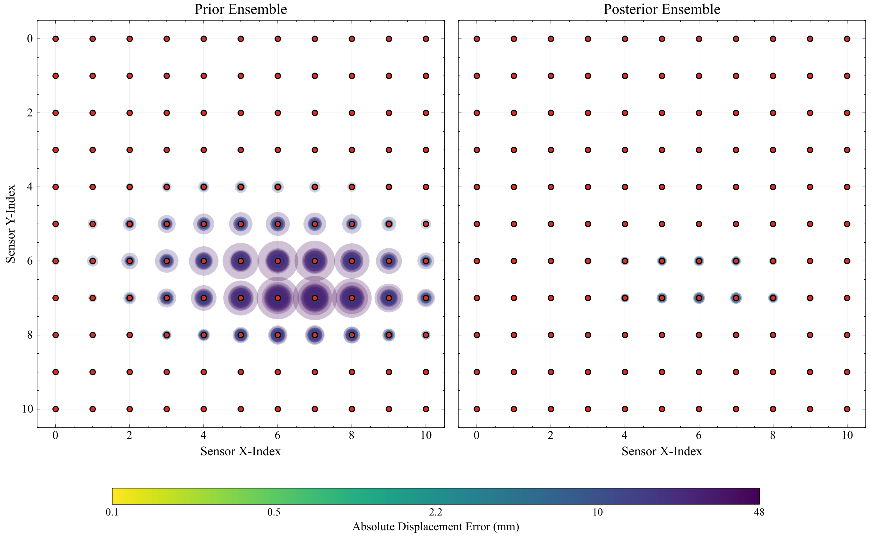


(b) High-density grid (121 points)

Figure 4.11: Absolute displacement error for different monitoring configurations in the 2D conceptual model. Left panels show prior ensemble errors; right panels show posterior ensemble errors after data assimilation.



(a) Low-density grid (36 points)



(b) High-density grid (121 points)

Figure 4.12: Absolute displacement error for different monitoring configurations in the Brugge model. Left panels show prior ensemble errors; right panels show posterior ensemble errors after data assimilation.

substantial differences between the two models: the 2D conceptual model showed higher overall entropy reduction compared to the Brugge model. This suggests that simpler, more symmetric reservoirs may be more effectively constrained by surface measurements than complex heterogeneous formations. The spatial pattern of uncertainty reduction also differs between configurations. In both models, the greatest reduction occurs near areas of maximum surface displacement rather than directly above the injection wells. This is particularly evident in Figure 4.11 and Figure 4.12, where the largest bubbles in the prior ensemble are concentrated in regions of peak displacement, and these same regions show the larger reduction in bubble size after assimilation. The contrast between our test cases shows that grid-based arrangements provided superior information gain in the symmetric 2D model, while the linear array aligned better with the dominant flow directions in the elongated Brugge reservoir. This illustrates and emphasizes that design monitoring systems require consideration of reservoir-specific characteristics rather than applying generic designs.

4.4.4. EFFECT OF TEMPORAL SAMPLING FREQUENCY

Temporal frequency of observations enhances the performance of data assimilation across all spatial configurations. Multiple-time measurements taken at years 1, 5, and 10 improve uncertainty reduction compared to single-time measurements at the final time step. The comparative analysis between multiple-time measurements (Configuration 1B) and single-time configuration (1A) for linear arrays illustrates this effect. Figure 4.13 illustrates how the assimilation process captures both spatial patterns and temporal evolution of displacements across all three time steps, demonstrating how tracking deformation progression provides more comprehensive subsurface characterization than end-point measurements alone. This temporal dimension adds constraints to the inverse problem, allowing better discrimination between different permeability realizations that might otherwise produce similar final-state observations.

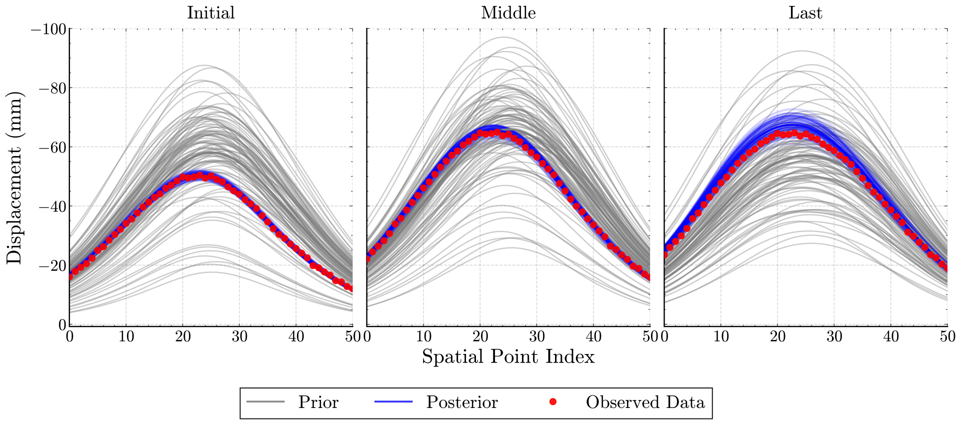


Figure 4.13: Displacement curves for linear array with multiple time steps (Configuration 1B) for the 2D model at years 1, 5, and 10. The plots show vertical displacements along the West-East transect crossing the injection well (see Figure 4.1a). Prior ensemble (gray), posterior ensemble (blue), and observed data (red dots).

The entropy reduction for the linear array improved from 0.50 (Configuration 1A) to 0.67 (Configuration 1B) with the addition of temporal measurements. This improvement demonstrates how tracking the evolution of surface deformations over time provides additional constraints on subsurface permeability and resulting CO_2 migration patterns. The Brugge model exhibited even greater sensitivity to temporal sampling. Figure 4.14 shows representative displacement results for the linear array configuration with multiple time steps.

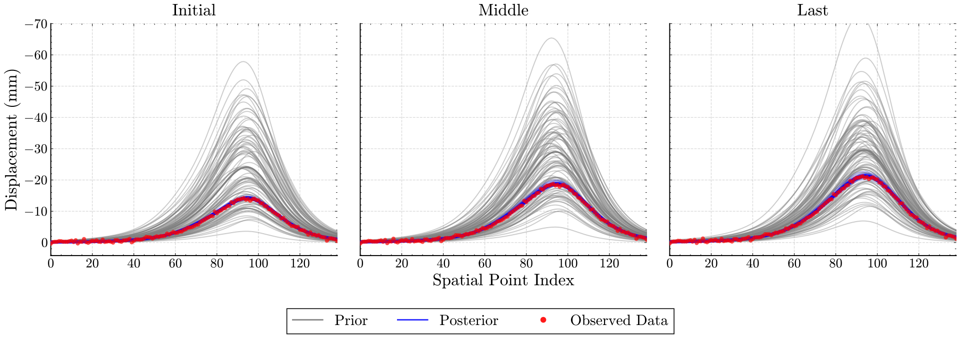


Figure 4.14: Displacement curves for linear array with multiple time steps (Configuration 2B) for the Brugge model at years 1, 5, and 10. The plots show vertical displacements along the West-East transect crossing the injection well (see Figure 4.1a). Prior ensemble (gray), posterior ensemble (blue), and observed data (red dots).

For the Brugge model's linear array, entropy reduction increased from 0.20 (Configuration 2A) to 0.63 (Configuration 2B). The high-density grid showed even more dramatic enhancement, from 0.18 (Configuration 2H) to 0.72 (Configuration 2I). This heightened sensitivity to temporal data in the more complex Brugge model suggests that heterogeneous reservoirs particularly benefit from time-series monitoring. This result is not surprising because the behavior of the conceptual models is easier to predict, hence the importance of adding observations becomes more significant in more realistic and complex cases. Similar improvements were observed across all spatial configurations. For the 2D model, the low-density grid entropy reduction increased from 0.64 (Configuration 1E) to 0.79 (Configuration 1F). The high-density grid increased from 0.68 (Configuration 1H) to 0.84 (Configuration 1I). For the Brugge model's low-density grid, entropy reduction increased from 0.15 (Configuration 2E) to 0.55 (Configuration 2F). An important insight from these results is that temporal sampling can be as influential as spatial density in reducing uncertainty. For example, in the 2D model, the low-density grid with multiple time measurements (Configuration 1F, 108 points) achieved greater entropy reduction (0.79) than the high-density grid with single-time measurements (Configuration 1H, 121 points, reduction of 0.68) despite using a lower total number of data points. This suggests that in some geological settings, operators may be able to compensate for limitations in spatial coverage by increasing measurement frequency over time. While the specific trade-offs would be site-dependent, these findings indicate a potential flexibility in monitoring design where practical constraints on spatial deployment might be offset by strategic temporal sampling. Such insights could prove useful for operators balancing monitoring effectiveness with logistical and economic considerations.

4.4.5. INTEGRATION OF BHP WITH SURFACE DISPLACEMENT DATA

Combining BHP data with surface displacement measurements enhances entropy reduction across all spatial configurations. For all spatial arrangements, the integration of BHP data provided modest but consistent improvement in entropy reduction. For the 2D conceptual model with a linear array, the addition of BHP data (Configuration 1C) increased entropy reduction from 0.67 to 0.71 compared to displacement data alone (Configuration 1B). Similar improvements were observed for the low-density grid, where entropy reduction increased from 0.79 (Configuration 1F) to 0.81 (Configuration 1G), and for the high-density grid, with improvement from 0.84 (Configuration 1I) to 0.86 (Configuration 1J). The Brugge model showed comparable enhancements when incorporating BHP data. For the linear array, entropy reduction increased from 0.63 (Configuration 2B) to 0.68 (Configuration 2C). The low-density grid improved from 0.55 (Configuration 2F) to 0.61 (Configuration 2G), and the high-density grid from 0.72 (Configuration 2I) to 0.79 (Configuration 2J). BHP and surface displacement measurements provide complementary information about subsurface processes. BHP data offer precise constraints on pressure conditions at the well, while surface displacements provide broader spatial coverage. Together, they enable more robust characterization of the CO_2 plume, capturing both localized and widespread subsurface dynamics that neither monitoring strategy could comprehensively reveal on its own.

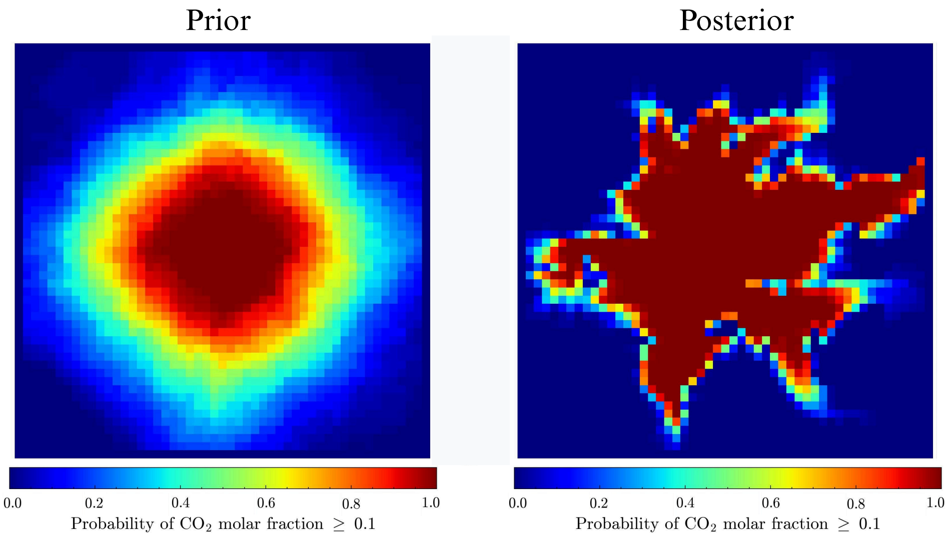
4.4.6. INTEGRATING ALL AVAILABLE DATA

The most complex monitoring configurations combined high spatial density, multiple temporal measurements, and integration of both surface displacement and BHP data. Figure 4.15 shows the probability maps for these configurations in both models. As described in Section 4.4.1, these probability maps are generated by calculating, for each grid cell, the probability of CO_2 molar fraction exceeding a threshold value across all ensemble members. While these probability maps illustrate the spatial distribution of information gain, a consolidated analysis of the normalized RMS error across all configurations will be presented at the end of this section to provide a comprehensive quantitative performance assessment.

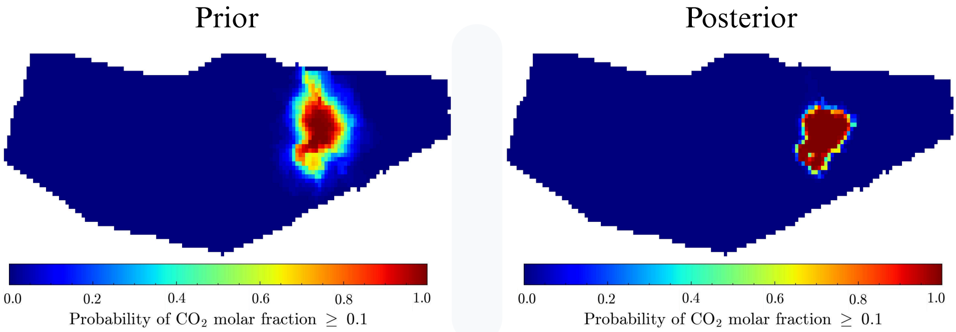
For the 2D model, Configuration 1J achieved a relative entropy reduction of 0.86, the highest across all configurations. This represents a substantial improvement over the baseline linear array with single-time measurements (Configuration 1A, reduction of 0.50). Similarly, the comprehensive strategy for the Brugge model (Configuration 2J) reached a relative entropy reduction of 0.79, compared to 0.20 for the baseline configuration.

To provide a comprehensive assessment of data assimilation performance across all configurations, Figure 4.16 and Table 4.3 present the normalized Root Mean Square Error (RMSE) for both the 2D conceptual and Brugge models. This normalized metric enables comparison between different data types (displacement measured in millimeters and BHP in bars) and configurations with varying numbers of measurement points.

The error metrics show that data assimilation reduced the normalized RMSE across all experimental configurations, with all posterior values lower than their corresponding prior values. This normalization was performed by dividing each configuration's RMSE by its prior RMSE value, allowing for direct comparison between different monitoring strategies. The Brugge model consistently shows higher normalized RMSE values than



(a) 2D model: High-density grid, multi-time + BHP (Configuration 1J).



(b) Brugge model: High-density grid, multi-time + BHP (Configuration 2J).

Figure 4.15: Probability maps of CO_2 molar fraction for comprehensive monitoring strategies. Left panels show prior probabilities; right panels show posterior probabilities.

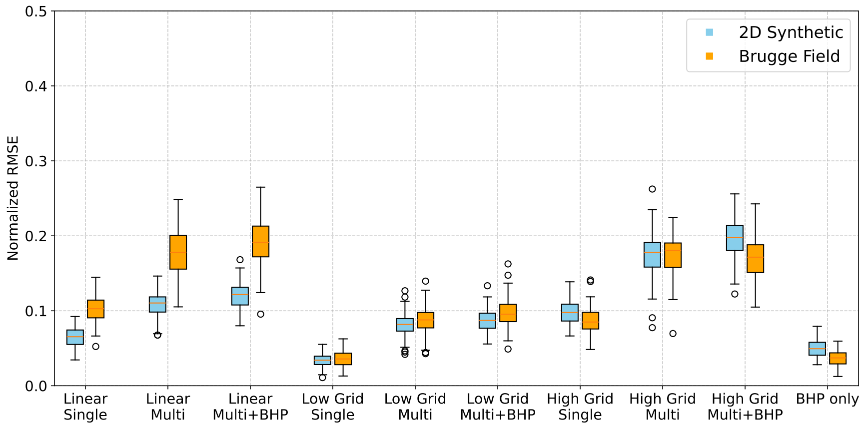


Figure 4.16: Normalized RMSE comparison between 2D Synthetic (blue) and Brugge Field (orange) models across different monitoring configurations. Box plots show the distribution of errors after data assimilation.

Table 4.3: Normalized RMSE Comparison Across Monitoring Configurations

Configuration	**2D Model**		**Brugge Model**	
	Points	**RMSE**	**Points**	**RMSE**
Linear Array (Final-Time)	51	0.08	139	0.13
Linear Array (Multiple-Time)	153	0.12	417	0.19
Linear Array (Multiple-Time + BHP)	193	0.14	457	0.21
Low-Density Grid (Final-Time)	36	0.06	36	0.06
Low-Density Grid (Multiple-Time)	108	0.10	108	0.10
Low-Density Grid (Multiple-Time + BHP)	148	0.12	148	0.12
High-Density Grid (Final-Time)	121	0.11	121	0.11
High-Density Grid (Multiple-Time)	363	0.18	363	0.18
High-Density Grid (Multiple-Time + BHP)	403	0.20	403	0.20
BHP only	40	0.04	40	0.04

the 2D conceptual model, reflecting the greater challenge in matching observations in a reservoir with more complex geometry. The "Low Grid Single" configuration exhibits lower error values for both models, while the BHP-only configuration shows moderate performance. The BHP-only run fits its own sparse data very closely (low RMSE) but reduces plume entropy by only 10 %, so its overall information gain is moderate when compared with the displacement-based configurations. These quantitative error metrics, combined with the entropy reduction analysis presented below, provide a systematic comparison of different observational strategies for CO_2 plume characterization, highlighting the relative effectiveness of various spatial configurations and temporal sampling approaches.

Entropy reduction metrics were employed to quantify the effectiveness of different monitoring configurations, offering an evaluation of uncertainty reduction in CO_2 plume characterization.Table 4.4 presents the summary of entropy reduction results for both the 2D conceptual and Brugge models. The ’Red.’ column indicates the relative entropy reduction, calculated as the fraction of prior entropy eliminated through data assimilation.

Table 4.4: Summary of Experimental Results for Both Models

Configuration	**Measurement Type**	**2D Synthetic**		**Brugge Field**	
		Points	**Red.**	**Points**	**Red.**
Linear array	Single time	51	0.50	139	0.20
	Multi-time	153	0.67	417	0.63
	Multi-time + BHP*	193	0.71	457	0.68
Low-density grid	Single time	36	0.64	36	0.15
	Multi-time	108	0.79	108	0.55
	Multi-time + BHP*	148	0.81	148	0.61
High-density grid	Single time	121	0.68	121	0.18
	Multi-time	363	0.84	363	0.72
	Multi-time + BHP*	403	0.86	403	0.79
BHP only*	All time steps	40	0.25	40	0.10

* Localization was applied in the 2D Synthetic model for configurations with BHP data.

Prior entropy was 932.68 nats for the 2D model and 260.84 nats for the Brugge model.

Our analysis revealed patterns in how monitoring design affected information gain. While increased spatial density generally improved uncertainty reduction in both models, the relationship wasn’t simply "more is better." Strategic placement and temporal sampling often outweighed raw quantity. Although high-density grid configurations generally achieved greater entropy reductions, we found cases where fewer optimally placed sensors outperformed more numerous but less strategically arranged ones. For instance, the low-density grid (36 points) achieved greater entropy reduction than the linear array (51 points) in the 2D model. The temporal dimension proved particularly impactful, often providing more value than equivalent increases in spatial density. In the 2D conceptual model with a high-density grid, entropy reduction increased when compared the single-time measurements to multiple-time measurements. These findings emphasize that optimizing the number, location, timing, and type of observations can be more effective than simply increasing measurement density.

The integration of different data types provided measurable improvements to monitoring effectiveness. The relative value of integrating additional data types varied depending on the existing monitoring configuration. Based on the results in Table 4.4, the combination of surface displacement and BHP data enhanced entropy reduction across all configurations, with specific improvements ranging from 2% to 7%. For the 2D model, these improvements were 4% for the linear array (0.67 to 0.71), 2% for the low-density grid (0.79 to 0.81), and 2% for the high-density grid (0.84 to 0.86). The Brugge model showed slightly larger gains of 5% for the linear array (0.63 to 0.68), 6% for the low-density grid (0.55 to 0.61), and 7% for the high-density grid (0.72 to 0.79). These improvements demonstrate that different data types provide complementary information for constraining subsurface uncertainty, as BHP measurements capture direct pressure responses at the well that surface measurements alone cannot detect. We observed diminishing returns as the number of data points increased. Figure 4.17 illustrates this relationship, showing that the marginal gain in entropy reduction diminishes significantly after approximately 150-200 data points for both models. For the Brugge model, the slope of improvement decreases from approximately 0.25% per additional 10 data points in the 0-150 range to less than 0.05% per additional 10 data points beyond 300 observations. This can be seen in the 2D conceptual model, where increasing from 363 points (Configuration 1I) to 403 points (Configuration 1J) only improved entropy reduction from 84% to 86%, representing a marginal gain of just 0.05% per 10 additional observations.

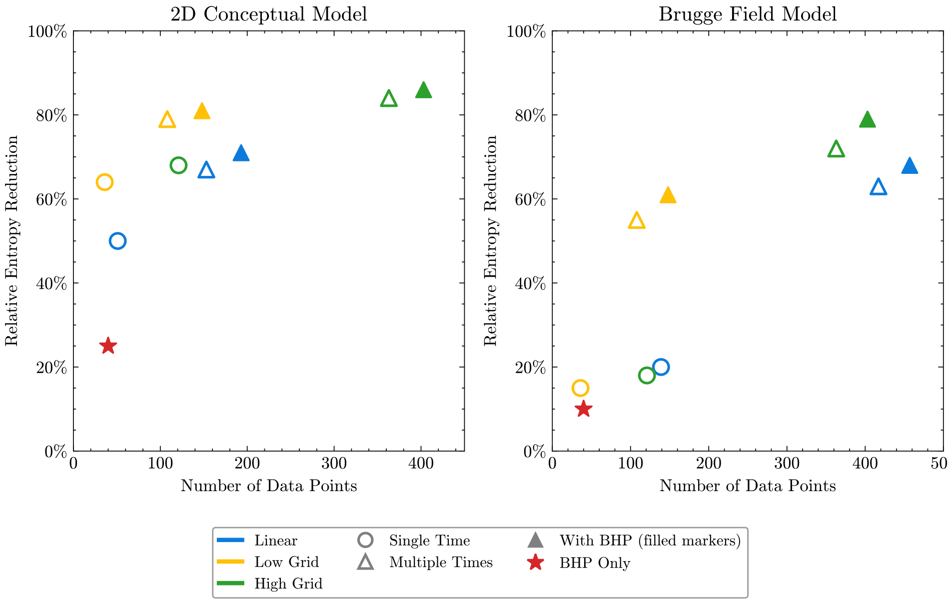


Figure 4.17: Relationship between number of data points and entropy reduction for both models.

Beyond these expected trends, our analysis revealed non-intuitive findings that pro-

Table 4.5: Entropy Reduction Comparison Across Monitoring Aspects

Aspect	**Comparison**	**2D**	**Brugge**
Spatial Configuration	Low-Density Grid vs. Linear Array	+14%	-5%
	High-Density Grid vs. Linear Array	+18%	-2%
Temporal Sampling	Multiple-Time vs. Final-Time (Linear)	+17%	+43%
	Multiple-Time vs. Final-Time (Low)	+15%	+40%
	Multiple-Time vs. Final-Time (High)	+16%	+54%
Data Types	Surface+BHP vs. Surface (Linear)	+4%	+5%
	Surface+BHP vs. Surface (Low)	+2%	+6%
	Surface+BHP vs. Surface (High)	+2%	+7%

vide insights into monitoring system design. In the 2D conceptual model, the low-density grid with only 36 points (Configuration 1E) achieved greater entropy reduction (64%) than the linear array with 51 points (Configuration 1A, 50%). This counter-intuitive result can be explained by the spatial coverage offered by the grid layout. While the linear array provides more measurement points, they are concentrated along a single transect, capturing displacement variations in only one direction. In contrast, the grid configuration, despite having fewer total points, samples the displacement field in multiple directions and provides more comprehensive coverage of the radial displacement pattern characteristic of the 2D model, resulting in more effective plume characterization. We also observed different sensitivity to spatial configuration based on geological complexity. The 2D conceptual model showed greater sensitivity to spatial arrangement, with entropy reduction varying from 50% to 68% across single-time configurations. In contrast, the Brugge model showed more consistent performance (15% to 20%) across these same configurations. This suggests that the geological structure in the Brugge model, with its heterogeneous permeability distribution, variable reservoir depth, and fault structures, may naturally constrain the range of plausible CO_2 distributions, reducing the impact of specific sensor arrangements. Although localization can certainly be applied to the Brugge BHP data, the prior-to-posterior entropy falls by only 10 % (from 260.8 to 234.3 nats), indicating sufficient ensemble spread; we therefore retained the un-localised solution to avoid introducing an additional tuning parameter.

Another interesting result is that when comparing configurations with similar total data points, adding temporal measurements often provided more entropy reduction than equivalent increases in spatial density. For example, in the 2D conceptual model, the linear array with multiple time measurements (153 points, 67% reduction) achieved similar results to the high-density grid with single-time measurements (121 points, 68% reduction) despite having different spatial densities. Similarly, in the Brugge model, the low-density grid with multiple time measurements (108 points, 55% reduction) outperformed both the high-density grid with single-time measurements (121 points, 18% reduction) and the linear array with single-time measurements (139 points, 20% reduction).

While specific insights may vary depending on the modeling approach and context, the framework presented here remains applicable across various fields. These insights provide a quantitative basis for evaluating monitoring strategies in GCS projects using geomechanics and flow data.

4.5. DISCUSSION

Our approach for evaluating monitoring strategies quantifies uncertainty reduction in CO_2 plume characterization using geomechanical data, a data type that remains underutilized in GCS monitoring despite its potential value. The entropy-based framework provides objective comparisons between configurations that traditional metrics cannot readily offer.

The absolute entropy reductions reported here are conditional on the particular reference models chosen. Different geological realities would yield different numerical values, yet the qualitative findings—such as the dominant influence of temporal sampling and the complementary nature of BHP and displacement data—are expected to hold because every monitoring scenario would be re-evaluated against the same new truth. Investigating the sensitivity of information gain to a wider ensemble of truths is a direction for future work once the computational budget permits

Surface displacement monitoring is not yet standard practice for CO_2 storage, where the industry primarily relies on well data, pressure measurements, and seismic surveys (Ringrose, 2020). Some initiatives have begun implementing tiltmeters and InSAR (Davis et al., 2023; Ringrose et al., 2013), reflecting a growing recognition of geomechanical data's value. Our quantitative results support these emerging approaches by demonstrating substantial entropy reductions from displacement measurements. The integration of this data with ensemble-based methods addresses the monitoring design gap identified by Bouqueta et al. (2022) and Seabra et al. (2024a), who noted that configurations are often determined through ad hoc decisions rather than systematic analysis.

Some unexpected findings are present in our analysis that can challenge conventional wisdom about monitoring system design with geomechanics. Although optimizing monitoring setups has been studied often (Kim & Vossepoel, 2024; Le & Reynolds, 2013; Shi et al., 2019), many real projects still just assume using more sensors automatically gives better results. Our framework provides a quantitative method for evaluating monitoring strategies without requiring computationally intensive optimization procedures. The results reveal important nuances that contradict simplistic "more is better" approaches. First, strategic sensor placement proved more important than quantity—in the 2D model, a low-density grid with 36 points achieved greater entropy reduction (64%) than a linear array with 51 points (50%). Second, multiple-time measurements outperformed single-time observations even with fewer spatial points. This balance between when and where to measure hasn't gotten enough attention in previous studies, even though it affects monitoring quality. The observed interaction between spatial and temporal sampling presents practical implications for GCS monitoring design. Our findings suggest that carefully choosing when to take measurements can sometimes make up for having fewer sensors on the ground. This can give operators more flexibility, especially in situations where it's hard to install many instruments—due to limited site access, equipment shortages, or tight budgets. In such cases, spreading out measurements over time might be a practical and cost-effective way to still gather useful information. Site-specific analyses using the framework presented here could help operators determine the most efficient allocation of monitoring resources between spatial coverage and temporal frequency for their particular geological setting, potentially optimizing both cost and information gain.

The different behavior between our two test cases highlights the importance of considering the geological context in monitoring design. The 2D model showed greater sensitivity to spatial configuration (50-68% across single-time configurations) than the Brugge model (15-20%). This suggests that heterogeneous geology may inherently constrain plausible CO_2 distributions, reducing the impact of sensor arrangement in more realistic settings. Our finding that localization techniques were necessary for the 2D model but not for the Brugge model further supports this relationship between geological complexity and data assimilation methods.

Our geomechanical proxy model, based on approaches by Segall (1992) and Geertsma (1973), enabled integration with ensemble methods but has limitations. The model assumes homogeneous, isotropic rock geomechanical properties and linear elasticity, neglecting the heterogeneous mechanical behavior of real formations (van Wees et al., 2019) and time-dependent effects like creep or consolidation (Rutqvist, 2012). A promising strategy for addressing the heterogeneity limitation would be to initially calibrate the proxy model using in-well measurements such as Distributed Strain Sensing (DSS) or formation testing, which can provide direct observations of spatially varying mechanical properties, and then proceed with surface deformation measurements for the data assimilation process. This two-stage approach could enable more accurate representation of heterogeneous mechanical behavior while maintaining computational efficiency. Despite these simplifications, the computational efficiency gained was important for the ensemble-based methodology—a necessary trade-off acknowledged in subsurface monitoring applications (Barbosa et al., 2022; Shi et al., 2019). Limitations of our study also include the examination of only two reservoir models with a single injection scenario, potentially restricting generalizability. Our temporal sampling strategy (years 1, 5, and 10) was selected to capture different stages of the CO_2 injection process rather than for optimality. We also did not incorporate economic factors or operational constraints that would affect real-world implementations.

Like other ensemble smoothers, ESMDA can underestimate posterior variance. We mitigated this with four iterations, an inflation schedule satisfying $\sum 1/\alpha_i = 1$, and localization only where necessary. Although the absolute entropies may therefore be slightly conservative, the bias is systematic across all monitoring scenarios, so the ranking of information gain is preserved. This trade-off—computational efficiency versus exact variance—matches current industry practice. Incorporating variance-inflation or hybrid particle–ensemble methods could be explored in future work to tighten the absolute entropy estimates once larger ensembles become affordable.

It is important to note that we do not implement optimization methods to determine the optimal monitoring design, as this is beyond the scope of this study. Our goal is on evaluating predefined configurations. This distinction is important, we aim to provide a comparative framework for assessing different monitoring strategies.

The framework presented here, combining open-DARTS (Voskov et al., 2024) with our geomechanical proxy and ESMDA, offers a path toward more systematic evaluation of monitoring configurations. This addresses a critical need identified by IEA (2022) and Ringrose et al. (2013) for ensuring long-term CO_2 containment through effective monitoring. As GCS projects expand globally to meet climate mitigation goals, optimization of monitoring resources will become increasingly important. Our quantitative approach enables stakeholders to develop tailored programs based on site-specific requirements, balancing measurement density, temporal frequency, and data types to maximize information gain within practical constraints.

4.6. Conclusion

This study developed a framework for evaluating monitoring strategies in Geological Carbon Sequestration (GCS) that combines Ensemble Smoother with Multiple Data Assimilation (ESMDA) with entropy-based quantification metrics. We integrated surface displacement data—a type data type not commonly utilized in GCS monitoring—with reservoir simulations of CO_2 injection using open-DARTS and a proxy approach based on the nucleus of strain method. This integration allowed us to assess information gain from different monitoring configurations using Shannon entropy reduction to quantify uncertainty in CO_2 plume characterization. The framework was applied to both a 2D conceptual model and the more complex Brugge model, testing various spatial configurations, temporal frequencies, and data types. Our analysis demonstrates that surface displacement measurements can significantly reduce uncertainty in CO_2 plume characterization.

The geomechanical proxy model enabled integration of displacement data with flow simulations, making ensemble-based uncertainty quantification computationally feasible. While simplified compared to full geomechanical simulations, the proxy approach captured the displacements response to CO_2 injection. We found the way sensors are arranged significantly influences how much uncertainty can be reduced. Grid-based configurations (36-121 points) outperformed linear arrays (51-139 points) in the 2D model, demonstrating that strategic sensor placement can be more important than maximizing sensor numbers. Temporal data proved very relevant for characterizing subsurface processes. For the 2D conceptual model, adding multiple time steps increased entropy reduction from 50% to 67% for a linear array, showing that temporal sampling can rival spatial coverage in importance for information gain. Geological complexity influenced monitoring strategy effectiveness, with the Brugge model showing less sensitivity to spatial configuration (15-20% reduction) than the 2D model (50-68%). This suggests that complex geology provides inherent constraints on plausible CO_2 distributions. Combining surface displacement and BHP measurements consistently improved entropy reduction in all configurations, though with more modest gains than those achieved through temporal sampling or spatial optimization. For the high-density grid in the Brugge model, adding BHP data increased reduction from 72% to 79%. Future studies should improve the geomechanical model by adding realistic rock variability and including how deformation changes over time. This approach can also be used to test the value of other data types, like seismic surveys. Studying how measurement uncertainty affects results would also make it more useful in practice. Field validation using data from operational GCS sites with surface monitoring infrastructure would be valuable to bridge the gap between synthetic studies and real-world applications.

While our specific findings are tied to the test cases examined, the quantitative framework itself is broadly applicable across diverse geological settings. This methodology enables practitioners to evaluate site-specific monitoring strategies, aiding the transformation of monitoring systems design for subsurface carbon storage projects worldwide.

CODE AND DATA AVAILABILITY

The developed code is open source and available at the gitlab repository https://gitlab.com/open-darts/subsidence-esmda together with input files of the numerical models. The open-DARTS simulator is available at https://gitlab.com/open-darts/open-darts. The dageo Python package for data assimilation in geosciences is available at https://github.com/tuda-geo/dageo.

5

AI ENHANCED DATA ASSIMILATION AND UNCERTAINTY QUANTIFICATION APPLIED TO GEOLOGICAL CARBON STORAGE

This chapter advances data assimilation for geological carbon storage by integrating machine learning surrogates with traditional methods while preserving physical consistency. The study evaluates Fourier Neural Operators (FNO) and Transformer U-Net (T-UNet) architectures as surrogate models for CO_2 *flow in channelized reservoirs. Two hybrid frameworks are introduced: Surrogate-based Hybrid ESMDA (SH-ESMDA), which accelerates ensemble-based data assimilation by over 50%, and Surrogate-based Hybrid RML (SH-RML), which enables gradient-based optimization without adjoint models. Both methods use ML surrogates for computational efficiency while employing high-fidelity simulators for final predictions, ensuring physical consistency. The SH-RML approach demonstrates superior uncertainty quantification compared to conventional ESMDA, particularly for non-Gaussian channelized permeability distributions.*

Parts of this chapter have been published in Gabriel Serrão Seabra, Nikolaj T. Mücke, Vinicius Luiz Santos Silva, Denis Voskov, Femke C. Vossepoel, AI enhanced data assimilation and uncertainty quantification applied to Geological Carbon Storage, *International Journal of Greenhouse Gas Control* **136** (2024) 104190 (Seabra et al., 2024b). Changes have been applied to make the text and figures consistent with the thesis.

5.1. INTRODUCTION

Effective pressure management in Geological Carbon Storage (GCS) projects is crucial to optimizing storage capacity and mitigating risks such as induced seismicity and caprock failure. Li and Liu (2016) identified that most GCS project risks originate from pressure management issues at storage sites, which can induce seismicity and lead to caprock failure at weak points when CO_2 injection pressures exceed the mechanical stability thresholds of the geological formations. These scenarios can result in microseismic events or mechanical failures that compromise the caprock's integrity and the overall containment security (Rutqvist, 2012; Zoback & Gorelick, 2012). The CO_2 injection process can alter subsurface stress states, potentially triggering seismic events ranging from microseismicity to significant earthquakes (magnitude 5+ events) (White & Foxall, 2016). Researchers have proposed several solutions to improve pressure management at storage sites. For instance, Machado et al. (2023a) suggests the application of horizontal wells to achieve a better distribution of the CO_2 plume and a smaller increase in reservoir pressure.

Robust UQ and forecasts in GCS projects typically rely on several components. These encompass robust geological models representing reservoir complexities accurately, high-fidelity reservoir simulators capturing intricate CO_2 injection dynamics and data assimilation (DA) techniques to combine these elements with field observations. It is important to note that effective uncertainty quantification can still be achieved even without data assimilation, depending on the specific requirements of the project and the available data.

DA techniques contain ensemble methods (e.g., Ensemble Kalman Filters and Ensemble Smoothers), variational methods (e.g., Randomized Maximum Likelihood and 4D-Var), and fully nonlinear DA methods (e.g., Particle Filters and Markov Chain Monte Carlo) (Evensen et al., 2022; Tarantola, 2005). These methods leverage data, prior knowledge, and physics-based models to predict reservoir behavior under uncertain conditions. While ensemble methods are computationally efficient and flexible, variational approaches can offer better convergence but require gradient computations. Fully nonlinear methods can offer high accuracy for systems characterized by nonlinearities, such as the CO_2 injection. However, the computational intensity of these methods can be prohibitive due to the substantial resources and time they require. In the context of CO_2 injection and DA, Tarrahi et al. (2015) integrated microseismic monitoring data of CO_2 injection with coupled flow and geomechanical models using the Ensemble Kalman Filter (EnKF), enabling the conditioning of heterogeneous rock permeability and geomechanical property distributions on microseismic data. Similarly, Li et al. (2017) employed one-step ahead smoothing for joint state-parameter estimation, important for addressing nonlinearities in CO_2 storage aquifers. Tadjer and Bratvold (2021) introduced a Bayesian Evidential Learning (BEL) for managing uncertainties in geological CO_2 storage, integrating Monte Carlo simulations with ensemble smoother direct forecasting to enhance decision support systems. Utilizing the Utsira saline aquifer in Norway as an example, the BEL approach improves the predictability of CO_2 sequestration by effectively managing uncertainties and potential leakages. These studies emphasize the significance of assimilating diverse data types into reservoir models, refining the understanding of subsurface properties, and optimizing CO_2 injection strategies, thereby contributing to the realization of effective

GCS projects in mitigating climate change.

Specialized softwares are often employed to model the heterogeneities in complex reservoirs, such as channelized formations. Among these, Alluvsim is an open-source option that generates multiple geological models with features like channel size, curvature, and shifts, using streamlines as building blocks to mimic natural deposition processes (Pyrcz et al., 2009). These detailed models reflect the heterogeneities commonly encountered in GCS projects. For high-fidelity reservoir simulation, numerical simulators designed to handle multiphase, multicomponent flow and transport in porous media are utilized. Examples of such simulators include CMG GEM (Computer Modelling Group Ltd., 2023), SLB Eclipse (Schlumberger, 2023), and open-source options like DuMux (DuMux Development Team, 2023) and GEOSX (GEOSX Development Team, 2023). Delft Advanced Research Terra Simulator (DARTS), recently released as an open-source reservoir simulator for energy transition applications, efficiently simulates CO_2 injection using advanced numerical techniques like the operator-based linearization approach (Khait & Voskov, 2017; Lyu & Voskov, 2023). However, integrating these simulators into DA frameworks can be challenging due to high computational costs.

The choice of Alluvsim and DARTS as our primary tools stems from their proven efficacy in handling complex geological formations and fluid dynamics simulations, respectively. Alluvsim's capability to accurately simulate channelized reservoirs (Delottier et al., 2023), coupled with DARTS's optimized computational efficiency (Khait & Voskov, 2017), makes them particularly suitable for our study's objectives.

The preceding chapters demonstrated a comprehensive data assimilation workflow using physics-based proxy models for geomechanics (Chapters 3 and 4). This chapter represents a progression in our research by applying the proxy-based approach through the employment of machine learning (ML) surrogates. This transition to ML-based methods allows us to address challenges of high computational cost and robust data assimilation. Where the physics-based proxies in previous chapters were semi-analytical formulations and homogeneity assumptions, ML surrogates can learn complex, nonlinear relationships directly from data without requiring explicit physical simplifications.

Recently, researchers have actively been exploring innovative strategies to merge machine learning (ML) and DA (Buizza et al., 2022; Silva et al., 2024). Buizza et al. (2022) provides a high-level overview of techniques for integrating DA and ML, called “Data Learning" for improving DA in several fields. Their focus is on approaches that leverage the strengths of ML's ability to uncover complex patterns in data and data assimilation's incorporation of physical models and dynamical constraints. Similarly, Cheng et al. (2023) explores how mixing ML and DA can make research on DA robust. The study sorts these methods into two main groups. The first group, called “DA using ML" looks at how ML can help solve problems in data assimilation. This includes fixing errors in DA models by adding ML, using it to estimate unknown variables in DA, and defining error rates using ML methods. The study also talks about how neural networks can help in learning DA systems from start to finish. The second group, “ML improved by DA/UQ" focuses on how DA and UQ can improve ML models. This covers topics like using Bayesian neural networks for uncertainty analysis in ML, fixing errors in simplified ML models with real-time data, and using DA to identify equations from noisy or incomplete data. Brajard et al. (2021) recently proposed an innovative approach that combines DA and ML to infer unresolved scale parametrization in models, helping overcome limitations from sparse and noisy observational data.

In the domain of deep learning for efficient surrogate modeling, UNets have long demonstrated their efficacy, particularly in tackling subsurface problems (Pintea et al., 2022; Wen et al., 2021; Zhang et al., 2021). Originating from biomedical image segmentation, UNets excel at capturing local features through specialized convolutional layers. The architecture comprises an encoder and a decoder connected by a “highway" system of channel concatenation, facilitating the transfer of multiscale spatial information. This has enabled outstanding predictive accuracy in diverse applications (Ronneberger et al., 2015; Taccari et al., 2022). More recently, advancements have been made by integrating UNets with transformers. Li et al. (2023) explored this method for robust medical images segmentation and Alsalmi and Elsheikh (2024) applied an attention UNet for seismic segmentation (see also (Li et al., 2023) for a related transformer–UNet design). As a related example in hydrology, attention U-Net surrogates enable image-to-image groundwater flow prediction, focus learning on salient regions, and significantly reduce runtime compared to state-of-the-art numerical solvers (Taccari et al., 2022). On the other hand, Fourier Neural Operators (FNOs) have recently emerged as a promising method to build surrogate models for reservoirs submitted to CO_2 injection. Employing Fourier basis functions, FNOs efficiently capture multiscale interactions and offer a novel way to overcome traditional limitations in surrogate modeling (Li et al., 2021; Wen et al., 2022; Witte et al., 2023). The original FNO learns mappings between function spaces, achieves zero-shot super-resolution on turbulent flows, and runs up to three orders of magnitude faster than traditional PDE solvers while maintaining high accuracy (Li et al., 2021). Building on this line, U-FNO extends FNOs to multiphase flow, providing faster and more accurate predictions of gas saturation and pressure buildup with improved data efficiency relative to CNN baselines (Wen et al., 2022).

Recently, Abbate et al. (2022), Crain et al. (2024), Tang et al. (2021), and Tang et al. (2022) developed surrogate models for CCUS DA, aiming to replace physics-based numerical models. However, these models often require a substantial amount of training data from high-fidelity simulations, posing practical challenges for real-world CCUS projects with limited computational resources. While deep learning has improved surrogate model accuracy, Dong et al. (2021) shows that it may struggle to capture subsurface complexities fully.

Hybrid models that combine physics-based and ML approaches have been explored by De Brito and Durlofsky (2021), Korondi et al. (2021), and Tang and Durlofsky (2022) to mitigate the specific limitations inherent to both physics-based and ML models, with the goal of forging a more balanced and comprehensive tool. ML models, while proficient at identifying patterns and correlations within large datasets, may lack the capability to infer the underlying physical processes governing the system. This limitation can lead to potential inaccuracies in predictions under unseen conditions or parameter ranges, especially when dealing with the intricate geological variations and non-linear fluid dynamics inherent in subsurface environments. On the other hand, physics-based models offer reliable insights into these underlying processes but may struggle with computational efficiency. However, significant challenges arise when incorporating ML surrogate models in DA due to the different parameterizations employed by ML surrogates and physics-based, high-fidelity simulators. This misalignment can impede the integration, limiting the applicability and efficacy of the resulting UQ in real-world GCS projects.

In our methodology, we initially employ a standard ESMDA approach utilizing DARTS for high-fidelity simulations of channelized reservoirs built with Alluvsim. This scenario poses a significant challenge for DA due to its highly nonlinear nature and the non-Gaussian distribution of parameters. To improve upon the standard ESMDA methodology, we evaluate two ML surrogate models for comparison: one rooted in the Fourier Neural Operators and the other adopting a Transformer UNet (T-UNet) architecture, which to our knowledge is the first application of these techniques to GCS subsurface problems. Our observations reveal that FNOs show a slight advantage over T-UNet, particularly for small datasets. Subsequently, we develop two hybrid techniques to integrate DA with these ML surrogates. The first, termed Surrogate-based hybrid ESMDA (SH-ESMDA), incorporates the ML surrogates models, expediting the ESMDA process by around 50% or more, and thereby facilitating quicker uncertainty evaluations. For the second technique, known as Surrogate-based Hybrid RML (SH-RML), we use the ML surrogates models specifically for gradient calculations within a variational framework and compute the posterior curves with high-fidelity physics simulator DARTS. The SH-RML achieves better history matching than ESMDA and SH-ESMDA.

In summary, our contributions are as follows:

- We train and test two different types of novel ML surrogates in a channelized reservoir setting for CO_2 storage, a FNO and a T-UNet.
- We introduce two novel hybrid methods, SH-ESMDA and SH-RML, that incorporate ML into both ensemble and variational DA techniques. The first, significantly expedites the DA process and the second allows one to perform variational DA.
- Both proposed methods ensure that posterior high-fidelity physics solution is respected.
- The proposed methods are versatile and can be adapted to various physical systems beyond CO_2 sequestration.

The paper is organized as follows: Section 5.2 discusses the creation of geological models, followed by CO_2 injection simulations with DARTS. In Section 5.3, we delve into the ML models used to build the surrogate models. Section 5.4 and Section 5.5 describes the DA methods discussed in the paper, ESMDA, RML and the hybrid methods SH-ESMDA and SH-RML. Finally, Section 5.6 presents the results and implications of these methods for enhancing DA in CO_2 storage projects.

5.2. OVERVIEW OF THE RESERVOIR MODEL FOR CO_2 INJECTION

5.2.1. GEOLOGICAL MODELING USING ALLUVSIM

We use Alluvsim, a specialized algorithm to simulate channelized reservoirs (Delottier et al., 2023; Pyrcz et al., 2009). This tool allows us to manipulate different geological variables that impact channel features. Following the guidelines by Pyrcz et al. (2009), we create multiple geological models by altering essential parameters within certain limits.

We consider as variables the likelihood of channel shifts, known as avulsion probability, and vertical sediment build-up, or aggradation levels, within statistically defined ranges. Parameters such as channel orientation, thickness, and geometric aspects like the width-to-thickness ratio are also modeled using various distributions to mimic natural variability. We similarly vary levee width to represent lateral sediment deposition and channel sinuosity to capture meandering behavior. Properties are distributed across different facies, to accurately represent variations in rock quality. This allows a comprehensive evaluation of the reservoir's attributes while accounting for uncertainty.

By randomly selecting values for the aforementioned parameters, we produce multiple realizations that capture the variability in the properties of channelized reservoirs. Figure 5.1 showcases the permeability distributions of six randomly sampled models from this dataset. Each model in the dataset has a grid dimension of 32 × 32 × 1 with a spatial discretization of 192 × 192 × 10 m. This grid resolution was carefully chosen, considering computational efficiency and memory requirements.

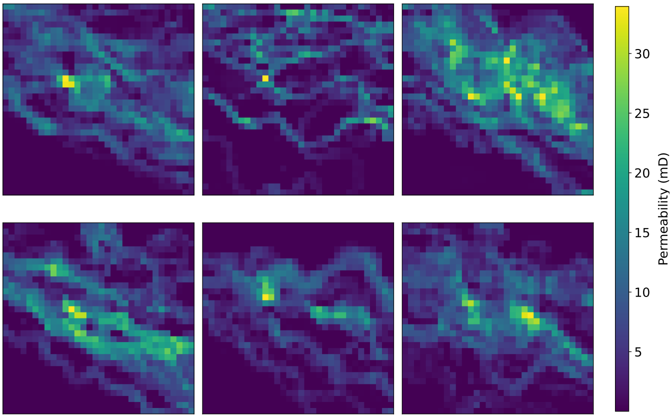


Figure 5.1: Permeability maps of six models from the dataset.

The histograms in Figure 5.2 clarify the permeability distribution of a model sample from our dataset. Note that the distributions are not Gaussian. These diverse geological models define the permeability distribution for subsequent CO_2 simulations which will, in turn, be used for the training of the ML methods.

It is worth mentioning that a 2D horizontal model was employed to simplify the study, focusing on the areal permeability distribution within channelized reservoirs. While acknowledging the influence of gravity in three-dimensional settings, this approach enables the examination of non-Gaussian and highly heterogeneous permeability patterns typical of natural depositional environments. The decision to use a 2D model was made to highlight how lateral variations in permeability impact CO_2 flow dynamics.

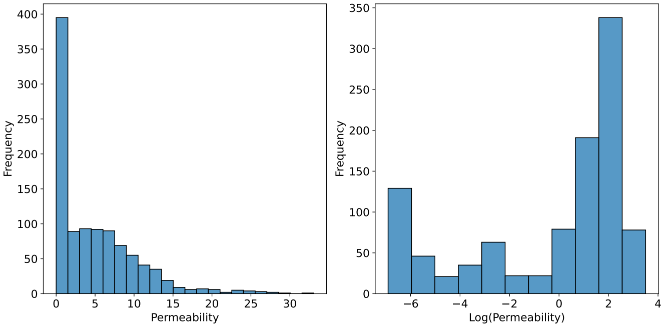


Figure 5.2: Histogram of permeability (left) and log of permeability (right) for one sample.

5.2.2. CO_2 INJECTION SIMULATION USING DARTS

Delft Advanced Research Terra Simulator (DARTS) is used to perform reservoir simulations of CO_2 injection into channelized reservoirs. DARTS is engineered for handling complex flow through porous media and is optimized for computational speed via techniques such as operator-based linearization (OBL) yielding a high-fidelity, physics-based representation (Chen & Voskov, 2020; Khait & Voskov, 2017; Pour et al., 2023). Detailed descriptions of the physics, analytical validations, and example models of DARTS for GCS modeling are documented in Wapperom et al. (2024), Lyu and Voskov (2023) and openly available in the software open repository (open-DARTS Development Team, 2024).

For the CO_2 flow simulation in this study, the gas phase viscosity is assumed to be solely that of CO_2, calculated using the model from Fenghour et al. (1998). The densities of the CO_2 and CH_4 mixture are computed using the Peng-Robinson equation of state, with properties determined by their composition. In terms of dissolution, all components are modeled simultaneously using a hybrid EOS with fugacity-activity formulation (Wapperom et al., 2024). A power-law model (Corey) is applied to describe relative permeability in the porous media, with exponent values $n_w = 2.0$ and $n_g = 1.5$, and terminal saturation points $S_{wc} = 0.25$ and $S_{gc} = 0.1$. Although aspects like diffusion and hysteresis can influence CO_2 migration and trapping, they are neglected to align the simulation objectives with broader reservoir behavior. Capillary pressure effects are also neglected, as they are considered to have minimal impact on the 2D model utilized. We assume this model captures the multiphase flow characteristics pertinent to CO_2 migration and trapping in porous media relevant this the present study.

In DARTS, the geological model is discredited as a computational grid. Within this grid, one well is placed in the center of the reservoir for CO_2 injection. The model simulates the injected CO_2 plume and conducts equilibrium flash calculations to determine phase partitioning.

Our simulation includes CO_2, CH_4, and H_2O as components, with initial conditions featuring a uniform gas saturation of 20% and a composition of 2% CO_2 and 98% CH_4. CO_2 is injected via one well in the center of the reservoir at a time-varying prescribed gas rate. In our simulations, the time frame comprises 61 time steps, each lasting 30 days. Although this is a shorter window than what is typically encountered in real-world CCUS projects, this duration is sufficient to induce overpressure in the reservoir, which is an aspect we aim to investigate as it can impact the further CO_2 distribution in the reservoir. Figure 5.3 illustrates our simulation results. The top row shows pressure distributions, and the bottom row displays CO_2 molar fractions, captured at the initial, intermediate, and final time steps.

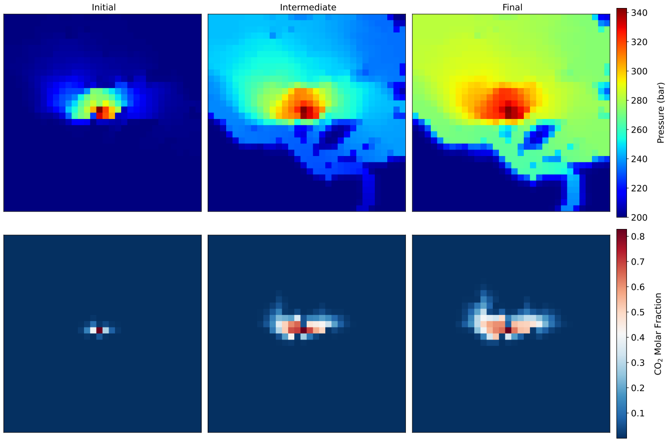


Figure 5.3: Simulation results using DARTS. Top row: Pressure distributions; Bottom row: CO_2 molar fractions.

A key observation is the differential progression of the pressure and CO_2 fronts. Pressure changes, governed by diffusion phenomena propagate more rapidly through the reservoir. In contrast, the CO_2 front advances more slowly, influenced by complex transport mechanisms. This differential movement highlights the potential value of monitoring pressure as an early warning system for subsurface changes, even before significant CO_2 migration occurs. Because of this, we choose pressure as a key variable for monitoring in our subsequent DA studies.

By applying DARTS to the Alluvsim realizations, we generate a comprehensive suite of high-fidelity simulations that form the basis for our subsequent ML model training and DA study. It's important to note that the combination of geological complexities modeled by Alluvsim and the fluid dynamics simulated by DARTS creates a highly nonlinear problem. Coupled with the non-Gaussian distribution parameters, as evident from Figure 5.2, this poses a substantial challenge for traditional DA methods. This complexity further emphasizes the need for advanced approaches, including ML-based techniques, to accurately perform DA.

5.3. NEURAL NETWORK AS SURROGATE FORWARD MODELS

A surrogate model is a model that replaces the high-fidelity model for the simulation of fluid behavior in a reservoir. The surrogate model should be computationally cheap to evaluate without sacrificing accuracy. To achieve this, we construct a surrogate model based on the output of the high-fidelity model. This is done in two stages: an *offline stage*, in which the surrogate model is trained, and an *online stage*, where the surrogate model is used in place of the high-fidelity model. Conventionally, surrogate models achieve speed-ups by lowering the dimension of the state and corresponding equations to be solved, such as in proper orthogonal decomposition (Hesthaven et al., 2016; Quarteroni et al., 2016). However, due to large Kolmogorov N-widths for highly nonlinear and/or hyperbolic problems (Ohlberger & Rave, 2015), such linear approaches have lately been replaced with equation-free methods. Here, one replaces the equations with a function that directly gives the output of interest. For this to work, one must typically use highly complicated functions to make up for the lack of physical knowledge in the online stage. The offline stage is typically significantly more expensive and requires more training data. Neural networks are immensely popular in this approach due to their capability of approximating highly nonlinear functions (Geneva & Zabaras, 2022; Li et al., 2021; Mücke et al., 2021).

The general setup for the offline stage is to first generate high-fidelity solutions and then train the neural networks on these solutions. In this paper, we specifically make use of the neural network architectures T-UNet and FNOs.

5.3.1. FORWARD MODEL

The forward model, that is, the simulator used to describe the behavior of fluid in a reservoir, maps input parameters to a state trajectory. In our case, it maps permeability and porosity to the state trajectory of pressure and fluid flow. Injection rate is the control that determines the behavior of the fluid. Let $\boldsymbol{k} \in \mathbb{R}^{N_x \times N_y}$ be the permeability, $\boldsymbol{\phi} \in \mathbb{R}^{N_x \times N_y}$ the porosity, $\boldsymbol{q} \in \mathbb{R}^{N_t}$ the injection rate, and $\boldsymbol{z} = (k, \phi, q)$. We then define the space of parameters $\boldsymbol{z} \in Z$. The state trajectories consist of pressure, $\boldsymbol{p} \in \mathbb{R}^{N_x \times N_y \times N_t}$, and CO_2 molar fraction, $\boldsymbol{f} \in \mathbb{R}^{N_x \times N_y \times N_t}$. We define the state trajectory space, $\boldsymbol{d} \in V$ and find that $\boldsymbol{d} = (\boldsymbol{p}, \boldsymbol{f})$. With this, the forward map, G, is defined by:

$$G : Z \to V, \quad \boldsymbol{z} \mapsto \boldsymbol{d}. \tag{5.1}$$

The surrogate forward model, $\hat{G}$, approximates G:

$$\hat{G} : Z \to V, \quad \boldsymbol{z} \mapsto \boldsymbol{d}, \quad \hat{G}(\boldsymbol{z}) \approx G(\boldsymbol{z}). \tag{5.2}$$

While G maps $\boldsymbol{z}$ to $\boldsymbol{d}$ implicitly by solving a set of PDEs, $\hat{G}$ directly maps $\boldsymbol{z}$ to $\boldsymbol{d}$. $\hat{G}$ is parametrized by a family of neural networks. As $\hat{G}$ is a neural network, it consists of a set of weights, $\boldsymbol{\omega}$. The weights are fitted in the offline stage.

The training of $\hat{G}$ is performed in the offline stage by first generating N_s training samples by using the high-fidelity forward model:

$$Z_{\text{train}} = \{\boldsymbol{z}_i\}_{i=1}^{N_s}, \quad V_{\text{train}} = \{G(\boldsymbol{z}_i)\}_{i=1}^{N_s} = \{\boldsymbol{d}_i\}_{i=1}^{N_s}, \quad S_{\text{train}} = (Z_{\text{train}}, V_{\text{train}}). \tag{5.3}$$

$\hat{G}$ is then trained by minimizing a loss function with respect to the weights of $\hat{G}$, ω:

$$L(\hat{G}, S_{\text{train}}) = \frac{1}{N_s} \sum_{i=1}^{N_s} l(\hat{G}(\boldsymbol{z}_i), \boldsymbol{d}_i) + \lambda ||\boldsymbol{\omega}||_2^2, \quad (\boldsymbol{z}_i, \boldsymbol{d}_i) \in S_{\text{train}}, \tag{5.4}$$

where l is some loss, typically the l^p or L^p norm of the residual, λ is a hyperparameter, and $|| \cdot ||_2^2$ is the squared l^2 norm and serves as a regularization term. The minimization of L is performed by stochastic gradient descent (SGD) with respect to $\boldsymbol{\omega}$. The specific SGD algorithm is often chosen to be the Adam optimizer (Kingma & Ba, 2014) or other variations thereof.

Below, we will present the particular neural network architectures we use in this paper.

5.3.2. TRANSFORMER UNET

The UNet architecture was introduced in Ronneberger et al. (2015). The idea is to reduce the dimension of the input data down to a bottleneck via a series of convolutional layers and then increase the dimension back to the original shape via upscaling convolutional layers. The bottleneck layers serves as a low-dimensional representation of the data that is rich in feature information. In the upscaling part of the network, the intermediate layers from the downscaling part are concatenated to the convolutions to provide context.

Originally developed and primarily utilized in the medical imaging field, the T-UNet architecture is being adapted in this work for GCS projects. To the best of our knowledge, this constitutes the first endeavor to apply the T-UNet architecture in the realm of GCS.

The choice of the T-UNet architecture, a variant of the well-established UNet model, is grounded in its substantial application and success in various fields, particularly in subsurface applications (Alsalmi & Elsheikh, 2024; Ronneberger et al., 2015; Zhang et al., 2021).

The UNet type of architecture allows one to add information to the predictions on various spatial levels. Specifically, we utilize this to inform the forward model with the injection rate in the bottleneck layers. By adding this information in the bottleneck layers, we effectively affect the rich feature encodings with additional information in an efficient manner. This makes the computations cheaper and makes it easier for the network to learn the relations between the input parameters and the output.

In the proposed architecture, we input the spatially distributed parameters, porosity and permeability, as a two-channel “image". Then we copy that N_t times and concatenate a channel consisting only of the time. This way we can compute the bottleneck encoding of the space for each time step as a batch consisting of 3D tensors (channels, height, width) rather than a single 4D tensor (channels, time steps, height, width), which enables us to use 2D convolutional layers instead of 3D convolutional layers. Since 3D convolutional layers are significantly more memory and compute-heavy, this gives us significantly more efficiency.

For the conditioning of the bottleneck layer, we use cross-attention in the shape of the transformer architecture (Vaswani et al., 2017). The transformer has been shown to provide state-of-the-art performance on multiple types of data and seems to be the superior choice for multi-modal data (Rombach et al., 2022; Xu et al., 2023). Unfortunately, the attention mechanism is very compute and memory intensive and scales poorly with the data dimension. By utilizing the transformer in the bottleneck layers, however, this problem is circumvented. For a description of the transformer neural network, see (Vaswani et al., 2017).

As mentioned, we utilize transformers to condition the forward model on the injection rate. We do this by first embedding the injection rate time series through dense layers, such that the dimensions match the encoded spatial dimensions. Then, we employ positional encoding which receives information from a dense embedding that originates from the gas injection rates. The encoded positions are then passed to the transformer decoder layers, effectively providing a richer context for each time step. This ensures that the transformer is not only aware of the feature information but also the sequence in which they occur. The embedding time series is passed through transformer encoder layers after which it is combined with the spatial data through transformer decoders. For a visualization of the full T-UNet, see Figure 5.4

5.3.3. FOURIER NEURAL OPERATORS

FNOs were introduced in (Li et al., 2021) for various parametric PDE problems. In contrast to conventional neural networks, FNOs learn operators between function spaces instead of Euclidean spaces. This makes FNOs resolution invariant. The general idea is to make use of the Fourier transform, followed by a series of operations in Fourier space, after which the data is transformed back to physical space. Hence, a single Fourier layer is given by:

$$a^{n+1}(x) = \sigma\left(Wa^n(x) + \mathscr{F}^{-1}(R\cdot(\mathscr{F}a^n))(x)\right), \tag{5.5}$$

where $a^n(x)$ is the output of the nth layer, W and R are affine transformations consisting of trainable weights, $\mathscr{F}$ is the Fourier transform, and σ is an activation function. Before applying R the number of modes is truncated to a pre-defined number of modes, k. The Fourier transform is in practice approximated by a discrete Fourier transform. W is typically a standard convolutional layer with a kernel of size one. While truncating the number of Fourier modes removes high-frequency information, the convolutional layer, W, compensates for that. For visualization of the FNO layer, see Figure 5.5

The FNO layers are preceded by a projection layer, that maps the number of input channels to the desired number of hidden channels. Similarly, the FNO layers are superseded by another projection layer that maps the number of hidden channels to the number of output channels.

For our specific application, we use the FNO to map parameters, $\boldsymbol{z}$, to corresponding state trajectories, $\boldsymbol{d}$. To capture the 3D structure of the data, we use 3D Fourier transform and 3D convolutions. The injection rate is encoded to have a 3D structure. The rate at each time step is copied onto all discrete spatial points:

$$\boldsymbol{q}_{\text{enc}} = [\boldsymbol{q}_0\mathbf{1}_{N_x\times N_y},\dots,\boldsymbol{q}_{N_t}\mathbf{1}_{N_x\times N_y}], \tag{5.6}$$

where $\mathbf{1}_{N_x\times N_y}\in\mathbb{R}^{N_x\times N_y}$ is a matrix consisting of ones. The subscript $()_{\text{enc}}$ signifies that the quantities are encoded to fit the 3D tensor format. Similarly, the spatial points coordinates, (x,y) are encoded and copied along the temporal dimension. The time steps are treated in the same way as the injection rate. Lastly, the porosity and permeability are also copied along the temporal dimension. Hence, the input to the FNO is:

$$(k,\phi,q,x,y,t)_{\text{enc}} = (k_{\text{enc}},\phi_{\text{enc}},q_{\text{enc}},x_{\text{enc}},y_{\text{enc}},t_{\text{enc}})\in\mathbb{R}^{N_c\times N_x\times N_y\times N_t}, \tag{5.7}$$

where N_c is the number of channels. In our case, $N_c=6$ – permeability, porosity, injection rate, x, y, and time step.

For the training of the FNO, we use the squared L^2-norm. This is an unusual choice for neural networks but a very common metric for PDEs. As neural networks typically

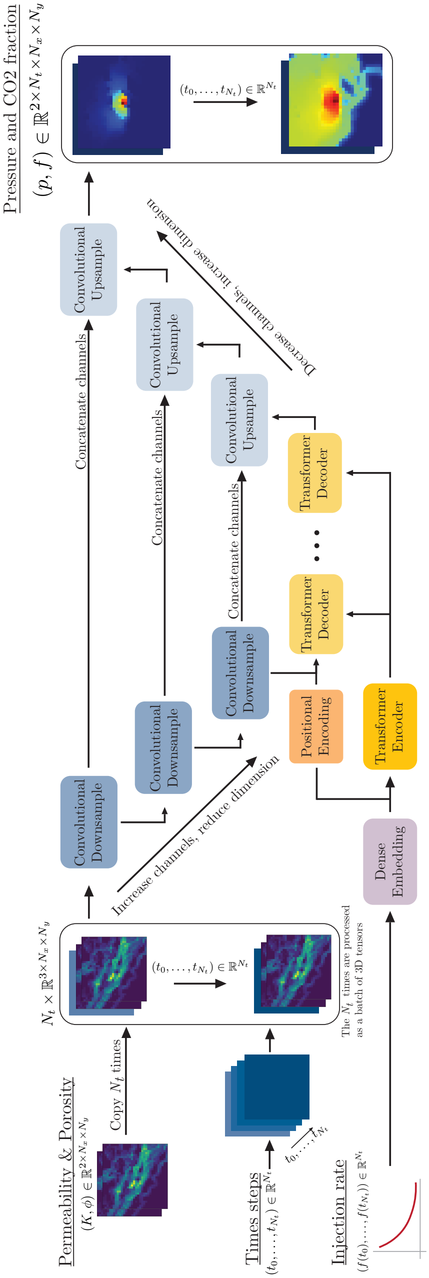


Figure 5.4: T-UNet architecture.

map tensors to tensors, the l^2-norm is the most frequent choice. However, since FNOs map functions to functions, we can make use of the squared L^2-norm, which is also a much more appropriate choice when dealing with PDEs. It's worth noting the nuanced difference between the L^2-norm and the l^2-norm, particularly when it comes to implementation. In both cases, you'll need to discretize the integral for computational purposes. However, the distinction lies in the underlying space over which the norms are computed. When using the L^2-norm, one is essentially approximating the integral over a function space, aiming to capture the "true" behavior of the function. On the other hand, the l^2-norm is computed over an Euclidean space, essentially summing up the squared differences in a point-wise manner. Hence, the loss function for the training is:

$$L(\hat{G}, S_{\text{train}}) = \frac{1}{N_s} \sum_{i=1}^{N_s} ||\hat{G}(\boldsymbol{z}_i) - \boldsymbol{d}_i||_{L^2}^2 + \lambda ||\boldsymbol{\omega}||_2^2, \quad (\boldsymbol{z}_i, \boldsymbol{d}_i) \in S_{\text{train}}. \tag{5.8}$$

Both FNO and T-UNet show promise in approximating the high-fidelity forward model, G. T-UNet capitalizes on the strength of 2D convolutional layers and transformer architectures to efficiently incorporate temporal information and spatial parameters. Its design allows for an efficient encoding of multi-dimensional data, making it computationally more lightweight. On the other hand, FNOs offer resolution invariance and the capability to operate directly in function spaces, making them highly suitable for parametric PDE problems. However, one caveat with FNOs is their higher memory consumption. This is largely due to their need to include time as an additional channel and their use of 3D Fourier transforms and convolutions, which significantly increase the size of input tensors.

The code for both the T-UNet and FNOs architectures is publicly accessible. The repository can be found at https://github.com/nmucke/subsurface-DA-with-generative-models.

5.4. DATA ASSIMILATION

5.4.1. CONVENTIONAL DATA ASSIMILATION METHODS

Ensemble-based DA techniques are computationally efficient, as they lend themselves well to parallelization. These methods offer a level of flexibility by requiring minimal alterations to the existing forward model code and avoiding the computation of adjoints gradients (Evensen et al., 2022). Among various ensemble-based DA techniques, the Ensemble Smoother (ES) serves as an effective method but it has limited ability to provide adequate data matches in complex problems, such as reservoir simulations, due to the application of a single Gauss–Newton correction for conditioning the ensemble to all available data (Emerick & Reynolds, 2013a). To address this, Emerick and Reynolds (2013b) introduced the ESMDA, an iterative version of ES, allowing for improved data matches by assimilating the same data multiple times with an inflated covariance matrix of measurement errors, which enables a more robust approach to updating the model. Its sampling behavior and performance characteristics in simplified settings have been investigated by Emerick and Reynolds (2013a). In spite of its inherent assumptions of Gaussianity, ESMDA is also applicable to weakly non-Gaussian problems. The method is easy to implement, leading to broad application in various scenarios.

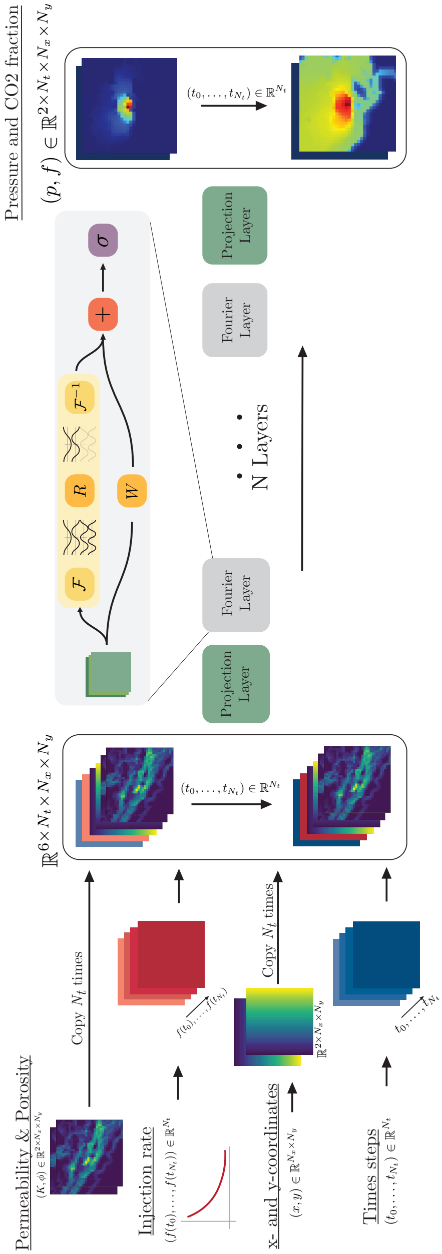


Figure 5.5: FNO architecture.

We can describe the analysis for each ensemble member's model parameter $\boldsymbol{z}_j^a$ as follows (Emerick & Reynolds, 2013a),:

$$\boldsymbol{z}_j^a = \boldsymbol{z}_j^f + C_{zd}^f \left(C_{dd}^f + \alpha_i C_{dd} \right)^{-1} \left(\boldsymbol{d}_j - G(\boldsymbol{z}_j)^f \right), \quad \text{for } j = 1, 2, \ldots, N_e. \tag{5.9}$$

Here N_e is the total number of ensemble members, $\boldsymbol{z}_j^a$ and $\boldsymbol{z}_j^f$ represent the analyzed and forecasted parameters of the j^{th} ensemble member, respectively, C_{zd}^f and C_{dd}^f are the cross- and auto-covariance matrices of the model parameters and data in the forecast step, α is a scaling factor, and $\boldsymbol{d}_j$ and $G(\boldsymbol{z}_j)^f$ are the perturbed and forecasted observations for the j^{th} ensemble member. Besides, $G()$ represents the forward model. For more details, see Equation 5.10:

$$\boldsymbol{d}_j = \boldsymbol{d}_{\text{obs}} + \sqrt{\alpha} C_{dd}^{1/2} \eta, \tag{5.10}$$

In this equation, $\boldsymbol{d}_{\text{obs}}$ is the observed data, C_{dd} is the measurement error covariance, and η is sampled from a standard normal distribution with zero mean and an identity matrix as the covariance. The factor $\sqrt{\alpha}$ scales the perturbations.

In this study, the model parameter $\boldsymbol{z}$ is defined as the permeability of the medium, and $\boldsymbol{d}_{\text{obs}}$ represents pressure values obtained from specific monitoring points.

Randomized Maximum Likelihood is a variational DA method for approximating the posterior pdf with a method introduced in S. Oliver et al. (1996). As a gradient-based method, it provides the advantage of better convergence and accuracy than an ensemble-based method, within a specified solution space. These benefits come at the cost of computational complexities. The need for adjoint models can be prohibitive and the computational efforts linked to the linearization of the model may introduce errors (García-Pintado & Paul, 2018).

RML employs a set of cost functions, often denoted as $J(\boldsymbol{z}_j)$. These cost functions aim to minimize the discrepancy between the ensemble's forecasted model states and the observed data, as well as the a priori model information. By optimizing these cost functions, RML generates multiple models that are consistent with the available observations, thus aiding in robust UQ. The core idea of RML is to use a set of cost functions, $J(\boldsymbol{z}_j)$, defined as:

$$J(\boldsymbol{z}_j) = (\boldsymbol{z}_j - \boldsymbol{z}_j^{\text{prior}})^T C_{zz}^{-1} (\boldsymbol{z}_j - \boldsymbol{z}_j^{\text{prior}}) + (G(\boldsymbol{z}_j) - \boldsymbol{d}_j)^T C_{dd}^{-1} (G(\boldsymbol{z}_j) - \boldsymbol{d}_j), \tag{5.11}$$

where $\boldsymbol{z}_j^{\text{prior}}$ and $\boldsymbol{d}_j$ represent the prior model parameters and the perturbed observed data for the j^{th} ensemble member, respectively. These cost functions are designed to produce an ensemble of models that are coherent with the observed data, thereby assisting in robust DA. For the minimization of the RML cost function $J(\boldsymbol{z}_j)$, we employ the Adam optimizer due to its effectiveness and computational efficiency (Kingma & Ba, 2014).

Applying RML in practice can be computationally intensive as each member of the ensemble necessitates a separate optimization process. Gradient-based optimization methods are commonly used for this purpose, and when adjoints are available for the forward model, they can be applied and increase efficiency. Algorithm 1 describes RML. It takes an initial set of model realizations and observations as inputs. It then computes the relevant covariance matrices and generates variations of the prior model and observed data for each ensemble member. Each ensemble member is then updated by optimizing its respective cost function.

Algorithm 2: RML Algorithm

```
Input  : Initial model set $\boldsymbol{z}_{\text{prior}}$ and perturbed observed data $\boldsymbol{d}_j$
Compute covariance matrices $C_{zz}$ and $C_{dd}$
Generate variations of the prior model $\boldsymbol{z}_j^{\text{prior}}$ and perturbed observed data $\boldsymbol{d}_j$
for $j = 1, \ldots, N_e$ do
    Compute $J(\boldsymbol{z}_j)$ using Equation 5.11
    Optimize $J(\boldsymbol{z}_j)$ using gradient-based optimization (e.g. Adam)
    Compute the final posterior results with DARTS with optimized parameters $\boldsymbol{z}^j$
end for
Output : Posterior history matched states for entire ensemble
```

5.5. HYBRID DATA ASSIMILATION

5.5.1. SURROGATE-BASED HYBRID ESMDA (SH-ESMDA)

ML surrogate models offer computational efficiency but may compromise robustness when used independently for DA. Conversely, ensemble-based DA methods like ESMDA are known for their accuracy but often come at a high computational cost. To preserve the trade-offs between efficiency and robustness, we introduce a surrogate-based hybrid approach called SH-ESMDA, which is within the “Data Learning” paradigm introduced by Buizza et al. (2022). A feature facilitating this integration is the use of the same parameters as input for both the ML algorithm and the reservoir simulator, allowing the surrogate model to serve as a direct substitute for the forward model in the intermediate steps of ESMDA. It is important to note that the surrogate model is only used in the intermediate steps of the ESMDA process. This ensures that the high-fidelity forward model is leveraged for both the initial and the computation of the posterior step, resulting in a more robust and accurate assimilation process. The surrogate model for SH-ESMDA requires training with only the number of forward simulations typically necessary for running the prior in standard approaches. This feature ensures no additional computational burden beyond conventional methods.

The development of SH-ESMDA has the primary objective of accelerating the ESMDA procedure, not achieving more accurate history-matching results than conventional methods. This is due to the fact that in this scheme, ESMDA is still the core DA method, so this hybrid approach also will keep its limitations in terms of DA. The main gain is the acceleration of the process, which potentially enables more iterations, which might otherwise be computationally prohibitive if relying solely on high-fidelity simulations. To achieve this acceleration, the following steps are proposed:

1. **Prior Dataset Generation**: Generate a prior dataset consisting of channelized permeability models with Alluvsim. Subsequently, perform CO_2 injection simulations on these models using the DARTS simulator.

2. **Surrogate Model Training**: Train a surrogate model, such as a FNO or T-UNet, on the generated dataset. This training can be conducted as an offline stage, allowing the pretrained model to be reused in multiple subsequent Hybrid-ESMDA-Surrogate runs, thereby obviating the need for repetitive training and enhancing computational efficiency.
3. **Initial Analysis Step**: Compute the first analysis step with ESMDA using prior forecasts.
4. **Intermediate ESMDA Steps**: Employ the trained surrogate model as a substitute for DARTS in the intermediate steps of the ESMDA process. This offers a computationally efficient approximation to the solution.
5. **Posterior**: Incorporate simulations from DARTS to compute the posterior to refine the solution and compute the final posterior states.

The innovation is leveraging the efficiency of the trained surrogate model to handle the computationally intensive ESMDA iterations. The final step with DARTS simulations acts as a physics-based regularizer to enhance robustness. Another significant advantage of a SH-ESMDA is the capability to pre-train the surrogate model in an offline stage. Once trained, this model can be reused across multiple SH-ESMDA runs without the need for retraining, thereby providing an additional layer of computational efficiency. This feature is particularly beneficial when dealing with a series of similar scenarios, as it eliminates the need to undergo the training process before each new ESMDA run. Consequently, this enables more frequent and rapid iterations, further enhancing the overall efficiency of the DA process. This SH-ESMDA approach is summarized in the algorithm below:

Algorithm 3: Hybrid-ESMDA-Surrogate Algorithm

```
Input  : Initial ensemble, observed data $\boldsymbol{d}_{\text{obs}}$, measurement error covariance $C_D$,
         and trained surrogate model
Determine $N_a$ and $\alpha$ for $i = 1, \ldots, N_a$
for $i = 1$ to $N_a$ do
    if $i = 1$ then
        Compute the prior using DARTS
    else
        Use surrogate model to compute predicted data $\hat{G}(\boldsymbol{m}_j)^f$ for each ensemble
          member
    end if
    Generate perturbed observations $\boldsymbol{d}_j$ using Equation 5.10
    Update ensemble members $\boldsymbol{m}^a$ using Equation 5.9
end for
Output : Compute the final posterior results with DARTS with optimized
         parameters $\boldsymbol{m}^a$
```

5.5.2. SURROGATE-BASED HYBRID RML (SH-RML)

In a similar manner to SH-ESMDA, we introduce a surrogate-based hybrid (SH-RML), a method that integrates ML surrogates into the RML variational framework. One of the critical features that make this integration possible is the use of consistent permeability parameterization for both the surrogate and the high-fidelity DARTS model. This uniformity allows for seamless transitions between the surrogate and the physics-based models during the optimization process. SH-RML employs a streamlined approach similar to SH-ESMDA, where the surrogate model training occurs in the same fashion. It requires only as many forward simulations as are typically necessary in the prior phase of conventional methods. This ensures the method's computational efficiency by aligning with the standard simulation demands, and avoids the need for additional computational resources.

A particular feature of this method is its ability to allow variational DA, even in cases where simulators lack adjoint capabilities. This capability is achieved through the computation of gradients using automatic differentiation from a neural network. The primary objective of the SH-RML approach is to facilitate the optimization of the RML cost function $J(\boldsymbol{m}_j)$. The algorithmic flow of the SH-RML is as follows:

1. **Prior Dataset Generation**: Generate a prior dataset consisting of channelized permeability models with Alluvsim. Subsequently, perform CO_2 injection simulations on these models using the DARTS simulator.

2. **Surrogate Model Training**: Train a surrogate model, such as a FNO or T-UNet, on this dataset. This training can be conducted as an offline stage, allowing the pre-trained model to be reused in multiple subsequent SH-RML runs.

3. **Parameter Initialization**: Initialize the permeability parameters $\boldsymbol{m}_j$ for each ensemble member within the RML framework.

4. **Initial Optimization**: Use gradients derived from the surrogate model to perform initial optimization of the cost function $J(\boldsymbol{m}_j)$.

5. **Posterior Parameter Computation**: Compute the posterior parameters after the surrogate-based optimization using RML.

6. **High-Fidelity Refinement**: Apply these posterior parameters to the DARTS model, running high-fidelity simulations to refine the solution and compute the final posterior states.

The initial steps of the RML optimization process are accelerated by leveraging the surrogate model, which offers both efficiency and automatic differentiation capabilities. The final steps employ the DARTS model to ensure high-fidelity, physics-based solutions as summarized in the algorithm below:

One of the innovations in SH-RML is the utilization of automatic differentiation capabilities provided by the surrogate model. This eliminates the need for manually deriving computationally expensive adjoint models, which are required in traditional variational DA methods. As a result, the SH-RML offers an efficient approach to DA in complex, nonlinear systems. However, it is important to acknowledge the inherent

Algorithm 4: Hybrid-RML-Surrogate Algorithm

```
Input  : Initial model set $m_{\text{prior}}$, perturbed observed data $d_j$, and trained
         surrogate model
Compute covariance matrices $C_{MM}$ and $C_{DD}$
Generate variations of the prior model $m_j^{\text{prior}}$ and perturbed observed data $d_j$
for $j = 1, \ldots, N_e$ do
    Use surrogate to compute initial $J(m_j)$ and gradients
    Perform optimization of $J(m_j)$ using gradient-based methods
end for
Output : Compute the final posterior results with DARTS with optimized
         parameters $m_j$
```

limitations due to the intrinsically ill-posed nature of the problem, affecting the overall results.

In summary, the proposed SH-ESMDA and SH-RML frameworks offer solutions for enhancing DA techniques in applications such as GCS. By adding the computational advantages of ML surrogates with the reliability of physics-based models, these hybrid methods pave the way for efficient and accurate DA and, ultimately, a better understanding and quantification of the model and data uncertainties.

5.6. RESULTS

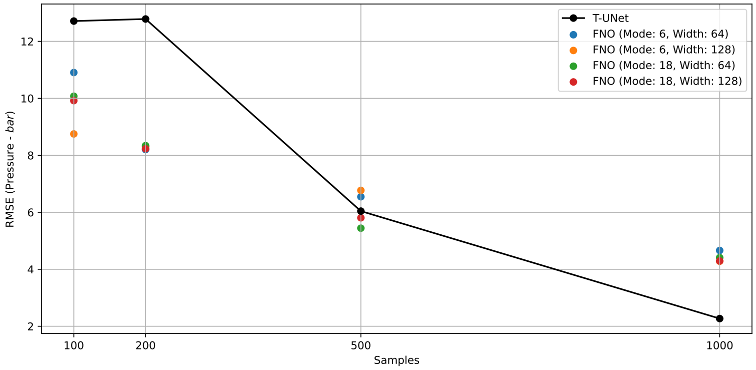
In this section, we present a comprehensive evaluation of surrogate models and history-matching methods for GCS applications. We begin with the training and evaluation of FNO and T-UNet as surrogate models. This is followed by an assessment of the ESMDA method for history matching. Subsequently, we evaluate the SH-ESMDA. Finally, we discuss the SH-RML results.

5.6.1. TRAINING AND EVALUATION OF THE FNO AND T-UNET

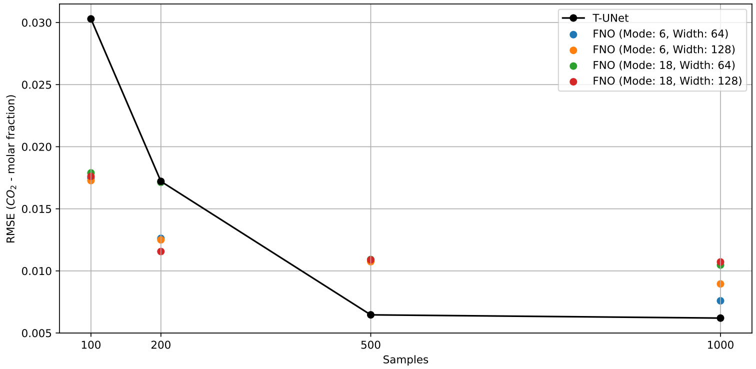
In this Section, the focus is on training the T-UNet and the FNO to serve as surrogate models that approximate the high-fidelity forward model G. As described in Section 5.3, these models are trained on a dataset derived from high-fidelity reservoir simulations generated with the DARTS simulator. This dataset includes critical unknowns such as permeability, porosity, and gas injection rates, along with corresponding state variables, pressure and CO_2 molar fraction. Given the developing state of understanding of optimal FNO configurations, we particularly investigated the impact of varying Fourier modes, considering both modes 18 and 6, and hidden channel widths of 128 and 64. This nuanced examination aims to contribute to the open question of how to best configure FNOs for subsurface modeling tasks.

To quantify the performance of the T-UNet and FNO surrogate models, we RMSE as a metric, focusing on both pressure and CO_2 molar fraction. Specifically, we consider training plus test sample sizes of 100, 200, 500, and 1000 to understand how the size of the training set affects the model's performance. Training and test data are split into

80% and 20%, respectively. The RMSE values presented as a function of the number of training samples in Figure 5.6 indicate that both neural network architectures yield accurate approximations of the high-fidelity forward model G. However, the FNO shows a slight advantage when the number of training samples is limited, particularly at the 100-sample size. This is a significant observation for subsequent data assimilation studies, as it suggests that FNO may require fewer training samples than the T-UNet, thus alleviating the need for additional high-fidelity simulations for neural network training. For context, it's important to note that the pressure range in the reservoir simulations is between 200 and 320 bars, and the CO_2 molar fraction varies from 0 to 1. In this range, the RMSE values indicate that the approximation errors are significantly small. However, these are still approximations and, although they are highly accurate, they can't entirely replicate the high-quality DARTS simulations.



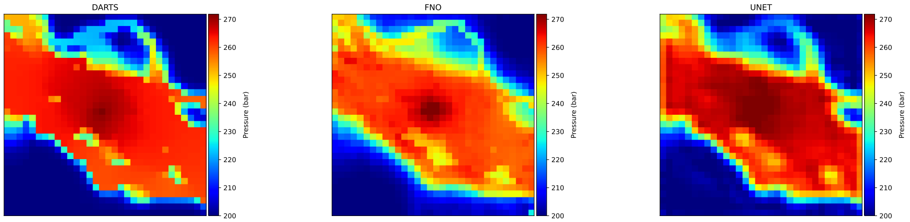
(a) Test RMSE Metrics for Pressure



(b) Test RMSE Metrics for CO_2 molar fraction

Figure 5.6: Test RMSE Metrics for FNO and T-UNet

We also examine the capability of FNO and T-UNet to represent the spatial distributions of pressure and CO_2 molar fraction. Figures 5.7 and 5.8 present the contour maps of CO_2 distribution for a specific test case (not part of the training dataset). At the final time step, we compare the true state variables against those predicted by FNO and T-UNet. Those are responses for the models with a sample size of 1000. Both models capture the overall behavior of the reservoir and what's particularly noteworthy is the ability of both models to approximate the shape of the CO_2 plume. We analyze the models' time evolution in addition to comparing their spatial distributions of pressure and CO_2 molar fraction at a specific point in the grid. Figure 5.9 illustrates the temporal variations at the injection point located at the grid position (16,16). These Figures show the predictive capabilities of both FNO and T-UNet in capturing dynamic behavior, as compared to the high-fidelity DARTS simulations. One observation that stands out is the relative smoothness in the time evolution generated by the FNO model, especially when compared to the more fluctuating curves from the T-UNet model. This difference is intriguing, and it prompts further discussion.

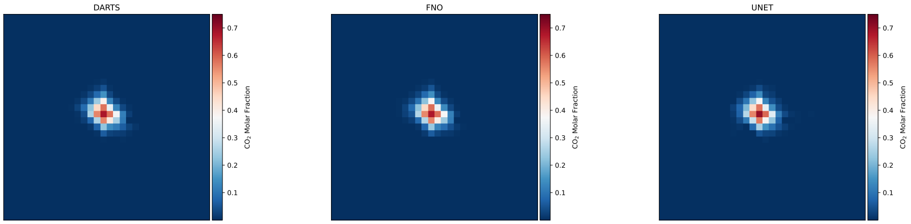


(a) DARTS Pressure

(b) FNO Predicted Pressure

(c) T-UNet Predicted Pressure

Figure 5.7: Pressure Distributions for a test case



(a) DARTS CO_2 molar fraction

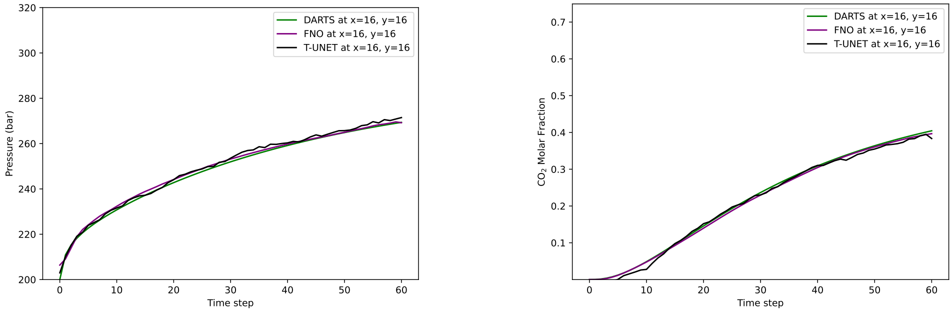
(b) FNO Predicted CO_2 molar fraction

(c) T-UNet Predicted CO_2 molar fraction

Figure 5.8: CO_2 distributions for a test case

The FNO's less noisy, and thus more physically plausible, representation of the temporal evolution at the injection point may be attributed to the way it handles time steps as additional channels of data. This treatment allows the FNO to account for interdependencies across time, thus yielding a smoother, more coherent dynamic response. On the other hand, the T-UNet model generates time evolution independently from each other through integration with the transformer, as discussed in Section 5.3. This approach seems to result in a more noisy representation of the pressure, possibly because the temporal dependencies are not as explicitly captured as in the FNO model.

The difference in noise levels between the FNO and T-UNet models might not just be a matter of the numerical accuracy of the methods, but could also have implications



(a) Time evolution of Pressure at grid (16,16)

(b) Time evolution of CO_2 molar fraction at grid (16,16)

Figure 5.9: Comparison of time evolution of states at grid (16,16) for DARTS, FNO and the T-UNet

for their respective usefulness in subsurface modeling and DA studies. The smoother temporal response of the FNO model may make it more suitable for cases that require a higher accuracy for the physical representation of the states. In summary, the ability of these models to adequately represent both the spatial and temporal complexities of the reservoir suggests their robustness and reliability for further analysis and DA studies. However, the observed differences in dynamic behavior between the FNO and T-UNet models could be a factor in deciding which model to employ for specific applications.

5.6.2. ESMDA HISTORY MATCHING

ESMDA is employed to conduct history matching on our reservoir model described in section 5.2. The primary objective is to evaluate ESMDA's efficiency and accuracy in modeling complex reservoir systems for this GCS application. The task at hand poses challenges due to the nonlinearity of the problem and the non-Gaussian nature of the reservoir's permeability field. To start our assessment, we utilized a distinct reference permeability model, generated outside our prior distribution, to produce synthetic observed data. This model, displayed in Figure 5.10, incorporated a central injection well, surrounded by four pressure monitoring points.

The ESMDA approach utilizes an ensemble of 100 prior permeability maps to represent the uncertainty before DA. These maps are drawn from geological realizations using the Alluvsim algorithm. Over 4 ESMDA steps, the results indicate that the posterior pressure distribution at each of the four monitoring points is closer to the observed data compared to the prior. Figure 5.11 displays this improvement, comparing pressures from the prior, the reference model, and the posterior. We further analyze the sensitivity of the method to the number of steps in ESMDA at the monitoring pressure points for both the prior and posterior models by varying the number of ESMDA steps: 4, 8, 16, and 32. Figure 5.12 reveals that increasing the number of iterations does not substantially improve the quality of history matching.

In terms of uncertainty quantification, ESMDA narrowed the range of uncertainty in pressure estimations. This is visualized in Figure 5.13, where the histograms demonstrate a post-ESMDA variance reduction of the width of the pressure distribution at the final

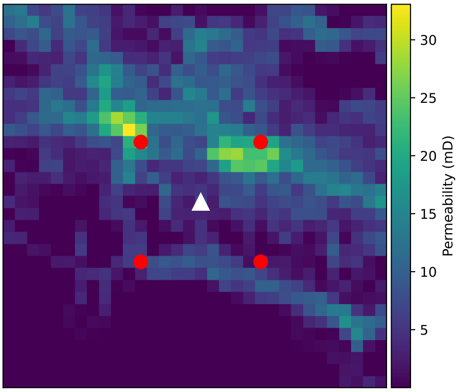


Figure 5.10: Reference permeability model showcasing the central injection well (triangle) and peripheral monitoring points (circles). Monitoring points are numbered from 1 to 4, starting from the top-left corner and proceeding clockwise.

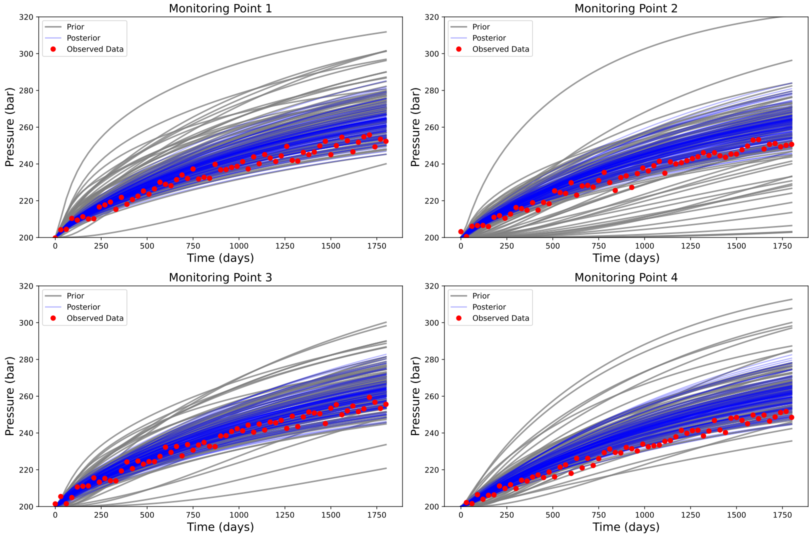


Figure 5.11: Comparison between the prior, reference model, and posterior pressures at each monitoring point for the ESMDA history matching. Red dots represent a realization of perturbed observed data.

injection step, as compared to the broader spread of the prior. The observed pressure values, indicated by the red dashed line, are now more centrally situated within the tightened posterior distribution, signifying how the observations reduce the uncertainty of pressure buildup within the reservoir. The true pressures observed at the monitoring points are as follows: 255.1 bar at Monitoring Point 1, 252.0 bar at Monitoring Point 2,

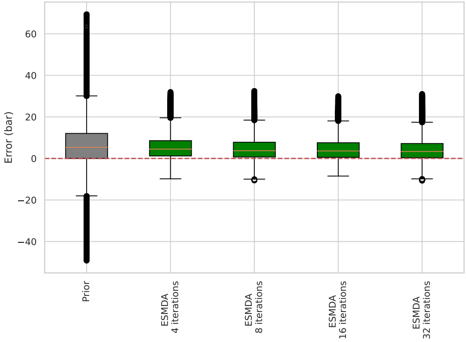


Figure 5.12: Comparison of monitoring pressure absolute error across different ESMDA iterations and the prior. Red dots represent a realization of perturbed observed data.

255.0 bar at Monitoring Point 3, and 251.4 bar at Monitoring Point 4. The P10-P90 pressure range, which serves as an indicator of uncertainty, is reduced after the history matching process, as quantified in Table 5.1.

Table 5.1: Uncertainty reduction in P10-P90 pressure range at monitoring points for the ESMDA history matching with 4 iterations.

Location	**Prior (bar)**		**Posterior (bar)**	
	P10-P90	**Difference**	**P10-P90**	**Difference**
1	213.9 - 268.8	54.9	212.6 - 261.1	48.5
2	202.7 - 261.2	58.4	211.0 - 259.9	48.9
3	209.9 - 265.1	55.2	211.5 - 260.2	48.7
4	208.7 - 264.0	55.3	208.8 - 259.0	50.2

Although ESMDA reduces errors related to measured pressure in comparison to the prior, it significantly overestimates reservoir permeability in comparison to prior permeability distributions, as illustrated in Figure 5.14. While undesirable, the discrepancy can be explained by the fact that history matching is an ill-posed problem, allowing for multiple solutions that can satisfactorily fit the data.

Given the inherent assumptions of ESMDA, namely, its reliance on Gaussian distributions and its better suitability for linear problems—none of which are present in our case—it is important to recognize its limitations in addressing the ill-posed problems we encounter. Our subsequent results delve into hybrid methods, as detailed in Section 5.5. These techniques serve dual purposes: one aims to accelerate the computational process, while the other focuses on enhancing the accuracy of history matching.

5.6.3. RESULTS FOR SH-ESMDA

In order to accelerate the computational efficiency of the standard ESMDA, we employ ML surrogates in the form of FNO and T-UNet following the algorithm outlined in Section 5.5.1. For this study, we employ 100 samples and conduct history matching for pressure

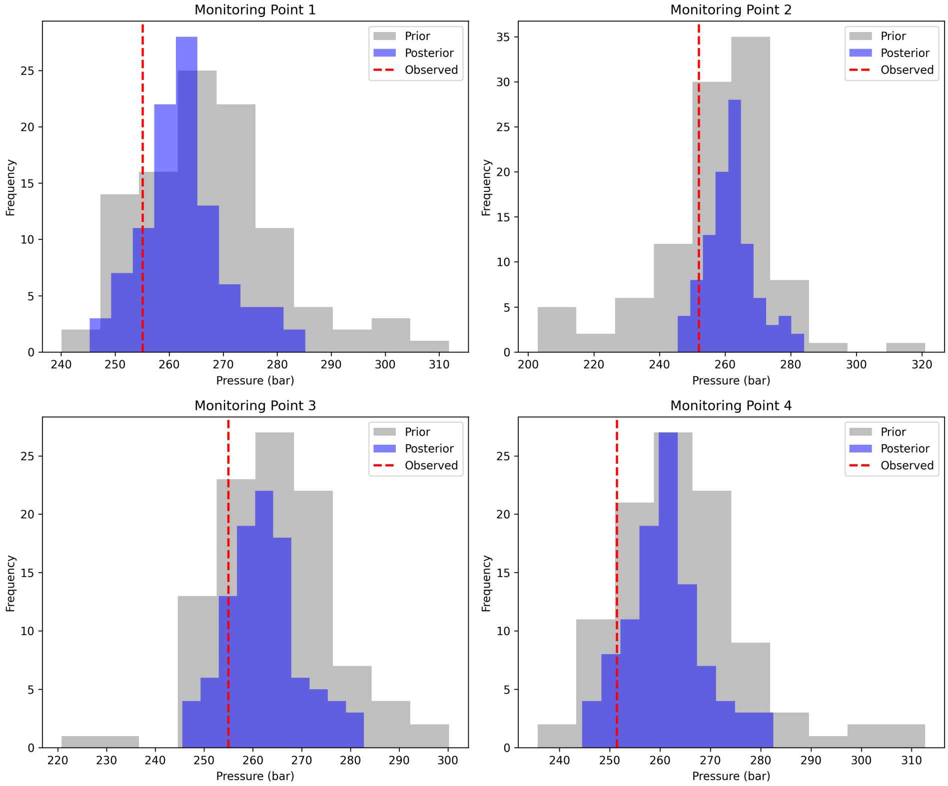


Figure 5.13: Histograms of prior and posterior pressure distributions at the last time step of injection for each monitoring point for the ESMDA history matching with 4 iterations. The observed reference pressures are marked with a red dashed line.

at four monitoring points over four iterations. Figures 5.15 and 5.16 display the results of these history-matching exercises. In this context, it should be highlighted that the larger-scale training experiments involving additional models were designed for performance benchmarking of SH-ESMDA method. These Figures reveal the performance when FNO and T-UNet are used as surrogate models for the intermediate steps in the ESMDA algorithm, respectively. As can be observed, the outcomes produced by both FNO and T-UNet are remarkably similar to each other and closely align with those obtained using the standard ESMDA methodology with 4 steps.

Figure 5.17 presents the histograms of pressure at the last time step of injection for SH-ESMDA using the FNO surrogate model with 4 iterations, illustrating a variance reduction of the pressure distributions around the observed values at each monitoring point, results similar to what was achieved with ESMDA. This is quantified further in Table 5.2, which presents the P10-P90 pressure range before and after applying SH-ESMDA with FNO and 4 iterations. The reduction in uncertainty ranges from 5.8 bars at Monitoring Point 1 to 3.9 bars at Monitoring Point 4.

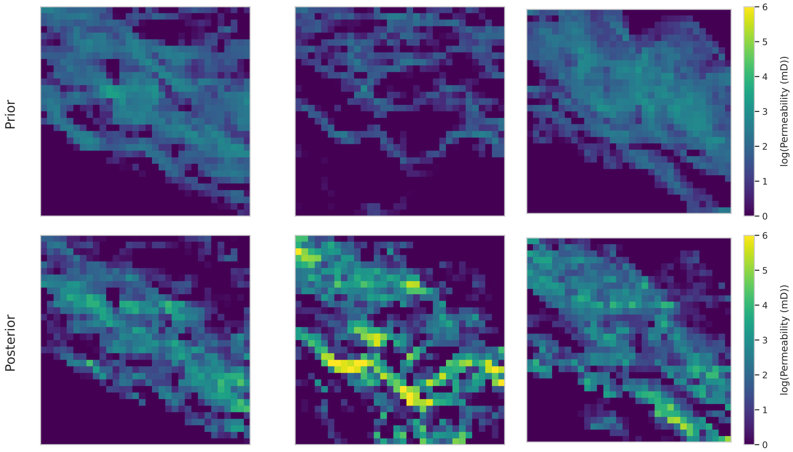


Figure 5.14: Comparison between prior and posterior permeabilities across three different samples for the ESMDA.

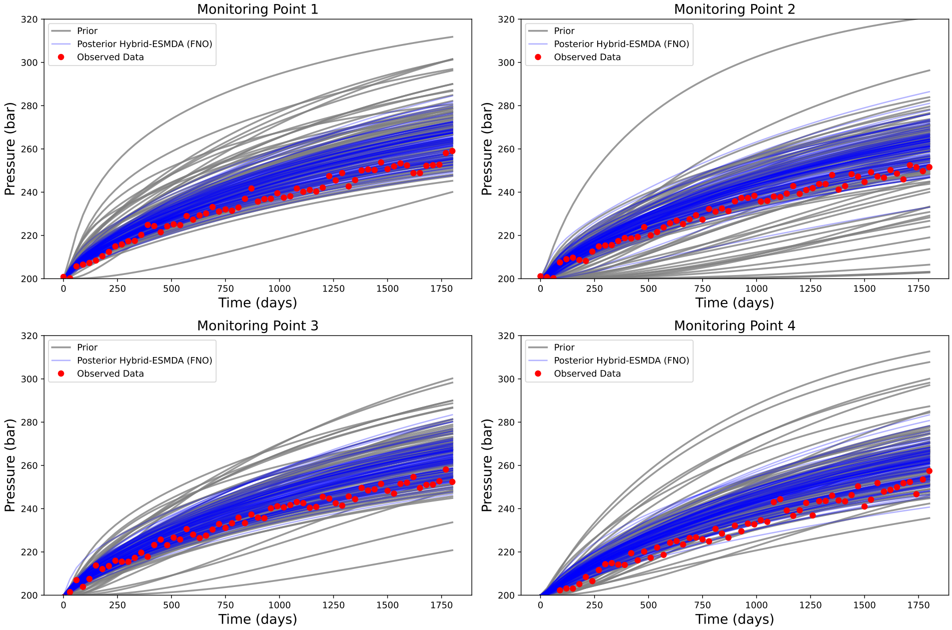


Figure 5.15: Comparison between the prior, reference model, and posterior pressures at each monitoring point for the Hybrid-ESMDA-Surrogate history matching using FNO. Red dots represent a realization of perturbed observed data.

Figure 5.18 presents box plots that offer a comparison of the monitoring pressure errors for both the standard ESMDA and the enhanced SH-ESMDA methodology, em-

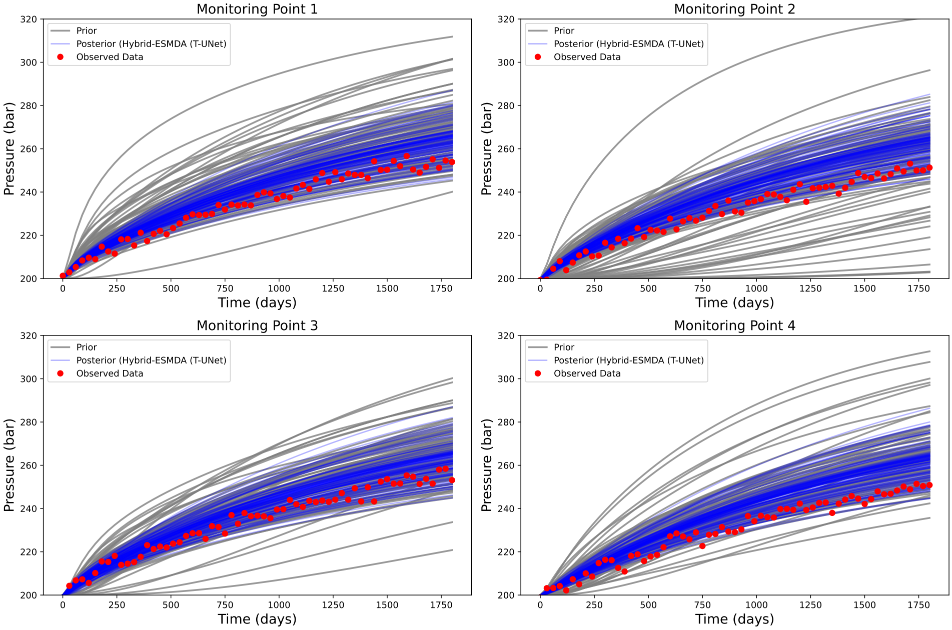


Figure 5.16: Comparison between the prior, reference model, and posterior pressures at each monitoring point for the SH-ESMDA history matching using T-Unet. Red dots represent a realization of perturbed observed data.

Table 5.2: Uncertainty reduction in P10-P90 pressure range at monitoring points for the SH-ESMDA history matching with 4 iterations.

Location	**Prior (bar)**		**Posterior (bar)**	
	P10-P90	**Difference**	**P10-P90**	**Difference**
1	213.9 - 268.8	54.9	212.96 - 262.1	49.14
2	202.7 - 261.2	58.4	210.93 - 260.8	49.87
3	209.9 - 265.1	55.2	211.65 - 261.0	49.39
4	208.7 - 264.0	55.3	209.10 - 260.1	50.97

ploying FNO and T-UNet as surrogates. The SH-ESMDA methods maintain a level of accuracy that is comparable to that of the standard ESMDA, while significantly reducing the computational time required. When the T-UNet model is employed as the surrogate, the errors observed are marginally higher compared to when the FNO is used. It is also important to highlight that, consistent with the observations made for the standard ESMDA, increasing the number of iterations does not result in a substantial reduction in errors.

The gains in computational efficiency are illustrated in Figure 5.19. Both versions of the SH-ESMDA method—utilizing either FNO or T-UNet as surrogates significantly reduce computation time. Specifically, there is a minimum speedup of around 50%. In all experiments, the ESMDA involved 4 steps. The computational times reported do not

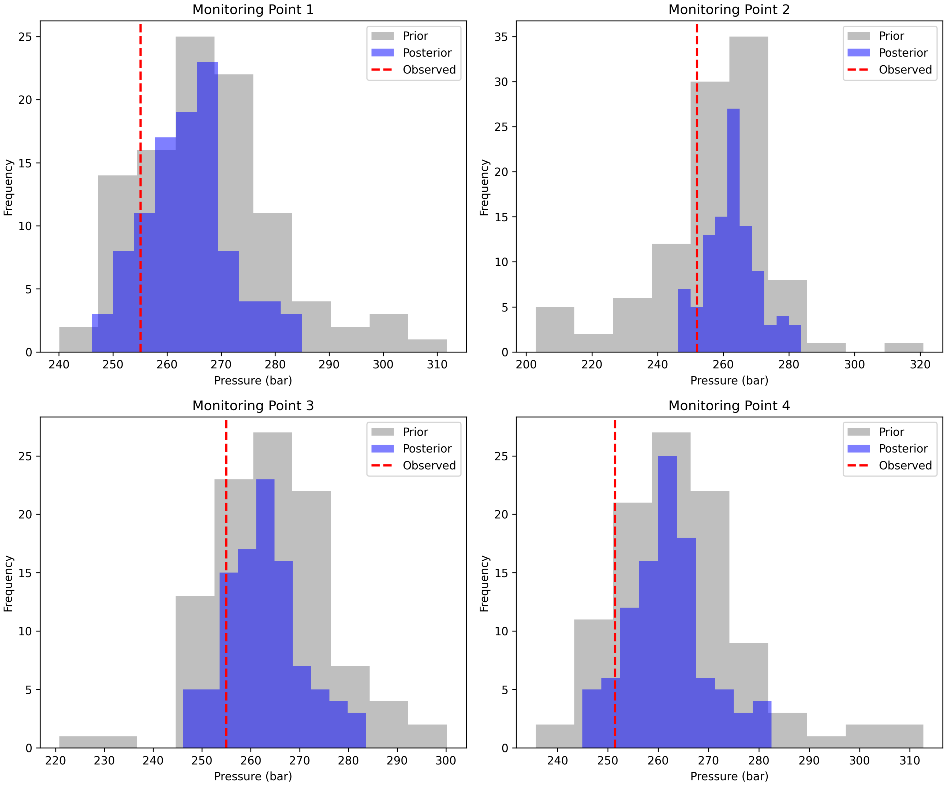


Figure 5.17: Histograms of prior and posterior pressure distributions at the last time step of injection for each monitoring point for the SH-ESMDA history matching with 4 iterations. The observed reference pressures are marked with a red dashed line.

include the time required to train the FNO and T-UNet models. The exclusion is justified for two main reasons. First, the training of these neural network models is generally considered a one-time computational expense. For the configuration used in our hybrid methods—comprising 100 samples for both FNO and T-UNet, along with 6 models and a width of 64—both models can be trained in less than 30 minutes. Once trained, these models can be reused for multiple iterations or different scenarios without the need for retraining. This is particularly beneficial in applications where the same or similar systems are studied multiple times. Second, the training can be performed offline and in parallel, taking advantage of high-performance computing resources. This minimizes its impact on the overall computational efficiency when amortized over multiple applications.

The posterior permeability distributions obtained through both the SH-ESMDA schemes employing FNO and T-UNet closely align with those achieved using the standard ESMDA approach. Given that ESMDA serves as the core method for DA in these hybrid algorithms, the permeability distributions inherently reflect the strengths and limitations

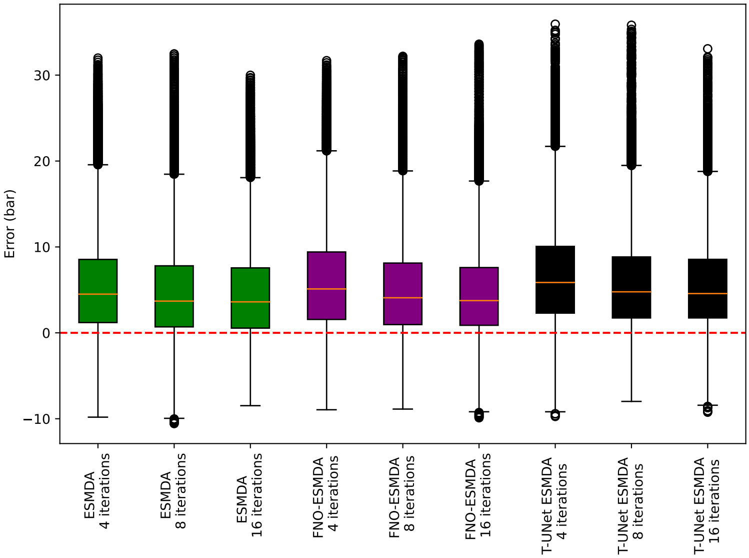


Figure 5.18: Box plots comparing monitoring pressure errors between standard ESMDA (green) and SH-ESMDA using FNO (purple) and T-UNet (black) surrogates for 4, 8 and 16 iterations.

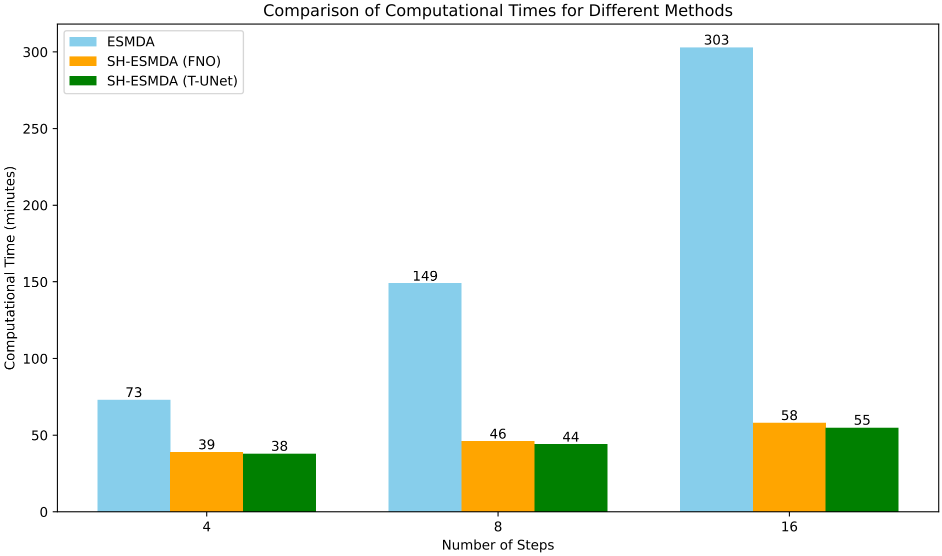


Figure 5.19: Comparison of computational times for ESMDA and SH-ESMDA methods using FNO and T-UNet surrogates. The bar chart shows the runtime in minutes for different numbers of assimilation steps, demonstrating the significant time savings achieved by the hybrid methods.

of ESMDA itself. Therefore, the hybrid schemes do not necessarily bring about a qualitative shift in the outcome; rather, their primary advantage lies in computational speedup. We acknowledge that the hybrid methods will inherit the constraints of ESMDA, and its inadequacies in encapsulating the uncertainties inherent in the system. For brevity and to avoid redundancy, we refrain from presenting additional Figures showcasing the similarity in the posterior permeability distributions across the different methods, as they would closely mirror the results already discussed for ESMDA.

In summary, SH-ESMDA successfully combines the computational efficiency of ML with the reliability of DARTS like ESMDA. The approach reduces the time needed for each computational step, allowing for less time for the same amount of iterations. However, the SH-ESMDA inherits the fundamental weaknesses of the core ESMDA algorithm. Therefore, although the method provides substantial acceleration, the accuracy of the history matching is still constrained by the inherent limitations of ESMDA. This makes this SH-ESMDA scheme an important tool for accelerating DA studies, especially for complex reservoir models.

5.6.4. RESULTS FOR SH-RML

As elaborated in Section 5.5.2, SH-RML offers a combination of factors that make it an effective choice for accurate DA for complex problems. Here, we explore further its comparative results over other methods like ESMDA. Our findings show that the SH-RML improves history matching and provides more accurate posterior permeabilities estimates. For a comprehensive perspective on the computational requirements of SH-RML, it's pertinent to discuss the computational time in the SH-RML. Both FNO and T-UNet models were subjected to 200 steps each for gradient evaluation in the optimization process for each one of the 100 prior models, and the full reservoir simulations with DARTS were also taken into account for the prior and the posterior curves. For SH-RML, it should be noted that the extensive training experiments, which included up to 1000 models, were primarily focused on performance benchmarking. These experiments are not a requirement for the framework but serve to showcase its adaptability and effectiveness in diverse application scenarios. The overall history-matching process requires approximately 8 hours, which is justified by the enhanced accuracy and reliability of the estimates obtained. Nonetheless, the optimization time could potentially be reduced by employing a more tailored approach to the cost function optimization without significantly compromising the quality of the results.

For SH-RML, we noticed an improvement in history matching with posterior pressure closer to the values of the observations. Figure 5.20 and Figure 5.21 show the pressure matching results at each monitoring point for SH-RML with FNO, and SH-RML with T-UNet, respectively. While surrogates are employed for gradient evaluations, the final posterior pressure curves are computed using the full reservoir simulator (DARTS). This allows SH-RML to leverage the efficiency of surrogates during optimization while still generating high-fidelity pressure forecasts with the simulator for final uncertainty quantification. Figure 5.22 displays box plots comparing monitoring pressure errors between the standard ESMDA and SH-RML methods. The results from SH-RML exhibit a better balance around zero when compared to ESMDA. Figure 5.24 demonstrates that the posterior permeabilities derived from SH-RML, employing FNO are consistent with the prior. In contrast to ESMDA, RML provides permeabilities within a similar range as the prior. For conciseness and to avoid repetition, we do not include additional Figures showcasing the posterior permeabilities using the T-Unet in SH-RML. The outcomes closely mirror those already presented for the FNO version, as well as the enhanced uncertainty quantification compared to ESMDA.

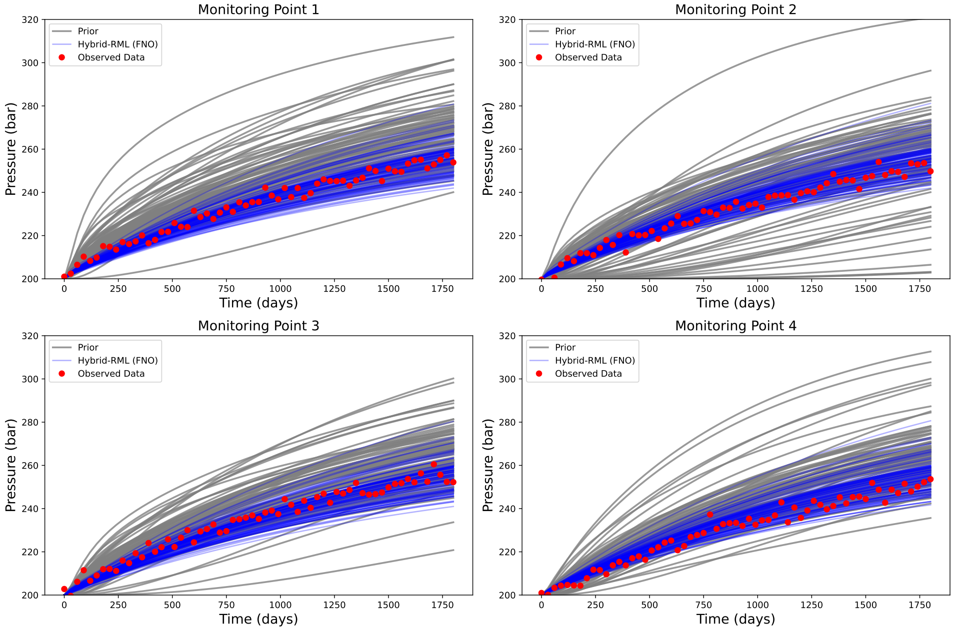


Figure 5.20: Comparison between the prior, reference model, and posterior pressures at each monitoring point for the Hybrid-RML-Surrogate history matching using FNO. Red dots represent a realization of perturbed observed data.

Figure 5.23 presents the histograms of pressure at the last time step of injection for SH-RML using the FNO surrogate model. These histograms demonstrate a better reduction in uncertainty than what was observed with both ESMDA and SH-ESMDA. This is quantitatively supported by the data in Table 5.3, which shows the P10-P90 pressure range reduction across the monitoring points. To maintain conciseness and since the outcomes between the FNO and T-UNet models are similar, we focus on presenting results from the FNO model only.

In summary, the results of this case study demonstrate that SH-RML offers improvements over ESMDA for history matching and uncertainty quantification. By leveraging the gradients of ML surrogates for optimization of the cost functions, SH-RML achieves better pressure matching while also providing more accurate estimates of posterior permeability distributions. The integration of surrogates enables the computations of gradient evaluations, enabling thorough optimization of the RML objective. Developing on these results, our study contributes to and extends the current understanding of machine learning integration with data assimilation in geological carbon storage. It aligns with the data

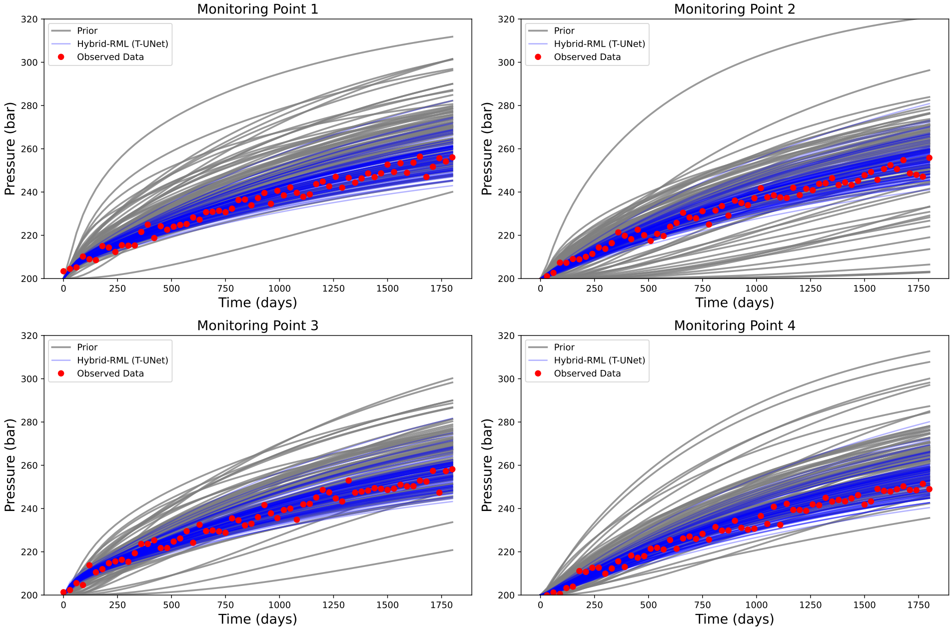


Figure 5.21: Comparison between the prior, reference model, and posterior pressures at each monitoring point for the Hybrid-RML-Surrogate history matching using T-UNet. Red dots represent a realization of perturbed observed data.

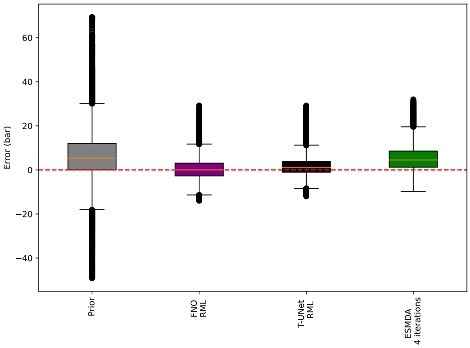


Figure 5.22: Box plots comparing monitoring pressure errors between standard RML and Hybrid-RML-Surrogate using FNO and T-UNet surrogates.

learning methods framework proposed by Buizza et al. (2022), contributing to the growing body of work on integrating machine learning with data assimilation. Similar to the approaches by Tang et al. (2022) and Tang et al. (2021), our methods apply machine learning models to enhance data assimilation processes. However, our approach distinguishes itself by ensuring that the posterior is physics consistent, addressing a common limitation

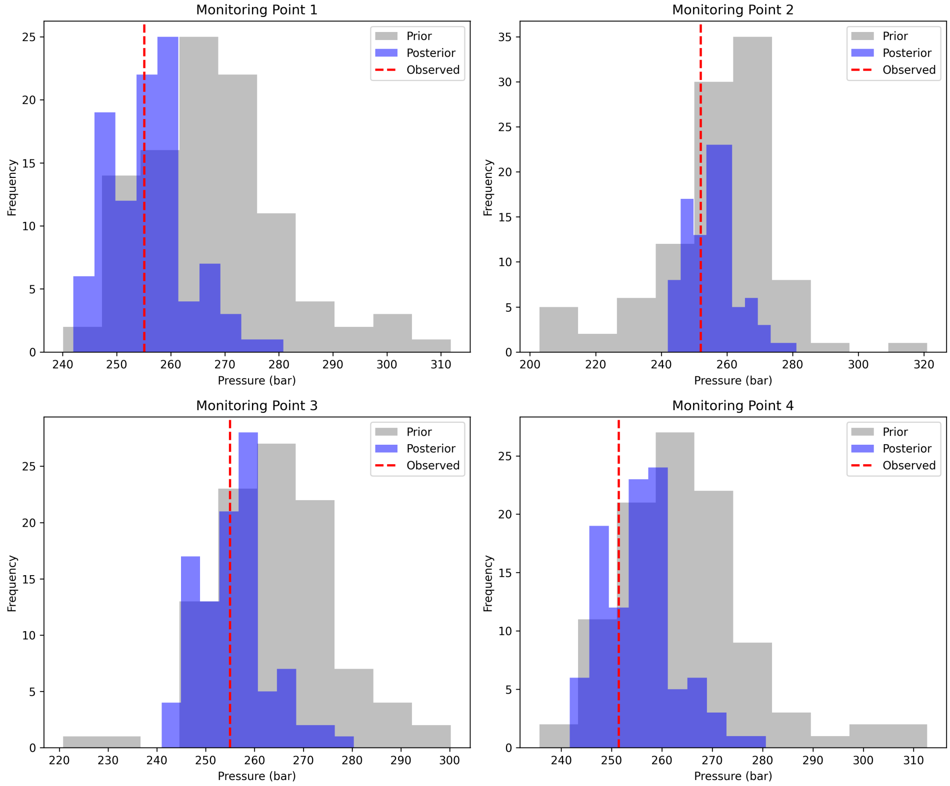


Figure 5.23: Histograms of prior and posterior pressure distributions at the last time step of injection for each monitoring point for SH-RML with the FNO surrogate model.

Table 5.3: Uncertainty reduction in P10-P90 pressure range at monitoring points for SH-RML with FNO.

Location	**Prior (bar)**		**Posterior (bar)**	
	P10-P90	**Difference**	**P10-P90**	**Difference**
1	213.9 - 268.8	54.9	209.0 - 253.9	44.9
2	202.7 - 261.2	58.4	208.7 - 253.7	45.0
3	209.9 - 265.1	55.2	208.2 - 253.5	45.3
4	208.7 - 264.0	55.3	208.2 - 253.5	45.3

in previous studies, which rely heavily on surrogate models for data assimilation without adequately ensuring physical consistency (Tang et al., 2021; Tang et al., 2022). In addition, our work demonstrates the feasibility of employing a combination of transformers with UNets in CO_2 storage problems, a method similar to that proposed by Li et al. (2023). This novel application within the field of geological carbon storage emphasizes the adaptability and potential of advanced neural network architectures in handling complex spatial and temporal data patterns typical of subsurface environments. Furthermore, while employ-

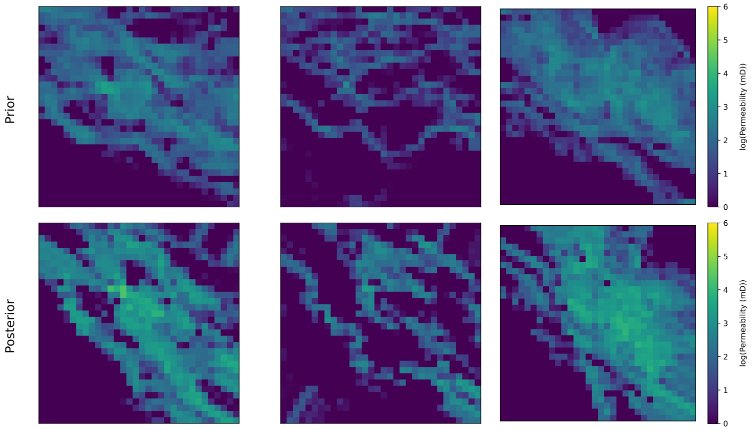


Figure 5.24: Posterior permeabilities for Hybrid-RML-Surrogate with FNO surrogates.

ing FNOs, we demonstrated the potential success of overcoming the challenge of data demand for training these models, which has been an issue in related works (**Wen2021**; Crain et al., 2024; Tang et al., 2022). Our approach not only reduces the amount of data required but also enhances the computational efficiency of the data assimilation framework, thereby offering a robust solution to one of the major hurdles in applying deep learning techniques to geoscience applications.

5.7. Discussion

This study presents advancements in combining DA and ML to enhance history matching in CO_2 storage projects, addressing challenges and limitations inherent in existing methodologies. Our innovative hybrid frameworks, namely SH-ESMDA and SH-RML, optimize both computational efficiency and accuracy in DA studies. The integration of ML surrogates enabled the application of ensemble-based DA methods like ESMDA efficiently to complex subsurface systems. Specifically, the SH-ESMDA method has accelerated ESMDA computations while maintaining the physical consistency of posterior physical responses. The SH-RML method has excelled in achieving superior history matching compared to standard ESMDA, attributed to the better approximation of the gradients for non-linear and non-Gaussian problems. This method leveraged the automatic differentiation inherent in ML models, enabling gradient-based optimization and overcoming the challenges posed by the absence of adjoint models in reservoir simulators.

A consideration during this study was the choice of grid resolution. In the preliminary stages, models with resolutions of 256 × 256, 128 × 128, and 64 × 64 were examined. The choice of a 32 × 32 resolution was influenced by extended computational times required for numerous runs at higher resolutions and substantial memory requirements of the FNO. Although subsequent updates to the computational tools have mitigated these challenges, allowing for exploration at higher resolutions, a substantial portion of the work presented was conducted using the initial setup and resolutions due to the constraints at the time. The chosen resolution effectively captures the variability of the permeability of channelized reservoirs, offering a balance between computational demand and model fidelity.

While we recognize that vertical gradients and gravity-driven flow are needed in more comprehensive CO_2 storage models, the 2D horizontal model allows the examination of the areal and non-Gaussian characteristics of permeability distributions typical of geological formations used for CO_2 storage. These formations often feature highly heterogeneous and non-Gaussian permeability distributions due to natural depositional processes, which are effectively represented in the simplified 2D channelized models. By focusing on horizontal pressure distribution within the reservoir, we address the primary dynamics of CO_2 flow, where lateral variations in permeability significantly influence the migration and trapping of CO_2. This approach facilitates the analysis of horizontal flow behaviors, providing essential insights into lateral connectivity and barrier effects on CO_2 plume behavior without the complicating factors of vertical segregation. The 2D model aligns with the goals of developing a new computational framework and sets a clear path for future 3D extensions, which can incorporate vertical dynamics and gravity effects, enhancing the framework's applicability to real CO_2 storage projects. The insights gained from these 2D horizontal simulations provide a baseline for the development of more complex models, ensuring that the underlying data assimilation and machine learning techniques are robust and well-understood before adding additional layers of complexity.

Reflecting on the characteristics and applications of each method, the SH-ESMDA is optimal for scenarios prioritizing computational speed. At the same time, the SH-RML is particularly suited for problems necessitating more accurate gradient computations, even when forward models lack adjoints. The unified parameterization of permeability between ML and physics-based models has facilitated integration, ensuring physical consistency and reliability in the posterior solutions computed using the high-fidelity DARTS simulator. However, it's important to acknowledge some limitations of the methods. The SH-ESMDA inherits the fundamental constraints from ESMDA in capturing uncertainties, and the accuracy of both hybrid methods is contingent on the precision of the ML surrogates, which may encounter challenges in extrapolation. Addressing these limitations could potentially be achieved by incorporating more robust and nonlinear DA methods, enhancing the reliability of uncertainty quantification. A limitation of the SH-RML is the dependency on the accuracy of the ML surrogates for gradient computation of specific points of the reservoir. If the surrogates are not well-trained or accurate, the gradients derived from them could be unreliable, leading to suboptimal or incorrect solutions and impacting the method's overall effectiveness.

The implications of this uncertainty reduction relate to the core objectives of this study since a reduction in the uncertainties of the pressure buildup in the reservoir is important to ensure operations that preserve the integrity of caprock to prevent CO_2 leakage and mitigate the risk of induced seismicity and fault reactivation.

Our comparative assessment between FNO and T-UNet has demonstrated that both architectures exhibit low RMSE compared to DARTS simulations, with FNO having a slight edge, particularly when training data is limited. This insight is essential for scientists to choose the most suitable ML architectures effectively, especially when considering data availability constraints. Beyond CO_2 storage, the versatility of the developed methodologies is evident, with potential adaptability to other applications including geothermal energy and nuclear waste disposal. The advancements made in this study are poised to catalyze the broader adoption and application of hybrid DA-ML methodologies across diverse scientific domains, marking a significant step forward in the field.

To provide a comprehensive overview of the methodologies assessed in this study, we have included a comparative summary in Table 5.4.

Table 5.4: Summary of Methods and Their Characteristics

Method	Category	Advantages	Disadvantages
Fourier Neural Operators (FNO)	ML Model	Slightly better performance with limited data. Efficient in capturing temporal dynamics.	Potentially higher computational cost than T-UNet.
Transformer UNet (T-UNet)	ML Model	Good at integrating spatial and temporal data. Effective in handling complex patterns.	Can be more sensitive to small training datasets compared to the FNO.
Ensemble Smoother with Multiple Data Assimilation (ESMDA)	DA Method	Doesn't require gradients. Easy implementation.	Assumes Gaussianity, which can limit its effectiveness in non-Gaussian systems.
Surrogate-based Hybrid ESMDA (SH-ESMDA)	DA Method	Significantly reduces computational time (by over 50%). Maintains physical consistency of the model.	Inherits limitations of the core ESMDA method, such as issues with non-Gaussian distributions.
Surrogate-based Hybrid RML (SH-RML)	DA Method	Provides improved history matching and uncertainty quantification. Avoids the need for adjoint derivations thanks to ML surrogates.	Dependency on the accuracy of ML surrogates for gradient computations.

5.8. CONCLUSION

This paper introduced novel frameworks for enhancing uncertainty quantification (UQ) in Geological Carbon Storage (GCS) projects through the integration of machine learning (ML) and data assimilation (DA) techniques. We evaluated two neural network architectures, the Fourier Neural Operators (FNOs) and Transformer-UNet (T-UNet), for generating accurate and efficient surrogate models of CO_2 injection simulations. Comparative analyses revealed the FNO's slight superiority when training data is limited.

Leveraging these surrogates, we proposed two hybrid methods. Surrogate-based hybrid ESMDA (SH-ESMDA) incorporates the ML models into the Ensemble Smoother with Multiple Data Assimilation (ESMDA), reducing computational time by over 50% while maintaining accuracy. Surrogate-based Hybrid RML (SH-RML) enables variational data assimilation by using automatic differentiation from the neural networks for gradient calculations in the Randomized Maximum Likelihood (RML) optimization. This avoids manual adjoint derivations.

Results showed that SH-RML achieved improved history matching and uncertainty quantification compared to standard ESMDA. The FNO surrogate enabled efficient computation of gradients for the RML optimization. The proposed frameworks thus enhance DA for GCS by integrating machine learning efficiency with the physical reliability of reservoir simulators.

In conclusion, while our study presents significant advancements in integrating AI with DA for GCS applications, it also opens avenues for future research. Particularly, exploring the scalability of our methodologies to larger, more complex reservoir systems, such as real carbonate formations (Burchette, 2012), and real field cases from current GCS projects and investigating the integration of additional neural network architectures, could yield further valuable insights in optimizing GCS operations. The framework shows versatility beyond CO_2 sequestration, presenting opportunities for adaptation to other subsurface modeling applications like geothermal energy and nuclear waste storage. Overall, this study demonstrates the potential of combining machine learning and physics-based models to tackle problems in uncertainty quantification for the energy transition.

CODE AND DATA AVAILABILITY

The Alluvsim-generated channelized reservoir dataset is available at https://doi.org/10.4121/fd03d033-1d73-4b9a-91b4-b47e9a102a04. The code for both the T-UNet and FNO architectures is publicly accessible at https://github.com/nmucke/subsurface-DA-with-generative-models. The open-DARTS simulator is available at https://gitlab.com/open-darts/open-darts. The dageo Python package for data assimilation in geosciences is available at https://github.com/tuda-geo/dageo.

6

INTEGRATING SCORE-BASED DIFFUSION MODELS WITH ML-ENHANCED LOCALIZATION FOR DATA ASSIMILATION IN GCS

Maintaining geological realism during data assimilation is difficult in channelized reservoirs used for carbon storage. We combine score-based diffusion models, which rapidly generate realistic ensemble members, with ML-enhanced localization that adapts to channel geometry. After training, diffusion supplies a large synthetic ensemble ($N_s = 5000$) used to estimate covariances; the ML proxy turns those samples into localization weights that follow connectivity where distance tapers do not. Tests on Gaussian and channelized fields show 15–40% better variance preservation with comparable data matches. This matters for storage projects in complex formations, where keeping non-Gaussian features is essential for reliable uncertainty quantification.

This chapter forms part of the manuscript: Gabriel Serrão Seabra, Nikolaj T. Mücke, Vinícius L. S. Silva, Alexandre A. Emerick, Denis Voskov, Femke C. Vossepoel, Integrating Score-Based Diffusion Models with Machine Learning-Enhanced Localization for Advanced Data Assimilation in Geological Carbon Storage, *International Journal of Greenhouse Gas Control* – preprint (2025). Portions also appear in Vinicius L. S. Silva, Gabriel S. Seabra, Alexandre A. Emerick, Integrating Score-Based Diffusion Models with ML-Enhanced Localization for Data Assimilation in GCS, *Journal of Computational Physics* – preprint (2025) (Silva et al., 2025a). Changes have been applied to make the text and figures consistent with the thesis.

6.1. INTRODUCTION

Wells typically sample a small part of the reservoir volume, yet decisions affecting CO_2 containment must be made based on this limited information. This data scarcity becomes particularly challenging in channelized formations, which can be attractive storage targets because of their high injectivity, where complex connectivity patterns defy simple statistical descriptions (Pyrcz & Deutsch, 2014).

Existing geostatistical methods face inherent limitations when confronted with channelized systems. Two-point statistics cannot capture curvilinear channel geometries, whereas multiple-point statistics require training images that may poorly represent the true subsurface heterogeneity (Mariethoz & Caers, 2014; Strebelle, 2002). When these prior models are updated through ensemble-based data assimilation, the limited number of ensemble members exacerbates these problems. Small ensemble sizes (typically $N_e \sim 50-200$ in operational settings) lead to poorly estimated covariance matrices with significant sampling errors, causing the assimilation update to introduce geologically implausible features, such as spurious high-permeability patches. In practice, N_e is capped by simulation cost per member, parallel scalability limits, and project turnaround requirements. The insufficient ensemble size worsens the results in artificial blurring of sharp geological boundaries, as the linear update mechanism cannot preserve the discrete nature of channel-background contrasts.

This geological degradation becomes more severe with smaller ensembles, where sampling errors dominate the true covariance structure, ultimately producing posterior realizations that bear little resemblance to the complex channelized patterns that control the subsurface flow behavior. Although object-based geological models, such as FLUVSIM (Deutsch & Tran, 2002), can generate geologically realistic channelized systems that address these representational limitations, they are computationally expensive and time-consuming to run, making it impractical to generate the hundreds or thousands of realizations required on-the-fly during data assimilation workflows. Across the thesis we employ two object-based geostatistical generators to create channelized priors: FLUVSIM (Deutsch & Tran, 2002) and ALLUVSIM (Pyrcz et al., 2009). The choice of generator is exploratory rather than prescriptive, reflecting our aim to probe the data-assimilation methods under different implementations of fluvial pattern synthesis. In this chapter we rely on FLUVSIM to produce the channelized training set and prior realizations used by the diffusion+ML-localization workflow, whereas Chapter 5 used ALLUVSIM within the surrogate-assisted experiments. Our focus here is the assimilation framework itself, not a comparison between generators.

Data assimilation methods provide the mathematical framework for combining these prior geological models with observed data to reduce uncertainty in reservoir characterization (Evensen et al., 2022; Oliver et al., 2008). The fundamental approaches can be broadly categorized into variational methods, which solve optimization problems to find the most probable model realization (Tarantola, 2005), and ensemble methods, which use an ensemble to approximate probability distributions (Evensen, 2009). Ensemble methods quantify uncertainty through their sample-based representation, making them suitable for GCS applications where understanding the range of possible outcomes is as important as identifying the most likely scenario (Aanonsen et al., 2009). The evolution from the original Ensemble Kalman Filter (Evensen, 1994) to smoothers and iterative variants reflects ongoing efforts to handle increasingly complex nonlinear problems while maintaining computational feasibility (Chen & Oliver, 2012; Emerick & Reynolds, 2013b). Advances address the challenge of limited ensemble sizes through localization techniques (Houtekamer & Zhang, 2016) and iterative schemes that linearize nonlinear problems (Oliver & Chen, 2011). However, these methods still face challenges when applied to highly heterogeneous channelized systems, where traditional assumptions about spatial correlation and Gaussian distributions break down.

The Ensemble Smoother with Multiple Data Assimilation (ESMDA) is a common method of choice for practical reservoir characterization due to its computational efficiency and ability to handle nonlinear forward models (Emerick & Reynolds, 2013b; Evensen, 2009). However, ESMDA has Gaussian assumptions that channelized permeability fields severely violate, and this can exacerbate the fact that small ensembles lose their variance, causing posterior models to be less uncertain than they should be, which affects the reliability of uncertainty quantification. The industry has widely adopted distance-based localization methods (Gaspari & Cohn, 1999; Hamill et al., 2001; Houtekamer & Zhang, 2016) as the primary solution to combat ensemble collapse. Covariance estimation in high-dimensional systems (Rukhin, 2007) and distance based methods, including the popular Gaspari-Cohn correlation function, are standard practice in several operational reservoir history matching systems.

While those approaches suppress spurious correlations arising from sampling errors, they rely on an assumption that correlation strength decreases with physical distance — an assumption that fails in channelized systems. A circular or elliptical localization kernel cannot distinguish between updating along a high-permeability channel versus across a low-permeability barrier, leading to geologically inconsistent modifications that compromise long-term predictions. Despite these limitations, distance-based localization remains the industry standard due to its computational efficiency and ease of implementation. Recent work has shown the benefits of applying covariance localization to field cases (Emerick & Reynolds, 2011) and comparing different adaptive localization methods for reservoir characterization (Lacerda et al., 2019).

Score-based diffusion models, generative AI models that learn to approximate the gradient of the log-density ($\nabla_x \log p_t(x)$) and can iteratively denoise random noise into realistic samples (Song & Ermon, 2019; Song et al., 2020), offer a promising approach for ensemble generation by training on established geological modeling outputs, such as those from FLUVSIM (Deutsch & Tran, 2002), to capture the complex spatial patterns and connectivity structures inherent in channelized systems. This capability addresses a need in ensemble-based data assimilation, where accurate covariance estimation benefits from much larger sample sizes than typically computationally feasible (Houtekamer & Mitchell, 2001; Rukhin, 2007). Once trained, diffusion models can rapidly generate large numbers of geologically consistent realizations on demand, providing super-ensembles for enhanced covariance estimation without requiring additional, expensive reservoir simulations. The generated realizations preserve essential geological features from the training data, including channel geometries and connectivity patterns that control fluid flow behavior in complex heterogeneous formations (Pyrcz & Deutsch, 2014). By leveraging these large synthetic ensembles within machine learning-enhanced localization frameworks (Lacerda et al., 2021; Silva et al., 2025a, 2025b), this approach can improve covariance estimation accuracy while maintaining the geological realism necessary for reliable uncertainty quantification in carbon storage applications.

This work places that idea in context. Data assimilation and machine learning have been connected in several communities—from early low-dimensional demonstrations (Brajard et al., 2020) to unified Bayesian views (Bocquet et al., 2020) and the broader "Data Learning" concept (Buizza et al., 2022). The evolution of DA-ML integration has progressed from early demonstrations on low-dimensional chaotic systems (Brajard et al., 2020) to frameworks that merge DA and ML through unified Bayesian formulations (Bocquet et al., 2020). The designers of these frameworks established that DA can provide complete state reconstructions from sparse observations, enabling ML models to learn dynamics that would otherwise be untrainable. The concept of "Data Learning"—combining ML's pattern recognition capabilities with DA's physics-based constraints—has emerged as a powerful paradigm (Buizza et al., 2022), with theoretical foundations showing how many DA-ML approaches can be unified under expectation-maximization frameworks.

In the subsurface domain, researchers explored two primary integration strategies: using ML to enhance DA processes and leveraging DA/UQ methods to improve ML models through physics-informed constraints (Cheng et al., 2023). For enhancing DA, ML is applied to learn model error corrections (Farchi et al., 2021b, 2021a), infer unresolved scale parametrizations directly from observations (Brajard et al., 2021), and enable training with partial observations through hybrid forecasting approaches (Wikner et al., 2021). Recent theoretical advances include Deep Data Assimilation (Arcucci et al., 2021), where neural networks learn to perform assimilation updates internally, embedding DA knowledge into the model itself. For geological carbon storage specifically, the integration of generative AI with DA represents a paradigm shift.

Recent work demonstrates that latent diffusion models can parameterize complex facies-based geomodels while maintaining geological realism during history matching (Di Federico & Durlofsky, 2024), working in latent space through VAE encoders. Score-based diffusion models have enabled real-time ensemble forecasts of 3D CO_2 plume evolution, generating hundreds of realizations in minutes rather than hours (Fan et al., 2025), although they require extensive training datasets. Beyond diffusion models, researchers employ conditional normalizing flows within Bayesian filtering frameworks to create uncertainty-aware "digital shadows" for continuous monitoring (Gahlot et al., 2025). They have also developed hierarchical DA approaches using sequential Monte Carlo methods to handle deep uncertainty in geostatistical parameters (Teng & Durlofsky, 2025). While these advances demonstrate that generative models can capture non-Gaussian uncertainties and maintain geological consistency, the challenge of adaptive localization that respects complex geological structures in ensemble-based methods remains an active area of research.

To establish the effectiveness of ML-enhanced localization, we first demonstrate its performance on a simpler test case with Gaussian-distributed permeability fields before progressing to the more complex channelized systems. This progression allows us to isolate the benefits of the ML approach from the additional complexities introduced by channelized geology. The validation case uses a smaller 32×32 grid with a Gaussian permeability distribution, providing a controlled environment to verify the variance preservation capabilities of our method.

To explore how machine learning can enhance data assimilation for GCS, we combine score-based diffusion models with machine learning-enhanced localization to overcome these limitations. Our investigation demonstrates that ML methods can improve data assimilation performance through an approach that operates on three levels:

1. **Efficient super-ensemble generation**: We employ score-based diffusion models trained on geostatistical realizations to enable the generation of thousands of permeability field samples. This overcomes the computational bottleneck of traditional methods by providing large ensembles (5000 members) for statistical estimation without running expensive forward simulations.
2. **Enhanced covariance estimation via ML proxies**: By training fast surrogate models on the working ensemble and applying them to large super-ensembles (5000 members) from the diffusion model, we improve covariance estimation accuracy without additional forward simulations.
3. **Adaptive localization respecting geology**: The ML-enhanced covariances capture how information propagates through specific geological structures, enabling localization that adapts to channel connectivity rather than imposing arbitrary geometric constraints.

We test our method on the methodology for 2D channelized systems with pressure observations. We do not address real-time monitoring, dynamic model updating, or non-channelized geological settings—each worthy of dedicated investigation, building upon this foundational work.

6.2. THEORETICAL BACKGROUND

This section presents the theoretical foundations of our integrated framework, covering score-based diffusion models, ensemble data assimilation with localization, and machine learning-enhanced covariance estimation.

6.2.1. SCORE-BASED DIFFUSION MODELS: MATHEMATICAL FOUNDATION

Score-based diffusion models add noise to data through a forward process and then learn to reverse it to generate new samples (Song & Ermon, 2019; Song et al., 2020). These methods evolved from foundational work on deep unsupervised learning using nonequilibrium thermodynamics (Sohl-Dickstein et al., 2015) and denoising diffusion probabilistic models (Ho et al., 2020), and have demonstrated better performance when compared to GANs and variational autoencoders for high-quality sample generation (Dhariwal & Nichol, 2021). Figure 6.1 illustrates the complete forward and reverse diffusion process, showing how channelized permeability fields are progressively transformed into noise and then reconstructed. The forward process transforms data samples into random noise through a continuous-time stochastic process defined by a stochastic differential equation (SDE):

$$\mathrm{d}\mathbf{x}_t = g(t)\mathrm{d}\mathbf{W}_t, \quad t \in [0, T] \tag{6.1}$$

where the initial condition, $\mathbf{x}_0 \in \mathbb{R}^D$, is a sample from the data distribution representing the permeability field in our context. The diffusion coefficient $g(t)$ controls the noise injection rate, $\mathbf{W}_t$ is the Wiener process (Brownian motion), and $t \in [0, T]$ is the time variable. As time progresses from $t = 0$ to $t = T$, $\mathbf{x}_t$ gradually transitions from a data distribution to a noise distribution.

In this work, we adopt the Score Matching with Langevin Dynamics (SMLD) variant of the Variance Exploding (VE) SDE formulation (Song & Ermon, 2019). In this framework, the forward diffusion is defined as $d\mathbf{x} = \sigma^t d\mathbf{W}$, where $\sigma = 25.0$ is the noise scale parameter. The diffusion coefficient is given by $g(t) = \sigma^t$, and the marginal probability standard deviation is:

$$\sigma_t = \sqrt{\frac{\sigma^{2t} - 1}{2 \ln \sigma}} \tag{6.2}$$

This formulation ensures that the variance of the noisy data grows exponentially with time, providing a transition from structured channelized patterns to isotropic Gaussian noise. The choice of $\sigma = 25.0$ was empirically determined to provide sufficient noise levels to destroy geological structures while maintaining numerical stability during training. The conditional distribution $p_t(\mathbf{x}_t|\mathbf{x}_0)$ follows a Gaussian distribution:

$$p_t(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0, \sigma_t^2 \mathbf{I}) \tag{6.3}$$

The key insight is that the time-reversed SDE can generate samples from the data distribution:

$$\mathrm{d}\mathbf{x}_t = -g^2(t) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) \mathrm{d}t + g(t) \mathrm{d}\bar{\mathbf{W}}_t, \quad t \in [T, 0] \tag{6.4}$$

where $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ is the score function—the gradient of the log-probability density. By solving this reverse SDE from $t = T$ to $t = 0$ with initial condition $\mathbf{x}_T \sim \mathcal{N}(0, \sigma_T^2 \mathbf{I})$, we obtain samples from the data distribution.

The score function is approximated using a neural network $s_\theta(\mathbf{x}_t, t) \approx \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$, trained using denoising score matching (Vincent, 2011):

$$\mathcal{L}(\theta) = \mathbb{E}_{t \sim \mathcal{U}(0,T)} \mathbb{E}_{\mathbf{x}_0 \sim p_{\text{data}}} \mathbb{E}_{\mathbf{x}_t \sim p_t(\mathbf{x}_t|\mathbf{x}_0)} \left\| s_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0) \right\|^2 \tag{6.5}$$

The trained diffusion model can also be adapted for posterior sampling to condition generated realizations on direct permeability observations. This approach follows recent advances in diffusion posterior sampling for general noisy inverse problems (Chung et al., 2022), which demonstrated how diffusion models can be adapted for Bayesian inference tasks. This capability, while not used in the main data assimilation workflow, provides an additional tool for integrating hard data constraints.

The reverse diffusion process in Equation 6.4 can be solved using two distinct approaches: stochastic (SDE) or deterministic (ODE) sampling. The probability flow ODE, which shares the same marginal distributions as the SDE, is given by:

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = -\frac{1}{2} g^2(t) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t), \quad t \in [T, 0] \tag{6.6}$$

We tested both SDE and ODE formulations for ensemble generation in data assimilation, and both approaches yield good results for generating geologically realistic channelized permeability fields. The choice between methods depends on the specific application requirements: SDE sampling provides more stochastic variability and exploration of the

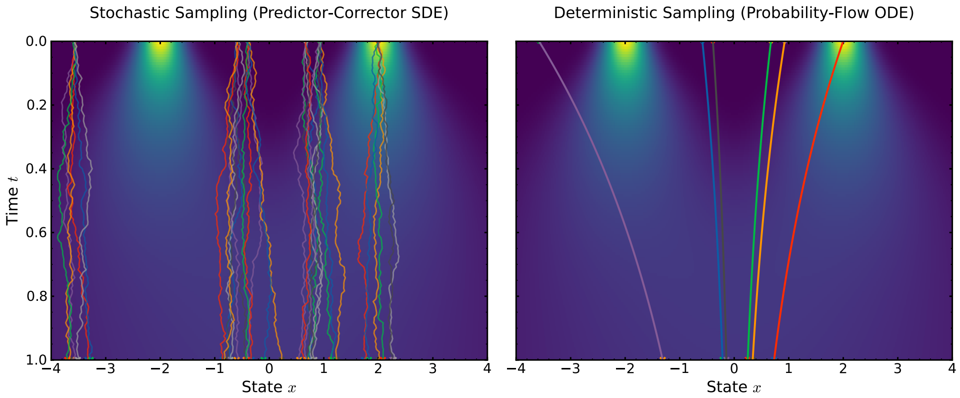


Figure 6.1: Score-based diffusion process. Forward diffusion (top) transforms channelized permeability to noise via VE-SDE. Reverse process (bottom) uses learned score function $s_\theta(\mathbf{x}_t, t)$ for denoising with SDE (stochastic) or ODE (deterministic) sampling.

learned distribution, while ODE sampling offers greater control and reproducibility. For our application in ensemble generation for data assimilation, we adopt the ODE formulation for several practical advantages. The deterministic nature ensures reproducibility. Given its deterministic nature and a fixed starting noise, the ODE always produces the same realization, enabling controlled experiments and sensitivity analyses. The ODE formulation typically requires fewer discretization steps than the SDE for comparable quality, improving computational efficiency. The smooth trajectories of the ODE solver also facilitate better preservation of geological structures during the denoising process. However, both the reverse-time SDE and the probability-flow ODE can generate large ensembles of geologically realistic samples suitable for enhanced covariance estimation in our setting, since they both are able to generate realistic samples.

6.2.2. ENSEMBLE-BASED DATA ASSIMILATION AND LOCALIZATION

ESMDA (Emerick & Reynolds, 2013b) provides a framework for integrating dynamic observations into reservoir models. Extending fundamental ensemble Kalman filter theory (Evensen, 1994, 2009) and the unified data assimilation framework (Evensen et al., 2022), the method performs multiple data assimilation steps with inflated observation error covariances, linearizing the nonlinear inverse problem through iteration.

For an ensemble of model parameters $\{\mathbf{z}_j\}_{j=1}^{N_e}$, where $\mathbf{z} \in \mathbb{R}^{N_z}$ represents the log-permeability field, the ESMDA update equation at iteration i is:

$$\mathbf{z}_j^a = \mathbf{z}_j^f + \mathbf{C}_{zd}^f \left(\mathbf{C}_{dd}^f + \alpha_i \mathbf{C}_{dd}\right)^{-1} \left(\mathbf{d}_{\text{obs}} + \sqrt{\alpha_i}\boldsymbol{\epsilon}_j - \mathbf{d}_j^f\right) \tag{6.7}$$

where superscripts f and a denote forecast (prior) and analysis (posterior) quantities respectively, $\mathbf{C}_{zd}^f$ is the cross-covariance between parameters and predicted data, $\mathbf{C}_{dd}^f$ is the auto-covariance of predicted data, $\mathbf{C}_{dd}$ is the observation error covariance matrix, α_i is the inflation coefficient for iteration i, $\mathbf{d}_{\text{obs}}$ represents the observed data, $\mathbf{d}_j^f = \mathbf{G}(\mathbf{z}_j^f)$ is the predicted data from forward model $\mathbf{G}$, and $\boldsymbol{\epsilon}_j \sim \mathcal{N}(0, \mathbf{C}_{dd})$ are random perturbations.

However, for practical ensemble sizes (typically $N_e \sim 100$), the sample covariance matrices contain large errors, leading to spurious correlations and subsequent ensemble collapse (Houtekamer & Mitchell, 2001). The industry standard solution to deal with these spurious correlations is distance-based localization, which modifies the Kalman gain through element-wise multiplication with a localization matrix $\mathbf{L}$. This matrix typically uses correlation functions such as the fifth-order piecewise rational function of Gaspari and Cohn (Gaspari & Cohn, 1999), which tapers correlations to zero beyond a specified distance:

$$\mathbf{z}_j^a = \mathbf{z}_j^f + \mathbf{L} \circ \left[\mathbf{C}_{zd}^f \left(\mathbf{C}_{dd}^f + \alpha_i \mathbf{C}_{dd} \right)^{-1} \right] \left(\mathbf{d}_{\text{obs}} + \sqrt{\alpha_i} \boldsymbol{\epsilon}_j - \mathbf{d}_j^f \right) \tag{6.8}$$

where $\circ$ denotes the Schur (element-wise) product.

6.2.3. MACHINE LEARNING-ENHANCED LOCALIZATION

Distance-based localization methods have become the industry standard for ensemble-based data assimilation, with implementations ranging from simple cutoff functions, like Gaspari-Cohn (Gaspari & Cohn, 1999), to adaptive correlation-based localization models (Vossepoel et al., 2025) and Schur product-based approaches (Houtekamer & Mitchell, 2001). These methods have been successfully deployed in operational weather forecasting, oceanography, and petroleum reservoir history matching for over two decades. However, while effective for many applications, they fail to capture the complex, non-local relationships between channelized permeability structures and pressure observations. In channelized systems, where pressure information travels along high-permeability pathways over long distances, distance-only tapering can be misleading, while correlation-based schemes can follow the geology, provided the estimated correlations are reliable. The challenge is that accurate correlation estimates typically require large ensembles, which are rarely affordable in practice. To address this, we use machine learning to augment classical localization: a fast proxy is trained on the working ensemble and then applied to a large super-ensemble generated by the diffusion model to produce more stable, correlation-aware covariance estimates even when the operational ensemble N_e is small (Lacerda et al., 2021; Silva et al., 2025a, 2025b).

Our ML-enhanced localization framework extends the approach proposed by (Silva et al., 2025a, 2025b) by training a fast proxy model $f^{ML} : \mathbf{z} \to \mathbf{d}$ using the working ensemble, then applying this proxy to a large super-ensemble generated by the diffusion model. The enhanced covariance estimate becomes:

$$\tilde{\mathbf{C}}_{zd}^{ML} = \frac{1}{N_s - 1} \sum_{k=1}^{N_s} (\mathbf{z}_k^s - \bar{\mathbf{z}}^s)(f^{ML}(\mathbf{z}_k^s) - \bar{\mathbf{d}}^{ML})^T \tag{6.9}$$

where $N_s \gg N_e$ is the super-ensemble size, and $\mathbf{z}_k^s$ are samples from the diffusion model. This approach increases the sample size for covariance estimation while maintaining computational feasibility.

Localization is effective when the taper $\mathbf{L}$ keeps parameter–data entries that represent true physical influence and zeros those that are spurious. With small ensembles, the sample cross–covariance $\mathbf{C}_{zd}$ is noisy: Monte Carlo error scales like $O(1/\sqrt{N_e})$, so many weak-but-real correlations are drowned out by sampling noise. Our approach trains a fast proxy $f^{ML}: \mathbf{z} \mapsto \mathbf{d}$ on the working ensemble and then evaluates it on a large super-ensemble ($N_s \gg N_e$) drawn from the diffusion model.

6.3. METHODS AND IMPLEMENTATION

This section presents the technical details of our integrated framework, including the channelized reservoir model, diffusion model implementation, ESMDA configuration, and machine learning-enhanced localization approach.

6.3.1. CHANNELIZED RESERVOIR GENERATION AND CO_2 INJECTION MODELING

We employ a 2D channelized reservoir model representative of fluvial depositional environments commonly encountered in CO_2 storage formations. The model domain consists of a 64×64 grid with 5-meter cell dimensions, resulting in a 320×320 meter area. Channelized permeability fields are generated using the FLUVSIM algorithm (Deutsch & Tran, 2002), which simulates fluvial channel systems through object-based modeling.

The FLUVSIM parameters are sampled from normal distributions $\mathcal{N}(\mu, \sigma)$ to ensure geological diversity across the training dataset. Table 6.1 summarizes the key parameters with their mean (μ) and standard deviation (σ) values.

Table 6.1: FLUVSIM parameters for channelized reservoir generation

Parameter	**Mean ± Std. Dev.**	**Description**
Channel orientation (degrees)	90 ± 30	Primary flow direction
Channel amplitude (m)	250 ± 10	Variable sinuosity
Channel wavelength (m)	2000 ± 50	Meandering patterns
Width-to-thickness ratio	50 ± 5	Channel aspect ratio
Channel proportion	0.40 ± 0.05	Net-to-gross ratio
Undulation wavelength (m)	250 ± 10	Thickness/width variations

This stochastic parameterization generates diverse channelized patterns while maintaining geological plausibility. The resulting permeability fields exhibit bimodal distributions with high-permeability channels (2000 mD) embedded in low-permeability background facies (50 mD), creating preferential flow paths characteristic of fluvial deposits. Figure 6.2 shows representative examples of the channelized permeability fields generated using FLUVSIM.

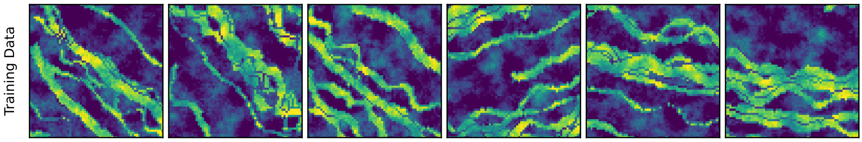


Figure 6.2: Representative FLUVSIM-generated channelized permeability fields with bimodal distribution: channels (2000 mD) in background (50 mD).

The CO_2 injection and storage simulations are performed using the Delft Advanced

Research Terra Simulator (DARTS), a compositional flow simulator specifically developed for subsurface flow applications (Voskov et al., 2024). DARTS uses operator-based linearization for efficient solution of complex multiphase flow problems (Khait & Voskov, 2017), and has been validated against benchmark cases including the FluidFlower experiment (Wapperom et al., 2024). The simulator has been successfully applied to optimize CO_2 injection using multi-scale reconstruction (Chen & Voskov, 2020) and to model enhanced CO_2 dissolution trapping in saline aquifers (Lyu & Voskov, 2023), demonstrating its capabilities for complex CO_2 storage scenarios.

The reservoir model in this study employs a structured Cartesian grid with dimensions of 64×64×1 cells, representing a two-dimensional horizontal layer. Each grid cell measures 5 meters in length and width, with a thickness of 10 meters, resulting in a computational domain of 320×320 meters horizontally. The simulated reservoir layer is situated at a depth of 2,000 meters, reflecting typical geological conditions for CO_2 storage, with an initial pressure of 150 bar and temperature of 50°C. The reservoir simulations involve a two-component, two-phase system comprising carbon dioxide (CO_2) and water (brine with 100,000 ppm salinity). The permeability field is generated using FLUVSIM as previously described, exhibiting a bimodal distribution with high-permeability channels (2000 mD) embedded in a low-permeability background (50 mD). The porosity remains constant throughout the reservoir at 0.3. CO_2 is injected through a single vertical well located centrally within the grid at a consistent rate of 1000 m^3/year over a two-year period. The injected fluid consists of nearly pure CO_2 into the saline aquifer. Pressure measurements are recorded monthly at four monitoring wells positioned in a five-spot pattern, each located at equal distances from the injection point in the cardinal directions. These observations provide spatial data points for subsequent data assimilation, with measurement errors of 1% reflecting realistic field monitoring uncertainties.

The relative permeability in the porous medium is modeled using the Brooks-Corey formulation (R. H. Brooks And A. T. Corey, 1964), with exponents $n_w = 2.0$ for water and $n_g = 1.5$ for CO_2. Residual saturations are set to $S_{wc} = 0.25$ and $S_{gc} = 0.1$, with end-point relative permeability values of $k_{rw} = k_{rg} = 1.0$. Capillary pressure is neglected due to its limited influence within the two-dimensional modeling framework. Diffusion and hysteresis effects, while potentially influential on CO_2 migration and trapping, are excluded to maintain consistency with the study's focus on large-scale reservoir dynamics. This modeling approach is considered sufficient to capture the key multiphase flow mechanisms governing CO_2 migration and entrapment in porous media. DARTS uses operator-based linearization Voskov, 2017 for efficient solution of the coupled multiphase flow and transport equations, accurately capturing the complex physics of CO_2 injection, including phase behavior and gravitational segregation. The simulator's computational efficiency enables the thousands of forward model evaluations required for ensemble-based data assimilation.

Figure 6.3 illustrates the channelized reservoir model and the corresponding pressure field simulated with DARTS. The left panel shows the log-permeability field with high-permeability channels, while the right panel displays the pressure distribution after injection. The injection well (triangle) and four monitoring wells (circles) are indicated. The pressure distribution correlates with the high-permeability channels and the injector–producer configuration: it is strongest along the dominant connected channel, with weaker propagation along the two minor channels.

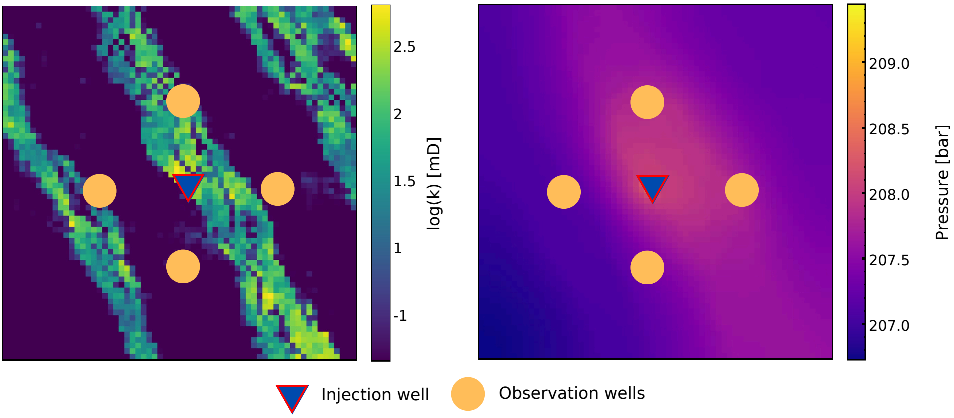


Figure 6.3: Channelized reservoir model (left) and DARTS-simulated pressure field (right) with injection well (triangle) and monitoring wells (circles). The pressure distribution correlates with the high-permeability channels and the injector–producer configuration: it is strongest along the dominant connected channel, with weaker propagation along the two minor channels.

6.3.2. DIFFUSION MODEL IMPLEMENTATION

The score-based diffusion model is implemented to learn the distribution of channelized permeability fields and generate new realizations that maintain geological realism.

NEURAL NETWORK ARCHITECTURE

We employ a hybrid UNet-Transformer architecture that combines convolutional neural networks with efficient attention mechanisms for learning the score function of channelized permeability distributions. The architecture is detailed in Figure 6.4. We use a U-Net with attention blocks to capture multi-scale features in permeability fields. The network begins with a time embedding module that uses Gaussian Fourier projections to encode the continuous diffusion time into high-dimensional representations. This temporal signal allows the network to adapt its predictions to the current noise level during the diffusion process.

The encoder path consists of four blocks that progressively downsample the input while expanding the feature channels. Each block uses `ConvNeXt`-style residual blocks (Liu et al., 2022): a 7×7 convolution to capture broader spatial context, followed by 3×3 convolutions and a residual connection; depthwise separable convolutions are not used unless explicitly stated. The time embeddings are injected at each resolution level, enabling the network to modulate its feature extraction based on the diffusion timestep.

At the bottleneck and intermediate resolutions, we incorporate Spatial Linformer attention modules (Wang et al., 2020a). Unlike standard self-attention, which scales quadratically with input size, Linformer achieves linear complexity through low-rank approximations. This builds on the transformer architecture's success in capturing long-range dependencies (Vaswani et al., 2017) while addressing computational constraints. This is valuable for geological applications where maintaining long-range spatial dependencies—such as channel connectivity across the domain—is important for realistic permeability field generation.

The decoder path mirrors the encoder with skip connections that preserve fine-scale details during upsampling. Through transposed convolutions and ConvNextBlock refinements, the network progressively reconstructs the score function at the original resolution. The final output represents $s_\theta(\mathbf{x}_t, t)$, the estimated score function that guides the reverse diffusion process.

TRAINING DATASET AND PROCEDURE

The training dataset comprises 3,242 channelized permeability realizations generated using sequential indicator simulation with systematically varied geostatistical parameters. This ensures a comprehensive coverage of plausible geological patterns, with each 64×64 field representing a different realization of channelized formations with varying orientations, widths, and connectivity patterns.

The model training employs standard practices for deep generative models. We use the Adam optimizer with a cosine annealing learning rate schedule that gradually reduces from 10^{-4} to 10^{-6}, promoting stable convergence. The training process minimizes the denoising score matching objective shown in Algorithm 5, which teaches the network to predict the gradient of the log-density at various noise levels.

The model converged in 525 epochs. The final loss of 308.76 represents an 83.6% reduction from the initial value of 1,882.96. The quick convergence shows that channelized patterns are well-structured and the UNet-Transformer architecture captures these geological features effectively.

Algorithm 5: Training Score-Based Diffusion Models

```
Input  : Dataset $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$, neural network $s_\theta$, noise schedule $\sigma(t)$
Output : Trained score network $s_\theta$
for $epoch = 1$ to $N_{epochs}$ do
    for batch $\mathcal{B} \subset \mathcal{D}$ do
        Sample $t \sim \mathcal{U}(0, T)$ uniformly;
        Sample $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$;
        Compute $\mathbf{x}_t = \mathbf{x}_0 + \sigma_t \boldsymbol{\epsilon}$;
        Compute loss: $\mathcal{L} = \|\boldsymbol{\epsilon} + \sigma_t s_\theta(\mathbf{x}_t, t)\|^2$;
        Update $\theta$ using gradient descent on $\mathcal{L}$;
    end for
end for
return trained $s_\theta$;
```

SAMPLING STRATEGY

The trained diffusion model supports two complementary sampling approaches for generating permeability fields. The Predictor-Corrector (PC) sampler combines Euler-

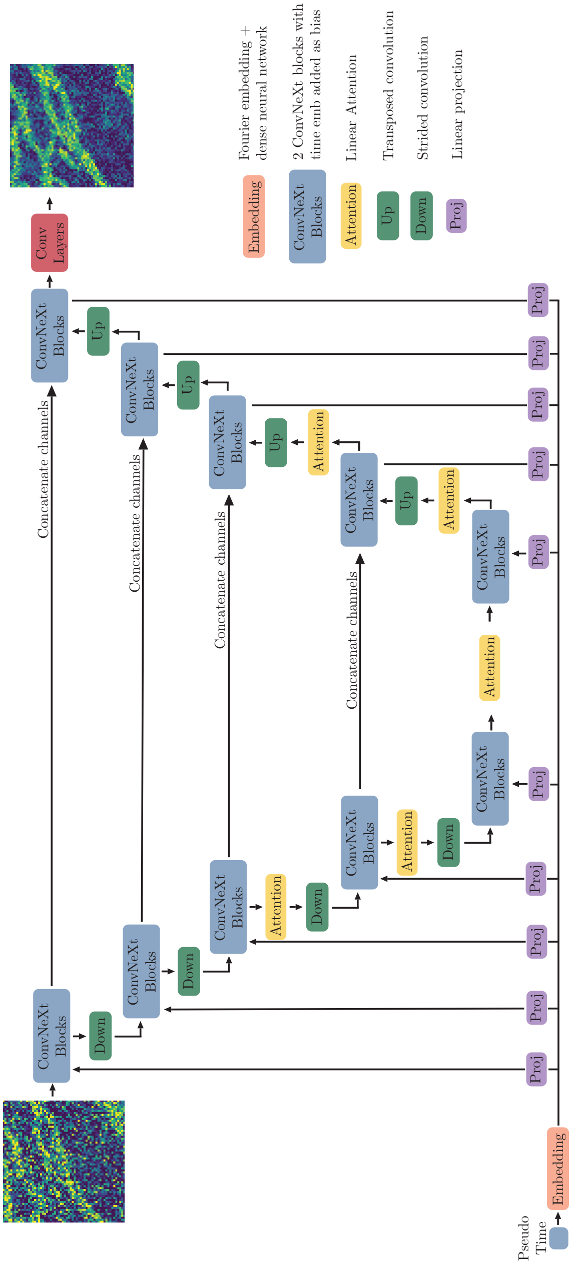


Figure 6.4: Visualization of UNet architecture with attention.

Maruyama steps for the reverse SDE with Langevin MCMC corrections, producing high-quality samples with stochastic diversity. This approach excels in exploring the full distribution of channelized patterns learned during training.

For our ESMDA application, however, we primarily employ the probability flow ODE formulation:

$$\frac{d\mathbf{x}}{dt} = -\frac{1}{2}g^2(t)s_\theta(\mathbf{x}_t, t) \tag{6.10}$$

Deterministic sampling helps because we can reproduce results exactly for ensemble-based data assimilation. Most importantly, it ensures reproducibility, which is essential for scientific validation and debugging. Given the same initial noise vector, the ODE always produces identical realizations. This property is used for debugging, validation, and ensuring consistent results across different computational environments. The ODE solver employs the midpoint method with 1000 integration steps, balancing numerical accuracy with computational efficiency.

Generated samples are not modified beyond reversing the preprocessing used for training. Specifically, inputs were standardized (per-pixel z-scoring of $\log k$) to ease optimization; at generation time we denormalize the network outputs using the training statistics and invert the log transform to recover permeability in millidarcies. We then compute and report simple diagnostics—channel proportion (0.40 ± 0.05) and value-range checks ($k \in [10, 10^4]$ mD)—but we do not filter or discard samples based on these checks.

Figure 6.5 illustrates the generation process, showing how both sampling methods progressively transform Gaussian noise into realistic channelized permeability fields while preserving distinct geological features.

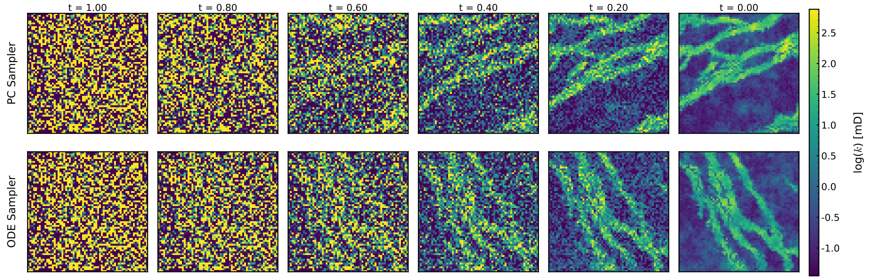


Figure 6.5: Evolution of channelized structures during generation. ODE (top) and PC (bottom) samplers progressively denoise from Gaussian to realistic permeability fields.

6.3.3. ESMDA Configuration

The data assimilation employs ESMDA with four iterations ($N_a = 4$), using uniform inflation coefficients of $\alpha_i = 4$ for all iterations. This choice balances computational efficiency with adequate linearization of the nonlinear inverse problem. The observation error is set to 1% of the true values, consistent with typical pressure measurement uncertainties. To maintain physical plausibility, parameter bounds constrain the log-permeability values to the range $\log_{10}(k) \in [1, 4]$, corresponding to permeability values between 10 and 10,000 mD. We evaluate the framework performance across multiple ensemble sizes (50, 100, 200, 500, and 1000 members) to understand scaling behavior and practical requirements for operational deployment.

With the inflation schedule $\{\alpha_i\}$ and observation–error covariance $\mathbf{C}_{dd}$ defined above, the localized ESMDA analysis step at iteration i reads:

$$\mathbf{z}_j^a = \mathbf{z}_j^f + \mathbf{L} \circ \left[\mathbf{C}_{zd}^f \left(\mathbf{C}_{dd}^f + \alpha_i \mathbf{C}_{dd} \right)^{-1} \left(\mathbf{d}_j - \mathbf{G}(\mathbf{z}_j^f) \right) \right] \tag{6.11}$$

where $\circ$ denotes the Schur (element-wise) product, $\mathbf{L}$ is the localization matrix, and $\mathbf{d}_j = \mathbf{d}_{\text{obs}} + \sqrt{\alpha_i}\,\boldsymbol{\epsilon}_j$ with $\boldsymbol{\epsilon}_j \sim \mathcal{N}(0, \mathbf{C}_{dd})$.

6.3.4. VALIDATION ON GAUSSIAN PERMEABILITY FIELDS

Before moving to channelized systems, we validate ML-enhanced localization on a simpler Gaussian permeability case. This shows the method's consistency across geology and isolates variance preservation benefits from channel effects.

The validation test case corresponds to a small reservoir model undergoing CO_2 injection, representing a CCS scenario similar to that presented in (Silva et al., 2025a). Simulations were performed using the same open-source Delft Advanced Research Terra Simulator (DARTS) used for the main channelized study. The reservoir model uses a structured Cartesian grid with dimensions of 32×32×1 cells, representing a two-dimensional horizontal layer. Each gridblock is uniformly sized at 192 meters in length and width and 10 meters in thickness. The simulated layer is located at a depth of 2,000 meters, consistent with typical geological settings for CO_2 storage in a saline aquifer, and is initially at a pressure of 200 bar.

The reservoir simulation considered a two-component, two-phase system comprising CO_2 and water. The prior log-permeability fields were generated from a Gaussian distribution with a mean of 3.0 and a variance of 1.0. The spatial correlation was modeled using a Gaspari-Cohn function with a correlation range corresponding to the size of 10 grid cells. Realizations were generated from the resulting covariance matrix via Cholesky decomposition. Porosity was assumed to be constant throughout the reservoir at 25%.

The data assimilation experiments used ensembles of realizations of log-permeability (a total of 1,024 model parameters), generated using the same prior settings as those used to construct the ground truth. The observations consisted of pressure measurements at four monitoring well locations over a two-year period, resulting in a total of 96 data points. We perform data assimilation using ESMDA with four iterations under four cases: no localization, distance-based localization (DB-localization), pseudo-optimal localization (PO-localization), and ML-localization. For the distance-based localization case, we used the standard Gaspari-Cohn function with a critical length corresponding to the size of 10 gridblocks. For ML-localization, we used the XGBoost method, consistent with our main study. Here, pseudo-optimal localization (PO) denotes a taper derived by minimizing the expected Frobenius-norm error between the true and localized covariance matrices, following the formulation of (Rukhin, 2007). This yields data-driven coefficients that balance removal of spurious correlations against preservation of true correlations.

6.3.5. MACHINE LEARNING-ENHANCED LOCALIZATION IMPLEMENTATION

Our ML localization works by training a fast proxy model on the working ensemble to improve covariance estimation. Adapting the machine learning-based distance-free localization introduced in recent studies (Silva et al., 2025a, 2025b), we tailor these methods for channelized reservoir characterization. Given ensemble pairs $\{(\mathbf{z}_j, \mathbf{d}_j)\}_{j=1}^{N_e}$ from DARTS simulations, we train:

$$f^{ML} : \mathbf{z} \rightarrow \mathbf{d} \tag{6.12}$$

where $\mathbf{z} \in \mathbb{R}^{4096}$ is the flattened log-permeability field and $\mathbf{d} \in \mathbb{R}^{96}$ contains pressure observations.

ENHANCED COVARIANCE ESTIMATION

The enhanced covariance estimation uses a super-ensemble of $N_s = 5000$ realizations:

$$\tilde{\mathbf{C}}_{zd}^{ML} = \frac{1}{N_s - 1}(\mathbf{Z}_s - \bar{\mathbf{Z}}_s)(\mathbf{D}_s^{ML} - \bar{\mathbf{D}}_s^{ML})^T \tag{6.13}$$

where $\mathbf{D}_s^{ML} = [f^{ML}(\mathbf{z}_1^s), \ldots, f^{ML}(\mathbf{z}_{N_s}^s)]$ contains ML predictions.

For the ML-enhanced localization, we compute localization coefficients using a modified pseudo-optimal approach. When model parameters lack direct spatial relationships with observations, conventional distance-based schemes become inadequate. We therefore employ an adaptation of the pseudo-optimal taper function introduced by (Rukhin, 2007), selected for its rigorous theoretical foundation and computational efficiency.

The pseudo-optimal localization approach seeks to find the optimal balance between preserving true correlations and removing spurious ones by minimizing the total squared error between the true and localized covariance matrices (measured using the Frobenius norm, which sums the squares of all matrix elements). For the (i, j)-th element of the localization matrix $\mathbf{L}$, the formula is:

$$\mathbf{L}_{ij}^{ML} = \frac{(\tilde{c}_{ij}^{ML})^2}{(\tilde{c}_{ij}^{ML})^2 + \frac{(\tilde{c}_{ij}^{ML})^2 + \tilde{c}_{ii}^{ML}\tilde{c}_{jj}^{ML}}{N_e}} \tag{6.14}$$

where $\tilde{c}_{ij}^{ML}$ represents the ML-enhanced covariance between the i-th model parameter and j-th predicted data point, computed from the super-ensemble. Since the true covariance is unknown, we use the ML-enhanced estimate $\tilde{c}_{ij}^{ML}$ from our large super-ensemble. To suppress spurious correlations, we apply a thresholding rule: $\mathbf{L}_{ij}^{ML} = 0$ if $|\tilde{c}_{ij}^{ML}| < \eta\sqrt{\tilde{c}_{ii}^{ML}\tilde{c}_{jj}^{ML}}$, with $\eta = 10^{-3}$ chosen to minimize over-aggressive correlation removal while maintaining effective spurious correlation suppression.

MACHINE LEARNING MODELS FOR STATE ESTIMATION

The relationship between permeability fields and pressure responses in reservoir simulation naturally forms a tabular dataset, where each grid cell's permeability serves as a feature and well pressures constitute the targets. This structure, combined with recent

Algorithm 6: ML-Enhanced Localization

```
Input  : Ensemble $\{\mathbf{z}_j, \mathbf{d}_j\}_{j=1}^{N_e}$, ML model type, super-ensemble size $N_s$
Output : Localization matrix $\mathbf{L}^{ML}$
// Step 1:  Train ML proxy
Train $f^{ML}$ using $\{(\mathbf{z}_j, \mathbf{d}_j)\}_{j=1}^{N_e}$ with 80/20 train-validation split;
// Step 2:  Generate super-ensemble
for $k = 1$ to $N_s$ do
    Sample $\mathbf{z}_k^s$ from prior (diffusion model);
    Predict $\mathbf{d}_k^{ML} = f^{ML}(\mathbf{z}_k^s)$;
end for
// Step 3:  Compute enhanced covariance
Calculate $\tilde{\mathbf{C}}_{zd}^{ML}$ using Equation 6.13;
// Step 4:  Calculate localization
for each parameter $i$ and observation $j$ do
    Compute $\mathbf{L}_{ij}^{ML}$ using Equation 6.14;
end for
return $\mathbf{L}^{ML}$;
```

evidence that classical machine learning methods can outperform deep learning on tabular data (Borisov et al., 2024; Gorishniy et al., 2021; Grinsztajn et al., 2022; Shwartz-Ziv & Armon, 2022), guides our model selection. We used three different ML models to approximate the nonlinear forward model. Linear regression provides a computationally efficient baseline, establishing whether nonlinearity is essential for accurate predictions. Despite its simplicity, linear models struggle with the complex multiphase flow physics inherent in CO_2 injection scenarios.

Random forests (Breiman, 2001) address these limitations through ensemble decision trees trained on bootstrap samples. The method naturally handles the sparse influence patterns characteristic of reservoir flow—distant grid cells often have negligible effect on local pressure observations. By averaging predictions across multiple trees and using random feature selection at each split, random forests capture nonlinear relationships while maintaining robustness against overfitting. XGBoost (Chen & Guestrin, 2016) extends the ensemble approach through gradient boosting, where each successive tree corrects errors from previous iterations.

The choice of tree-based methods for our application comes from two data characteristics identified by Grinsztajn et al. (2022). First, many grid cells represent uninformative features for specific pressure observations due to flow barriers and distance effects. Second, pressure responses can exhibit non-smooth behavior with sharp transitions at permeability contrasts and channel boundaries. These properties explain the performance gap observed in our experiments, where tree-based methods achieve prediction errors 10-15 times lower than linear regression while preserving geological patterns in the localization coefficients. All models employ default hyperparameters from their respective implementations (scikit-learn (Pedregosa et al., 2011) and XGBoost library (Chen &

Guestrin, 2016)), reflecting operational constraints and ensuring reproducibility.

6.3.6. INTEGRATED WORKFLOW

The workflow consists of two phases: (1) initial ensemble generation using geostatistical models, and (2) iterative ESMDA with ML-enhanced localization. Algorithm 7 details the complete process.

Algorithm 7: Integrated Diffusion-ML-ESMDA Workflow

```
Input  : Ensemble size $N_e$, observations $\mathbf{d}_{\text{obs}}$, trained diffusion model, ML model type
Output : Posterior ensemble $\{\mathbf{z}_j^a\}_{j=1}^{N_e}$
// Phase 1:  Initial Ensemble Generation
Generate initial ensemble $\{\mathbf{z}_j^0\}_{j=1}^{N_e}$ using geostatistical models (e.g., FLUVSIM);
// Phase 2:  ESMDA with ML-Enhanced Localization
for $i = 1$ to $N_a$ iterations do
    // Forward simulations
    for $j = 1$ to $N_e$ do
        Run DARTS: $\mathbf{d}_j = \mathbf{G}(\mathbf{z}_j^{i-1})$;
    end for
    // ML proxy training
    Train ML model: $f^{ML} = \text{fit}(\{\mathbf{z}_j^{i-1}\}, \{\mathbf{d}_j\})$;
    // Super-ensemble generation
    Generate $\{\mathbf{z}_k\}_{k=1}^{N_s}$ using diffusion model;
    Predict: $\mathbf{d}_k^{ML} = f^{ML}(\mathbf{z}_k)$ for $k = 1, \dots, N_s$;
    // Enhanced localization
    Compute $\tilde{\mathbf{C}}_{zd}^{ML}$ using super-ensemble;
    Calculate localization matrix $\mathbf{L}^{ML}$ (Algorithm 6);
    // ESMDA update
    for $j = 1$ to $N_e$ do
        Perturb observations: $\mathbf{d}_j^{\text{pert}} = \mathbf{d}_{\text{obs}} + \sqrt{\alpha_i}\boldsymbol{\epsilon}_j$;
        Apply localized update to get $\mathbf{z}_j^i$;
    end for
end for
return Posterior ensemble $\{\mathbf{z}_j^{N_a}\}_{j=1}^{N_e}$;
```

The framework's modular design allows flexibility in component selection. Different ML models can be employed based on problem characteristics—Random Forest for capturing local effects, XGBoost for handling complex nonlinearities, or Linear Regression when computational resources are extremely limited. Similarly, the diffusion model architecture can be adapted to different geological environments by training on appropriate datasets, whether channelized, fractured, or layered systems. This combination addresses important problems in geological carbon storage characterization: maintaining geological realism throughout the assimilation process, preserving uncertainty estimates needed for risk assessment, and achieving computational efficiency suitable for operational deployment. The framework provides a method for combining machine learning with physics-based modeling, ensuring that each component enhances rather than compromises the others.

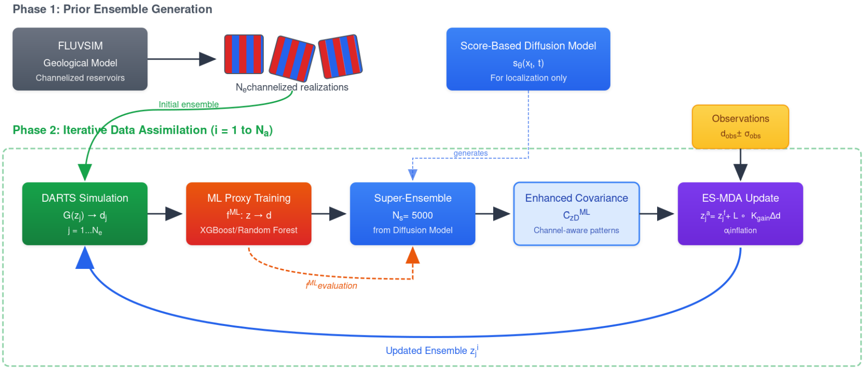


Figure 6.6: Integrated workflow combining score-based diffusion models, ML-based localization, and ESMDA. Phase 1: prior ensemble generation via geostatistical models. Phase 2: iterative data assimilation with ML-enhanced localization.

6.4. RESULTS AND ANALYSIS

We evaluate the framework for uncertainty quantification in GCS. We assess diffusion model performance, analyze ML proxy models for localization, and evaluate assimilation results across ensemble sizes. The focus is how well the method captures channelized structures and improves pressure predictions through ML-enhanced localization. To quantify variance preservation, we report the normalized variance (NV), defined as the ratio between the posterior and prior ensemble variances of log–permeability: $\mathrm{NV} = \mathrm{tr}(\mathbf{C}_{\mathrm{post}})/\mathrm{tr}(\mathbf{C}_{\mathrm{prior}})$, where $\mathbf{C}_{\mathrm{post}}$ and $\mathbf{C}_{\mathrm{prior}}$ are the ensemble covariance matrices over all grid cells. An NV of 1 indicates no variance loss, while smaller values indicate stronger variance collapse.

6.4.1. GENERATING GEOLOGICAL MODELS WITH DIFFUSION

The diffusion model converged rapidly during training on 3,242 channelized permeability realizations. The loss decreased from an initial high value to a stable level after about 500 epochs, leading to early termination well before the planned 5,000 epochs. The checkpoint near convergence was used for subsequent experiments.

Training dynamics showed exponential decay in the raw training loss during initial epochs, with excellent agreement between training and validation losses, indicating no overfitting. The learning rate schedule with cosine annealing contributed to stable convergence, and the smoothed loss curve confirmed the absence of oscillations.

Figure 6.7 demonstrates the quality of generated permeability fields compared to the original FLUVSIM training data. The comparison shows realizations generated using both ODE and SDE sampling approaches alongside true training samples. Both sampling approaches produce geologically plausible variations that respect the learned distribution while avoiding direct replication of training samples. This demonstrates that the diffusion model has successfully learned the underlying geological patterns rather than simply memorizing the training data, enabling generation of novel realizations suitable for uncertainty quantification in data assimilation.

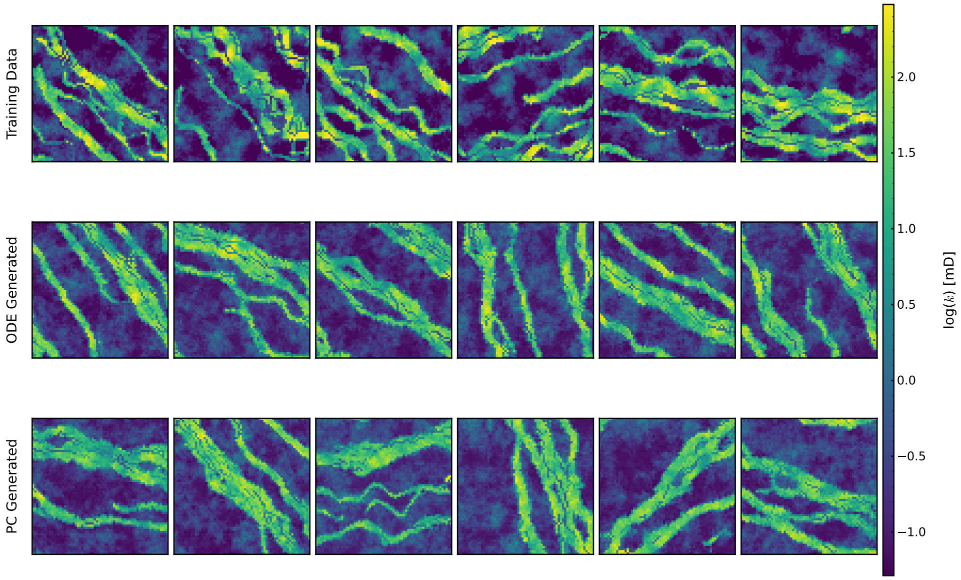


Figure 6.7: Visual comparison of training data (top row) with generated realizations using ODE sampling (middle row) and SDE sampling (bottom row).

6.4.2. VALIDATION RESULTS: GAUSSIAN PERMEABILITY CASE

Figure 6.8 shows the reference (ground truth) permeability field for the validation case, highlighting the location of a single CO_2 injection well at the center of the model and four pressure monitoring wells. Unlike the channelized systems in our main study, this field exhibits smooth Gaussian spatial correlations.

Figure 6.9 shows the normalized variance (NV) values obtained after data assimilation with different ensemble sizes. ML-localization consistently achieved the highest NV across all ensemble sizes, with improvements ranging from 15% to 35% compared to no localization. This demonstrates that the variance preservation benefits of ML-enhanced localization extend beyond channelized systems to general heterogeneous permeability fields.

In terms of data match quality, all methods resulted in similar values of data mismatch. This is illustrated in Figure 6.10, which shows the predicted well bottom-hole pressure for all localization cases with an ensemble of 100 realizations. The comparable data matching quality across methods confirms that improved variance preservation does not come at

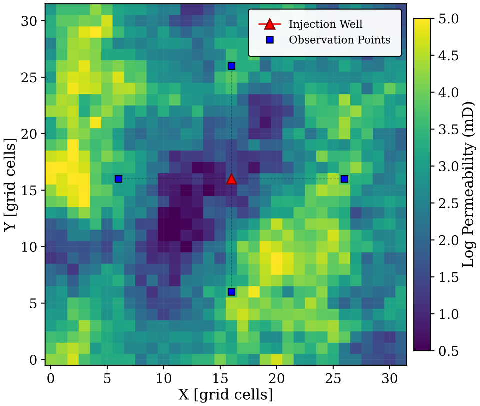


Figure 6.8: Reference permeability field for the CCS validation model. The red star indicates the CO_2 injection well, and the blue squares denote the positions of the four pressure monitoring wells.

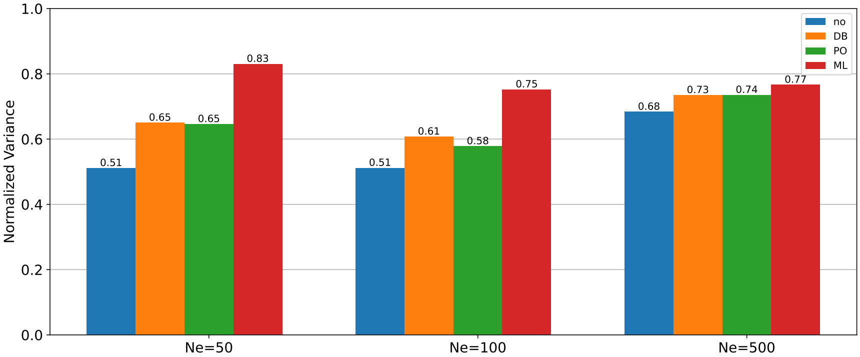


Figure 6.9: Normalized variance results with varying ensemble sizes for the CCS validation model. This figure includes the results of data assimilation without localization (no), DB-localization (DB), PO-localization (PO), and ML-localization (ML).

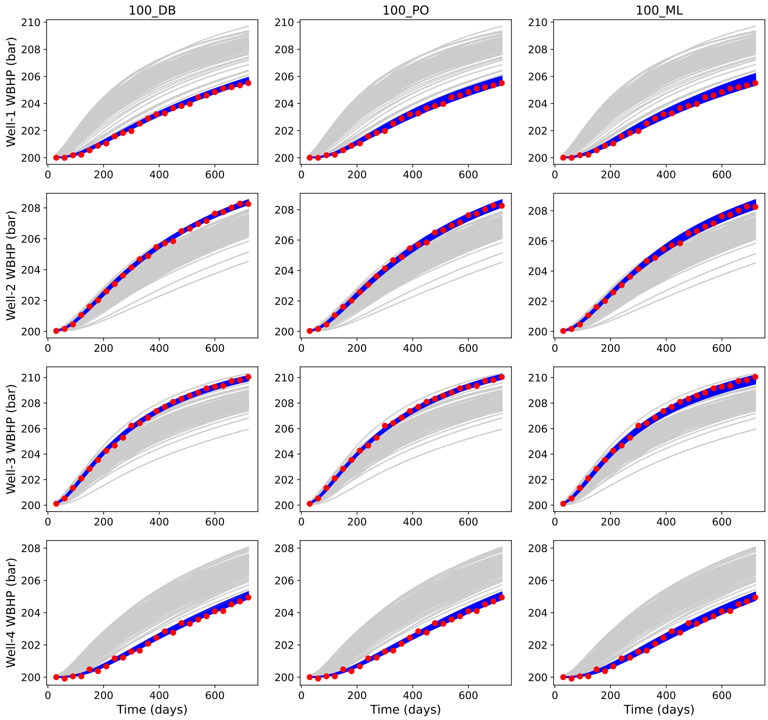


Figure 6.10: Predicted WBHP data before and after data assimilation for an ensemble size of $N_e = 100$. Each row represents a monitoring well and each column a different localization scheme. The red dots represent the observed data, the gray curves the prior ensemble, and the blue curves the posterior ensemble.

the cost of history matching accuracy.

Figure 6.11 reveals the fundamental differences in how various localization methods distribute parameter updates. Distance-based localization produces smooth, radially symmetric patterns centered on each monitoring well, with values decaying from one at the data location to zero at distant regions. Pseudo-optimal localization emphasizes the regions between injection and monitoring wells but also shows high variability in distant areas, likely due to spurious correlations from limited ensemble sizes.

ML-localization demonstrates superior performance by effectively suppressing distant correlations while maintaining strong updates along the flow paths connecting injection and monitoring wells. The localization patterns are neither constrained to be centered on the data location (as in DB-localization) nor do they exhibit spurious correlations in distant regions (as in PO-localization). ML-localization exhibits much sharper decay in localization values, concentrating updates where they are most geologically meaningful.

These validation results confirm that ML-enhanced localization provides consistent benefits across different permeability distributions, setting the stage for its application to the more complex channelized systems presented in the following sections.

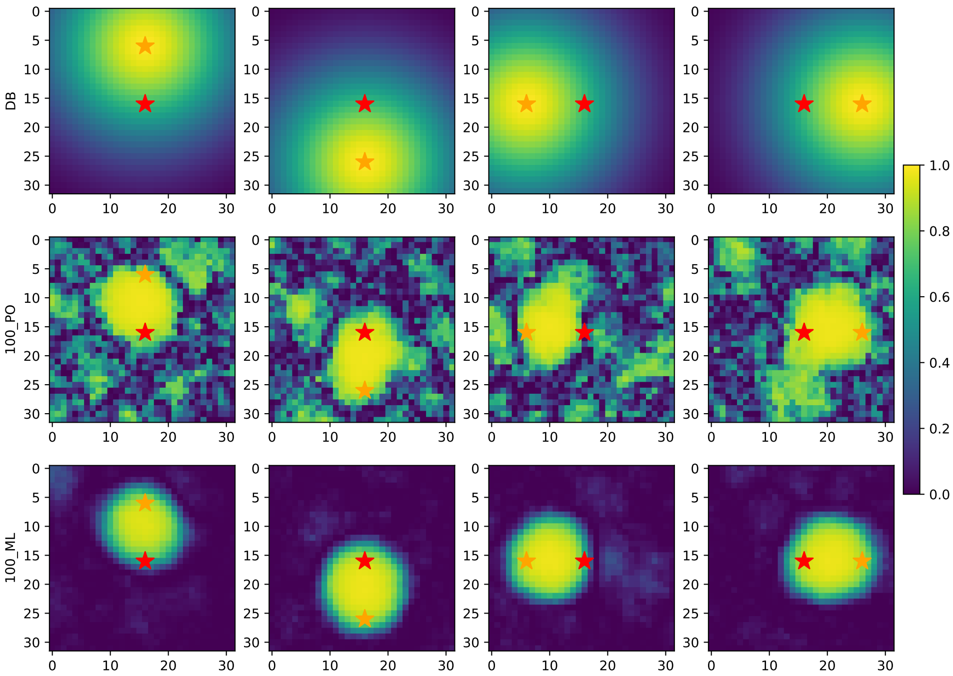


Figure 6.11: Localization values ($N_e = 100$) for the CCS validation model. Each column corresponds to a different monitoring well, and each row shows results from a different localization scheme. The red star indicates the CO_2 injection well, and the orange stars denote the positions of the monitoring wells.

6.4.3. MACHINE LEARNING PROXY FOR LOCALIZATION IN CHANNELIZED SYSTEMS

We evaluate three machine learning algorithms for predicting localization coefficients: linear regression, random forest, and XGBoost. Table 6.2 summarizes their performance across different ensemble sizes, revealing distinct patterns that inform our framework design.

Table 6.2: Machine Learning Model Performance for Localization Coefficient Prediction

(a) Root Mean Square Error (RMSE)

Model	**Number of Ensemble Members** (N_e)				
	50	100	200	500	1000
Linear Regression	2.351	2.828	0.681	1.262	0.268
Random Forest	0.254	0.175	0.191	0.177	0.159
XGBoost	0.310	0.198	0.216	0.188	0.159

(b) Training Time (seconds)

Model	**Number of Ensemble Members** (N_e)				
	50	100	200	500	1000
Linear Regression	0.016	0.027	0.067	0.141	0.40
Random Forest	3.528	6.238	15.678	38.890	106.8
XGBoost	1.301	1.849	4.372	15.767	70.9

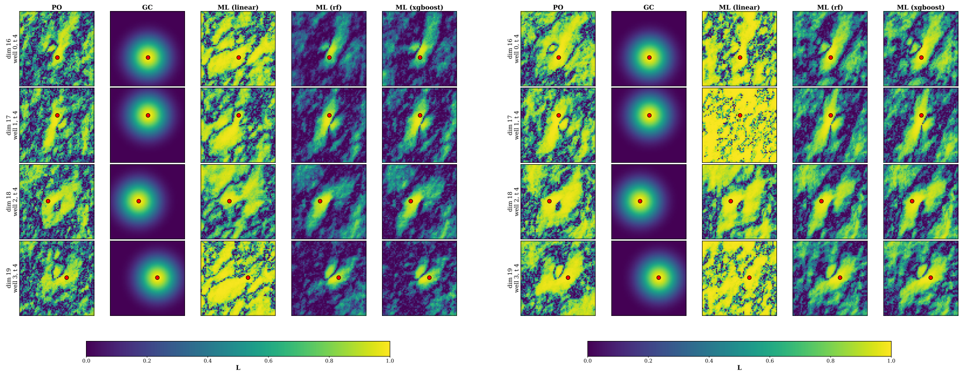
Tree-based methods (Random Forest and XGBoost) consistently generate better results than linear regression, achieving RMSE values below 0.31 across all ensemble sizes. This performance confirms the highly nonlinear relationship between permeability structure and pressure response in channelized systems—a relationship that linear models fail to capture, as evidenced by their erratic RMSE values ranging from 0.268 to 2.828.

While Random Forest achieves the lowest prediction errors for most ensemble sizes, computational cost analysis reveals important practical considerations. Training time for Random Forest increases with ensemble size, reaching 106.8 seconds for $N_e = 1000$. XGBoost offers an attractive alternative, providing comparable accuracy (RMSE within 0.02 of Random Forest) with 33% lower computational cost for large ensembles.

6.4.4. DATA ASSIMILATION PERFORMANCE IN CHANNELIZED SYSTEMS

ML localization's main benefit shows up when examining the spatial patterns of parameter updates. Figure 6.12 illustrates representative localization patterns for medium ($N_e = 200$) and large ($N_e = 500$) ensemble sizes, demonstrating how ML-enhanced methods adapt to the channelized reservoir structure. A comprehensive comparison across all ensemble sizes and methods is presented in Figure 6.13.

Traditional distance-based localization (Gaspari-Cohn) produces circular update regions centered on observation wells, indiscriminately modifying both channel and background permeabilities. This approach fails to respect the fundamental geological structure where pressure information propagates preferentially along high-permeability



(a) $N_e = 200$

(b) $N_e = 500$

Figure 6.12: Representative localization patterns for medium and large ensemble sizes showing channel-aligned parameter updates for different methods. The data point location is marked with a red circle. Unlike traditional distance-based methods (GC) that produce circular patterns, ML-enhanced localization (ML methods) adapts to the channelized permeability structure, concentrating updates along high-permeability pathways.

channels. In contrast, ML-enhanced methods generate updates that follow channel geometries, concentrating modifications where they are geologically meaningful. This channel-aligned updating is maintained consistently across all ensemble sizes, demonstrating the robustness of the ML approach.

The impact of these different localization strategies on the geological realism of the posterior permeability fields is illustrated in Figure 6.14. The figure displays a representative posterior permeability sample for different ensemble sizes and localization methods. For small ensemble sizes ($N_e = 50$), the absence of localization or the use of pseudo-optimal localization leads to severely blurred permeability fields. The updates affect the background facies indiscriminately, and even the channel structures lose their sharpness. In contrast, the ML-based localization methods (Random Forest and XGBoost) preserve the channel connectivity and overall geological structure remarkably well, even with a small ensemble. The updates exhibit a channel-like form, which is a direct consequence of the geologically-aware localization patterns shown previously. The Gaspari-Cohn localization produces updates that are distinctly circular, reflecting its distance-based nature and leading to less geologically consistent modifications. As the ensemble size increases, the benefits of ML-based localization become less pronounced, as expected. However, even for larger ensembles ($N_e = 500, 1000$), the permeability fields generated with ML-localization appear geologically sharper and more realistic compared to other methods. This demonstrates that even with larger ensembles, ML-enhanced localization provides an advantage in preserving the fine-scale geological details that are critical for accurate flow prediction.

To visualize where and how strongly each method actually modifies the model, Fig. 6.15 shows the absolute change in log–permeability (posterior – prior) for the same realization shown in Fig. 6.14. The machine-learning localizations concentrate updates along the high-permeability channel that connects to the observation well (green marker), whereas the un-localised cases disperse changes into the background. This effect is

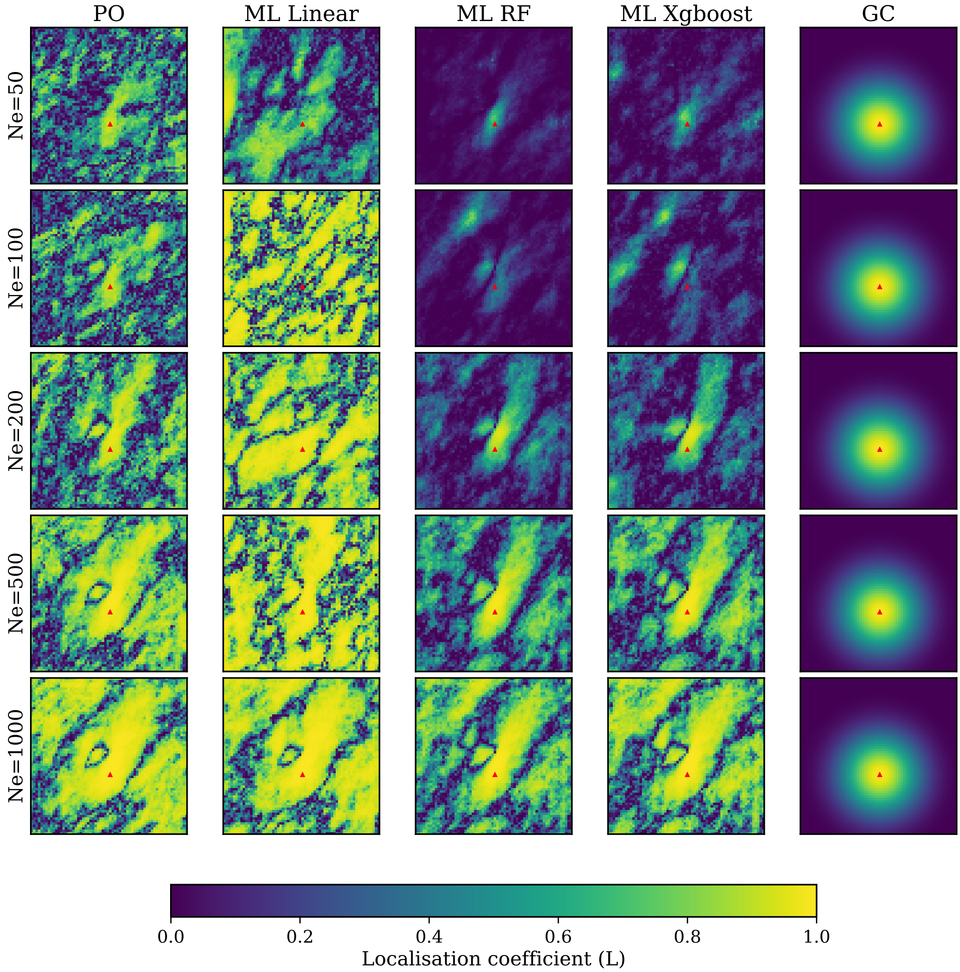


Figure 6.13: Localization patterns for different ensemble sizes (N_e = 50-1000) and methods. ML methods produce channel-aligned patterns; traditional methods show circular patterns.

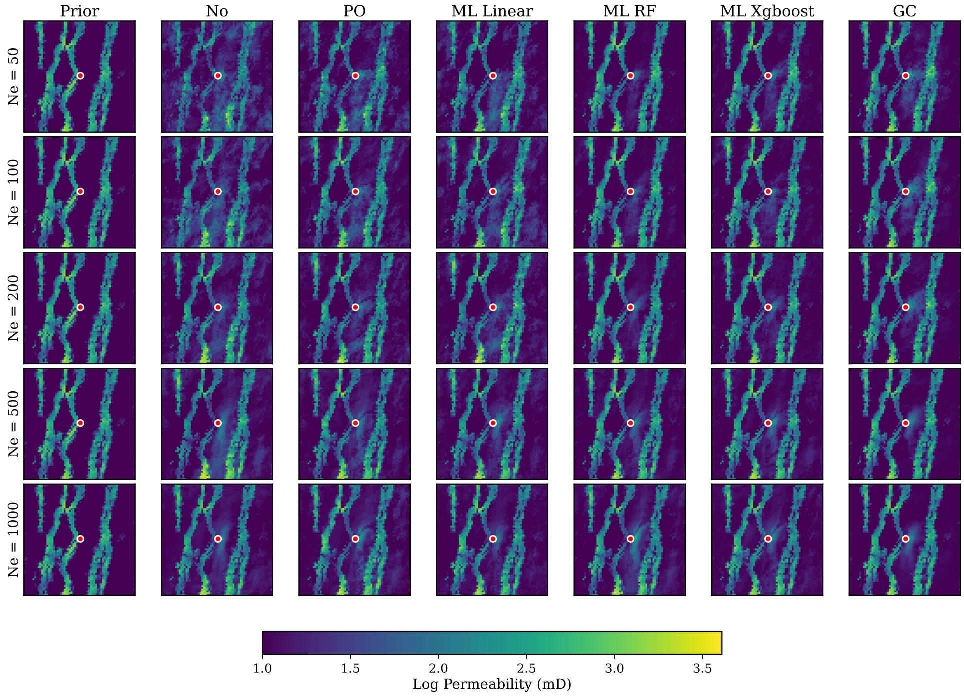


Figure 6.14: Sample of posterior permeability fields for different ensemble sizes (N_e) and localization methods. ML-based methods preserve channel structures better than other methods, especially for small ensembles.

larger for N_e=50 and 100 members. This update pattern illustrates the ML approaches preserving channel connectivity and variance more effectively than the alternatives.

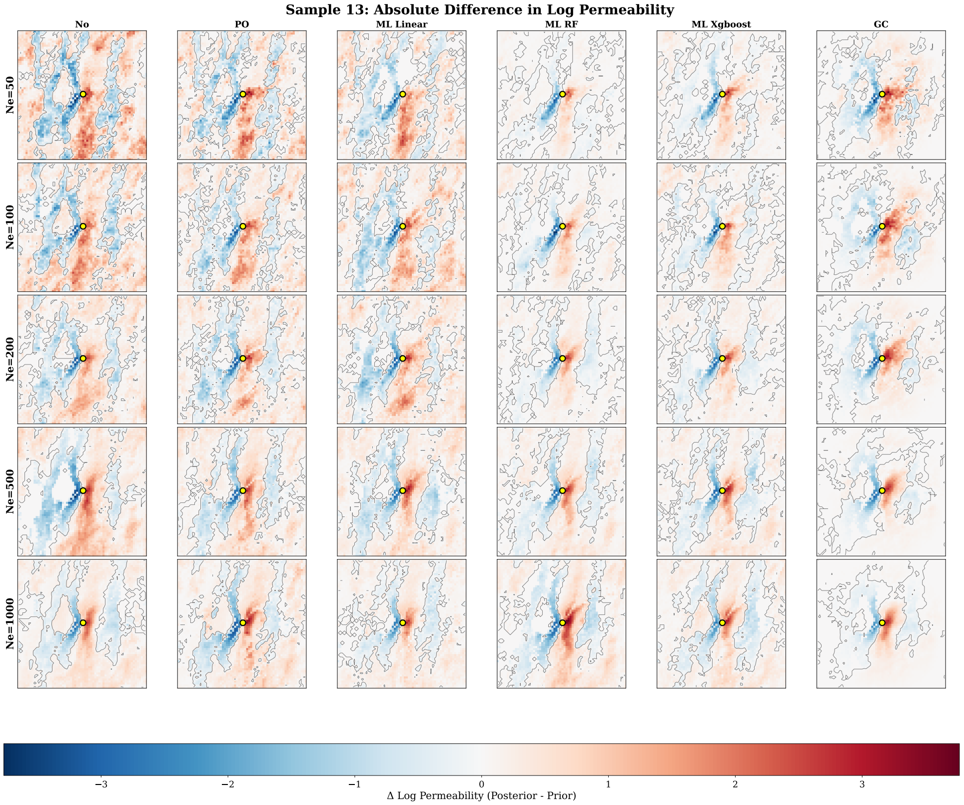


Figure 6.15: Log-permeability change (posterior – prior) by N_e and localization.

To quantify the quality of data matching across different ensemble sizes, Figure 6.16 presents the RMSE values achieved by each method, providing a comprehensive measure of data matching quality.

The RMSE analysis reveals distinct performance patterns across ensemble sizes. For larger ensembles ($N_e \geq 100$), all localization methods achieve acceptable data matching quality. However, very small ensembles ($N_e = 50$) show degraded performance across all methods, suggesting this ensemble size may be insufficient for reliable data assimilation in channelized systems, even with enhanced localization techniques. Despite these limitations at $N_e = 50$, ML-enhanced methods consistently outperform traditional approaches within each ensemble size category.

To illustrate the history matching in detail, Figure 6.17 presents pressure evolution at four monitoring wells throughout the 10-year CO_2 injection period for the $N_e = 500$ case, comparing different localization methods. All methods track the observed pressure evolution, with ensemble mean predictions closely following the true data.

The effectiveness of different localization methods in preventing ensemble collapse

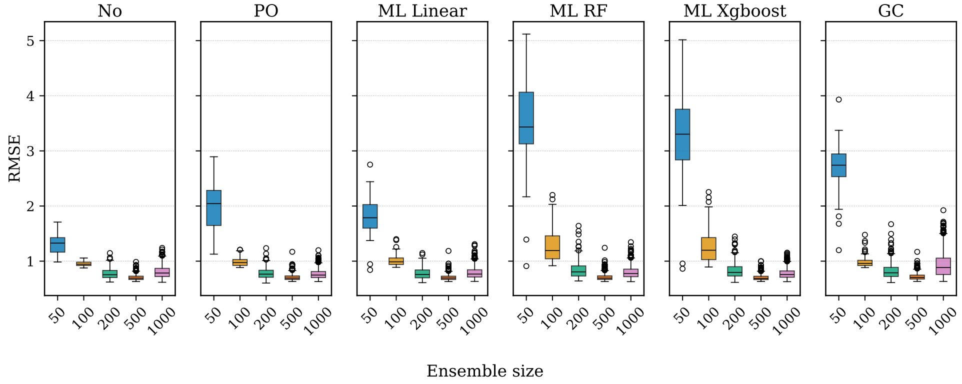


Figure 6.16: RMSE comparison across ensemble sizes.

is quantified through normalized variance (NV), our primary performance metric. Table 6.3 and Figure 6.18 present comprehensive results across all tested configurations.

Table 6.3: Normalized Variance (NV) Results for Different Localization Methods Across Ensemble Sizes

Method	N_e=50	N_e=100	N_e=200	N_e=500	N_e=1000
No Localization	0.400	0.632	0.714	0.801	0.765
Pseudo-Optimal	0.577	0.740	0.797	0.835	0.787
Gaspari-Cohn	0.749	0.857	0.904	0.927	0.882
ML-Linear	0.570	0.708	0.780	0.825	0.791
ML-Random Forest	0.864	0.885	0.871	0.861	0.811
ML-XGBoost	0.821	0.872	0.858	0.857	0.805

Random Forest and XGBoost win across all ensemble sizes. The improvements are most pronounced for small ensembles where sampling errors are severe. For $N_e = 50$, ML-Random Forest achieves NV = 0.864 compared to only 0.400 without localization—a 116% improvement that transforms a collapsed ensemble into one that maintains meaningful uncertainty quantification.

The framework exhibits distinct performance characteristics across different ensemble size regimes:

Small ensembles ($N_e < 100$): ML-enhanced localization may be necessary, preventing variance collapse. Without proper localization, small ensembles lose all geological structure and underestimate uncertainty. ML methods restore viable uncertainty quantification with minimal computational overhead.

Medium ensembles ($N_e = 100 - 500$): ML methods continue providing clear benefits, with variance preservation improvements of 20-40% over no localization. For $N_e =$ 200, ML methods achieve NV values of 0.86-0.87 compared to 0.71 without localization, maintaining better geological realism throughout the assimilation process.

Large ensembles ($N_e > 500$): Benefits diminish as natural sampling errors decrease, though ML methods still show marginal improvements. For $N_e = 1000$, all methods

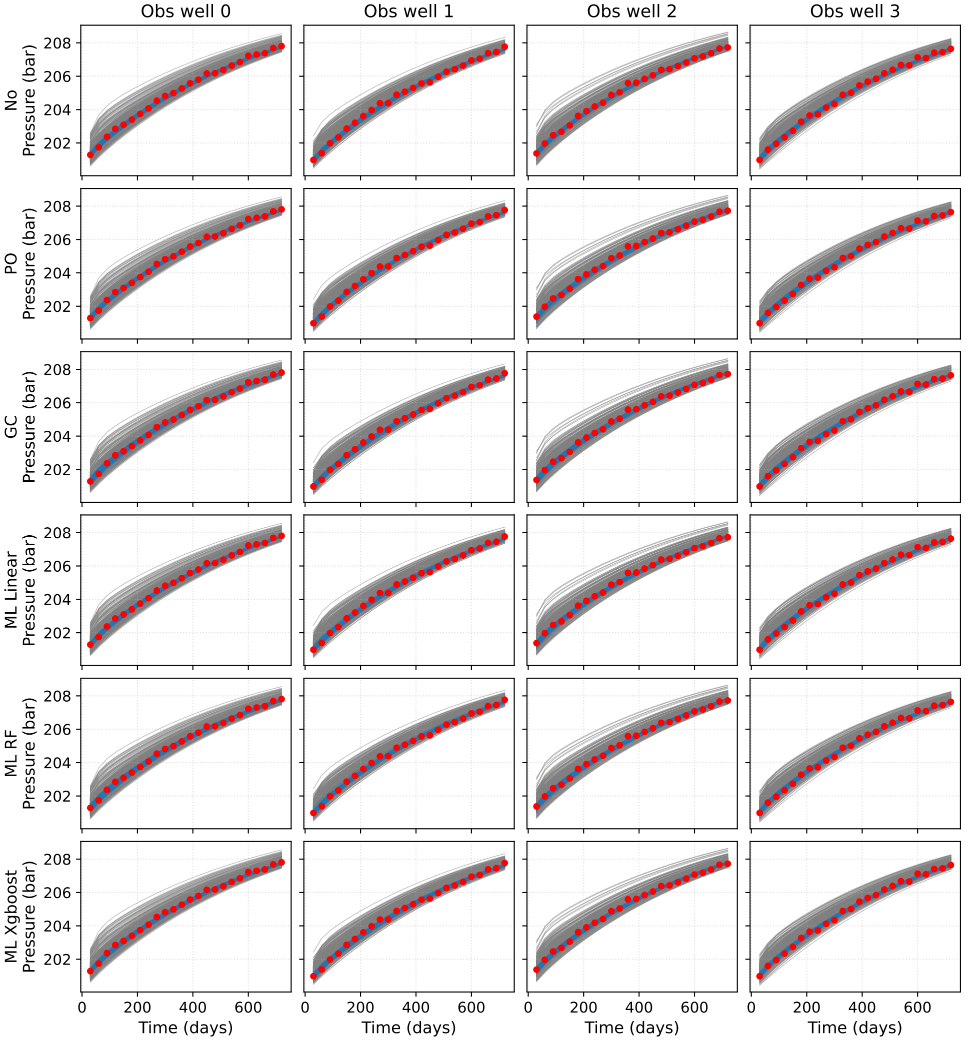


Figure 6.17: Data match for $N_e = 500$ across four monitoring wells. True data (red), posterior (blue), prior (gray) curves.

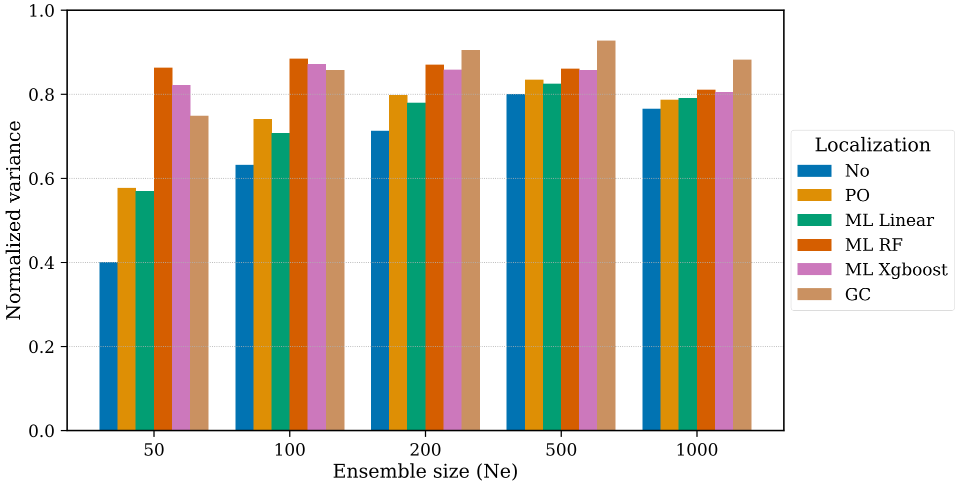


Figure 6.18: Normalized variance comparison across ensemble sizes.

achieve acceptable variance preservation (NV > 0.78), with ML approaches reaching 0.81.

6.5. DISCUSSION

This study investigates whether ensemble smoothers can retain geological realism and usable uncertainty in channelized CO_2 storage by combining diffusion-generated super-ensembles with ML-based covariance estimation for localization. The numerical experiments indicate progress on both known issues: (1) variance loss with small N_e and (2) connectivity-blind distance tapers (Emerick & Reynolds, 2013b; Evensen et al., 2019; Lacerda et al., 2021; Silva et al., 2025a, 2025b).

Across ensemble sizes, Random Forest and XGBoost localizations preserved more prior variance than linear, pseudo-optimal, or Gaspari–Cohn tapers. With N_e between 100 and 200, the added variance was roughly 20–40%, a practical regime for many studies. The advantage narrowed for $N_e > 500$, consistent with reduced sampling noise at larger N_e. These gains did not degrade data match quality; pressure misfits stayed comparable across methods.

An important design choice is that the ML proxy does not replace the simulator inside the update. It is used only to compute covariances on a large synthetic set, where speed matters and approximation error is acceptable. Localization depends on correlation structure rather than exact trajectory reproduction, so lightweight proxies suffice. This separation preserves the physics where it matters—the forward simulations—while enabling much larger effective sample sizes for covariance estimation.

Qualitatively, the spatial update patterns match these findings. Distance-based kernels create circular footprints around wells and update channels and background alike. ML-derived localization concentrates updates along flow paths connecting injection and monitoring wells and avoids spreading corrections into low-permeability regions. Posterior permeability fields retain sharper channels, especially for small ensembles, which supports more reliable long-term predictions.

The computational overhead is modest. Training the proxy at each assimilation step added only a few seconds for $N_e \leq 200$ and under half a minute for $N_e = 500$, negligible compared to a forward DARTS run.

The method still has limits. All tests are 2D; moving to 3D raises memory demands for diffusion training (practical limits near 128^3 on common GPUs) and increases forward costs. We focus on pressure data; extending to seismic, electromagnetic, or deformation observations is a logical next step in future work. We train proxies offline at each step; online learning that updates the proxy as new data arrive could further improve adaptability.

6.6. CONCLUSIONS

We combined score-based diffusion models, fast ML proxies, and ESMDA to build localization that preserves uncertainty and geological structure in channelized CO_2 storage problems. On Gaussian fields, the approach preserved 15–35% more variance than no localization; in channelized cases, improvements reached about 40%. After training, the diffusion model generates thousands of realistic permeability fields at low cost; the ML proxy converts those samples into parameter–data covariances; the resulting localization mitigates ensemble collapse without extra flow simulations. Across tests, the method outperformed distance-based tapers, maintained sharper channel structure in posteriors, and added only modest runtime overhead.

These results help bridge the gap between the small ensembles affordable in practice and the large ensembles needed for accurate covariance estimation. Next steps include 3D extensions, multi-physics data integration, and online learning for the proxy, aiming at uncertainty-aware workflows suitable for real-time management of large-scale CO_2 storage projects.

CODE AND DATA AVAILABILITY

The FLUVSIM-generated channelized reservoir dataset is available at https://doi.org/10.4121/a8ad7808-b923-4335-ba7a-898c8c1232be. The DARTS simulator used for reservoir simulations is available at https://gitlab.com/open-darts/open-darts. The dageo Python package for data assimilation in geosciences is available at https://github.com/tuda-geo/dageo.

7

CONCLUSIONS AND FUTURE PERSPECTIVES

This thesis addresses the challenge of uncertainty in Geological Carbon Storage (GCS). The central goal was to develop, integrate, and validate a methodology that combines physics-based modeling, data assimilation, and machine learning to improve uncertainty quantification, guide monitoring design, and make GCS a more predictable and reliable engineering practice. By progressing from foundational geomechanical analysis to advanced AI-enhanced data assimilation, we have assembled a suite of tools that systematically reduce uncertainty and provide a clearer understanding of subsurface processes.

7.1. CONCLUSIONS

The research objectives of this thesis were met through a series of interconnected studies. While each chapter builds upon insights from previous work, some methodologies (such as the geomechanical proxy model and diffusion-based priors) are also developed as independent contributions that can be applied separately or combined with other data assimilation frameworks.

In Chapter 2, we began by establishing the value of geomechanical monitoring. Through detailed simulations of a Brazilian offshore reservoir, we demonstrated that ground deformation provides a clear and measurable signal of subsurface pressure evolution and CO_2 plume behavior. A finding was that the location of maximum uplift can migrate away from the injection well as pressure dissipates, which challenges conventional wellhead-focused monitoring. We also showed that deformation continues long after injection ceases, highlighting the importance of post-injection monitoring for assessing long-term containment security.

In Chapter 3, we developed a way to integrate this geomechanical data into subsurface models. We combined a physics-based geomechanical proxy model and the Ensemble Smoother with Multiple Data Assimilation (ESMDA). By assimilating surface deformation data, we showed that we could constrain the uncertain permeability field. This was quantified using entropy reduction, where the method achieved a 53.1% decrease in plume uncertainty in a two-dimensional conceptual case and an 18.5% reduction in the more complex Brugge field dataset.

Building on this integrated setup, in Chapter 4 we introduced an information-theoretic method to quantitatively compare different monitoring strategies. By using Shannon entropy as a consistent metric for information gain, we evaluated the trade-off between the spatial density of sensors and the temporal frequency of measurements. The results showed that increasing the measurement frequency can provide an information gain that is comparable to, or even greater than, deploying a denser spatial array of sensors. This provides a quantitative basis for designing more cost-effective and efficient monitoring programs.

In Chapter 5, we extended the ESMDA framework established in Chapter 3 by the integration of machine learning surrogates to overcome the computational barriers that limit ensemble-based methods. Building on the proxy-model approach from earlier chapters, we developed two novel hybrid methods, SH-ESMDA and SH-RML, which use Fourier Neural Operators (FNOs) and a Transformer U-Net (T-UNet) as surrogates. These methods accelerated the data assimilation process by over 50% while ensuring the final predictions remained physically consistent by using the high-fidelity simulator for the posterior step. An innovation was the SH-RML method, which enabled the use of gradient-based variational data assimilation for simulators that lack an adjoint model.

Finally, in Chapter 6, we addressed the specific challenge of data assimilation in complex, non-Gaussian channelized reservoirs. We developed a method that combines score-based diffusion models for generating large, geologically realistic ensembles with a machine learning–enhanced localization strategy. The method retained 15–40% more of the original ensemble variance than standard techniques and made updates that followed the actual channel geometry. This preserved geological realism, a feature often lost when using conventional methods.

7.2. RECOMMENDATIONS FOR FUTURE RESEARCH

7.2.1. ON GEOMECHANICAL PROXY MODELING

The geomechanical proxies from Chapters 3 and 4 were fast because they made a key simplification: they assumed the rock properties were uniform. To make them more realistic, the next step is to adapt these proxies to handle rocks with properties that change from place to place. This can be approached by developing layered models, or by using machine learning to train a surrogate on a modest number of full-physics simulations, allowing the surrogate to learn the complex responses of heterogeneous rocks.

7.2.2. ON MONITORING STRATEGY DESIGN

The evaluation method from Chapter 4 provides a good foundation for cost-benefit analysis in monitoring design, but it did not factor in costs. A practical extension is to incorporate an economic analysis that weighs the information obtained from a monitoring plan against the real-world cost of deploying and operating sensors. It would also be valuable to investigate adaptive monitoring, in which measurement frequency and spatial focus are adjusted in real time as new data arrive.

7.2.3. ON HYBRID DATA ASSIMILATION AND ML SURROGATES

The hybrid methods in Chapter 5 demonstrated computational speedups of ensemble-based data assimilation, achieving over 50% reduction in time compared to conventional ESMDA on two-dimensional models. The next step is to apply and test these methods on full three-dimensional reservoir models, ensuring that the surrogates scale and remain accurate when accounting for vertical flow and gravity. Testing with field data from operational GCS projects—which are often noisy, sparse, and temporally discontinuous—will be needed for evaluating performance under operational conditions and assessing readiness for deployment.

7.2.4. ON GENERATIVE MODELS AND ADVANCED LOCALIZATION

The diffusion models in Chapter 6 were highly effective at generating realistic two-dimensional channelized reservoirs. Future work should extend these generative models to three dimensions, which better represent field conditions and capture the full complexity of geological structures including vertical heterogeneity and gravity segregation. In addition, the ML-enhanced localization was tested only with pressure data; the next step is to evaluate performance with other data types, such as time-lapse seismic or electromagnetic measurements. Different observation types provide complementary information about distinct physical properties and spatial scales—seismic data constrain large-scale elastic properties and fluid saturation, while electromagnetic measurements are sensitive to fluid conductivity and water content—enabling more robust multi-physics characterization.

7.3. FINAL THOUGHTS

At the beginning of this thesis, I included a quote from Richard Feynman:

> “I have approximate answers, and possible beliefs, and different degrees of certainty about different things, but I’m not absolutely sure of anything...”

This sentiment captures the core challenge of subsurface science and, in particular, Geological Carbon Storage. We can never be absolutely sure of the conditions deep underground. The work in this thesis was motivated by the need to move beyond this state of ambiguity and toward a more quantitative understanding of our uncertainty.

This research shows that we can make progress by combining our physical knowledge of the subsurface with observational data and the pattern-finding ability of machine learning. The goal of these methods is not to find a single “correct” answer. Instead, they provide quantitative estimates of the uncertainties, narrowing the range of possibilities in a way that is both physically sound and computationally efficient. We have shown that it is possible to design monitoring programs that are more effective than ad-hoc configurations based solely on regulatory requirements or engineering judgment, accelerate simulations without sacrificing physical fidelity, and preserve geological realism throughout the data assimilation process.

As GCS transitions from the realization of a small number of pilot projects to a global-scale climate mitigation technology, the ability to reliably quantify and manage subsurface uncertainty will be paramount. Quantitative uncertainty estimates provide the foundation for risk mitigation strategies in GCS operations, enabling informed decisions about injection rates, monitoring frequency, and long-term site management.

The methods presented in this thesis provide a set of practical tools to help make GCS a more predictable and transparent engineering practice. These tools help turn “possible beliefs” into probabilistic forecasts, transforming uncertainty from a vague concern into a measurable quantity that supports safer, better-informed decisions.

Beyond the specific application to GCS, this work advances data assimilation science in several ways. We demonstrated that ML-enhanced localization can preserve ensemble variance in strongly non-Gaussian systems where conventional distance-based methods fail, providing new insights into covariance estimation with limited samples. The hybrid surrogate-physics frameworks (SH-ESMDA, SH-RML) offer a template for integrating machine learning with ensemble and variational methods while maintaining physical consistency—a challenge in many inverse problems beyond subsurface applications. The systematic comparison of monitoring configurations using information-theoretic metrics establishes a methodology for quantifying observational value that extends to other fields requiring optimal sensor placement under resource constraints. By combining diffusion-based prior generation with adaptive localization, we showed that ensemble methods can maintain geological realism in channelized systems, addressing a long-standing limitation of Gaussian-assumption-based assimilation techniques.

A

DATA MATCH PLOTS FOR VARYING ENSEMBLE SIZES

This appendix presents the history-matching performance for all ensemble sizes considered in Chapter 6. Each figure shows the evolution of pressure at the four monitoring wells over the two-year injection period for the corresponding ensemble size N_e.

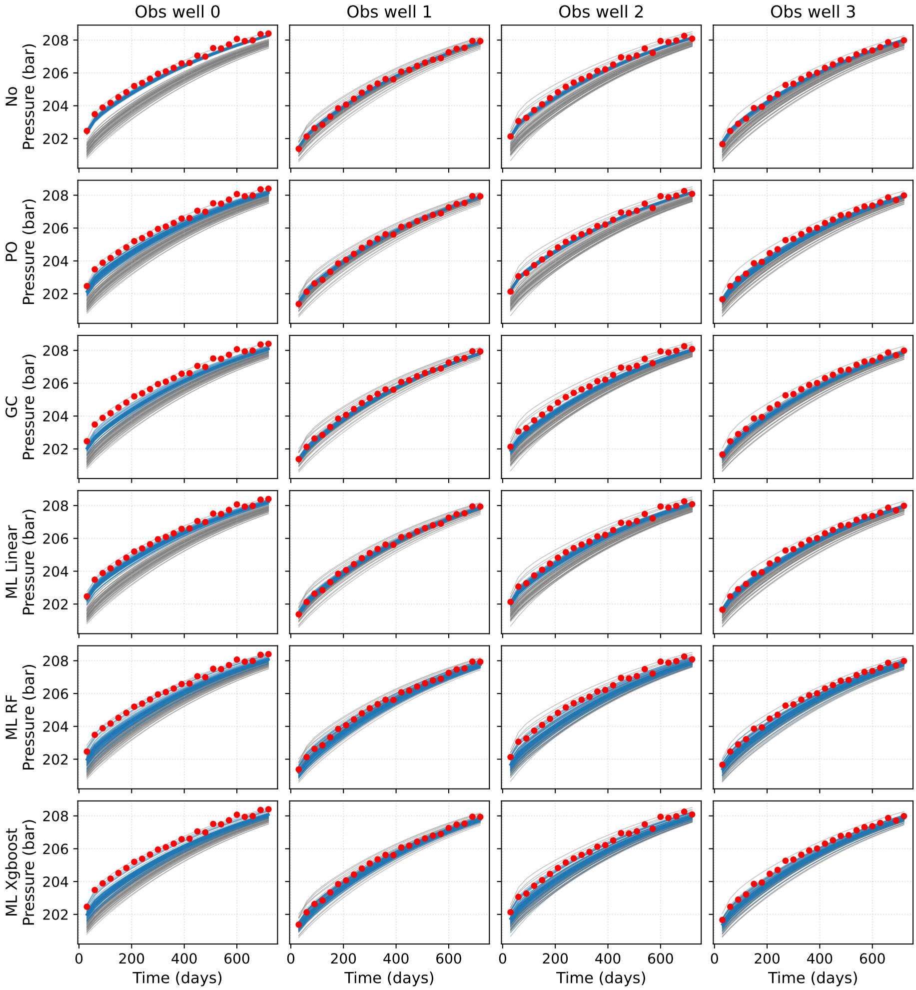


Figure A.1: Data match results for $N_e = 50$.

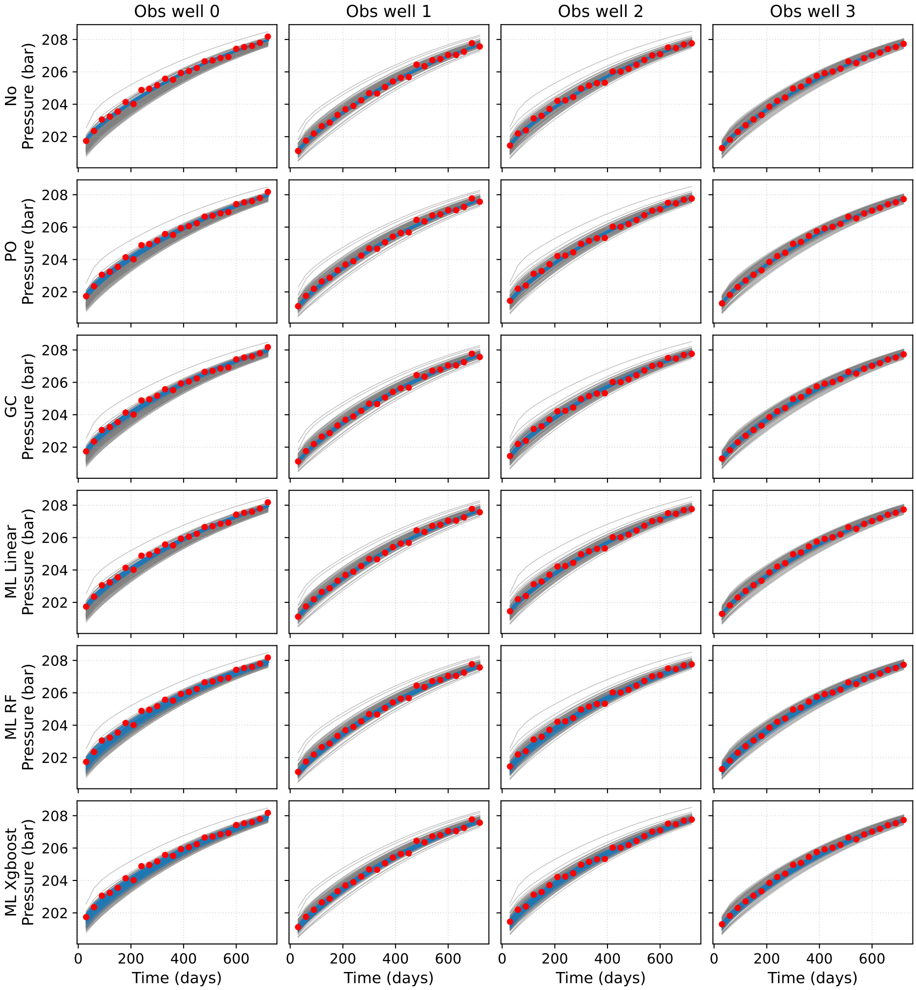


Figure A.2: Data match results for $N_e = 100$.

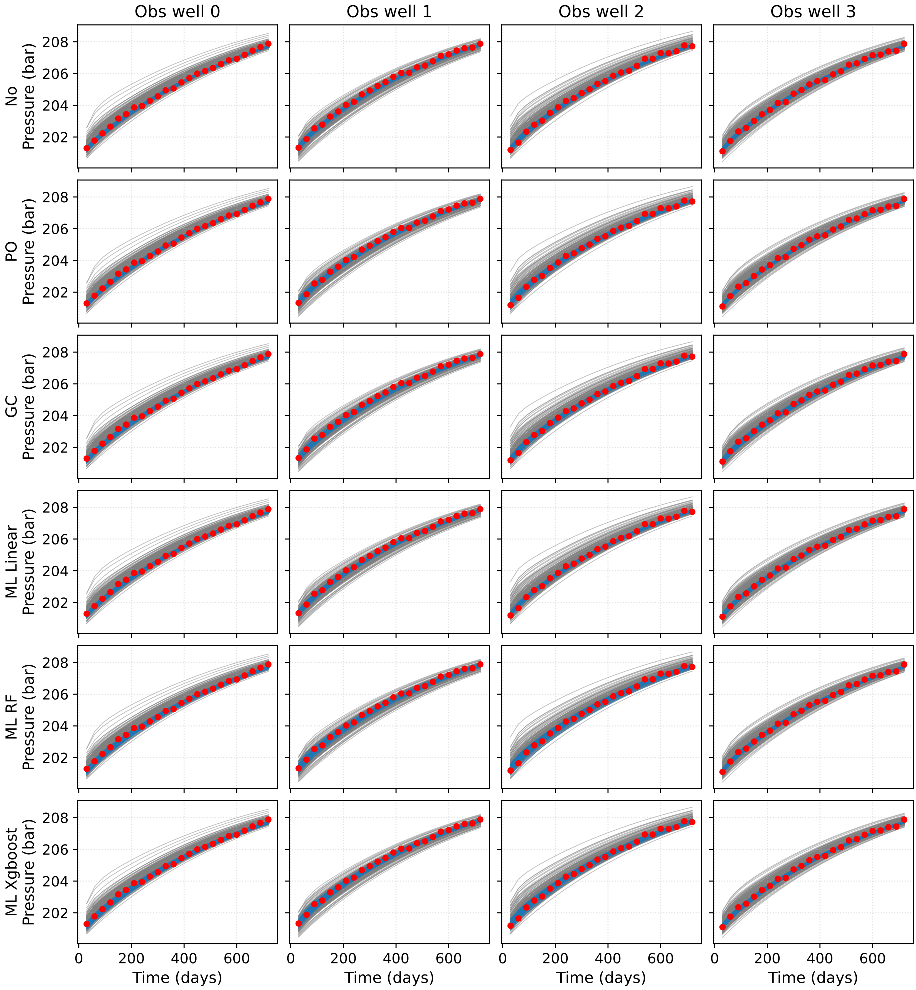


Figure A.3: Data match results for $N_e = 200$.

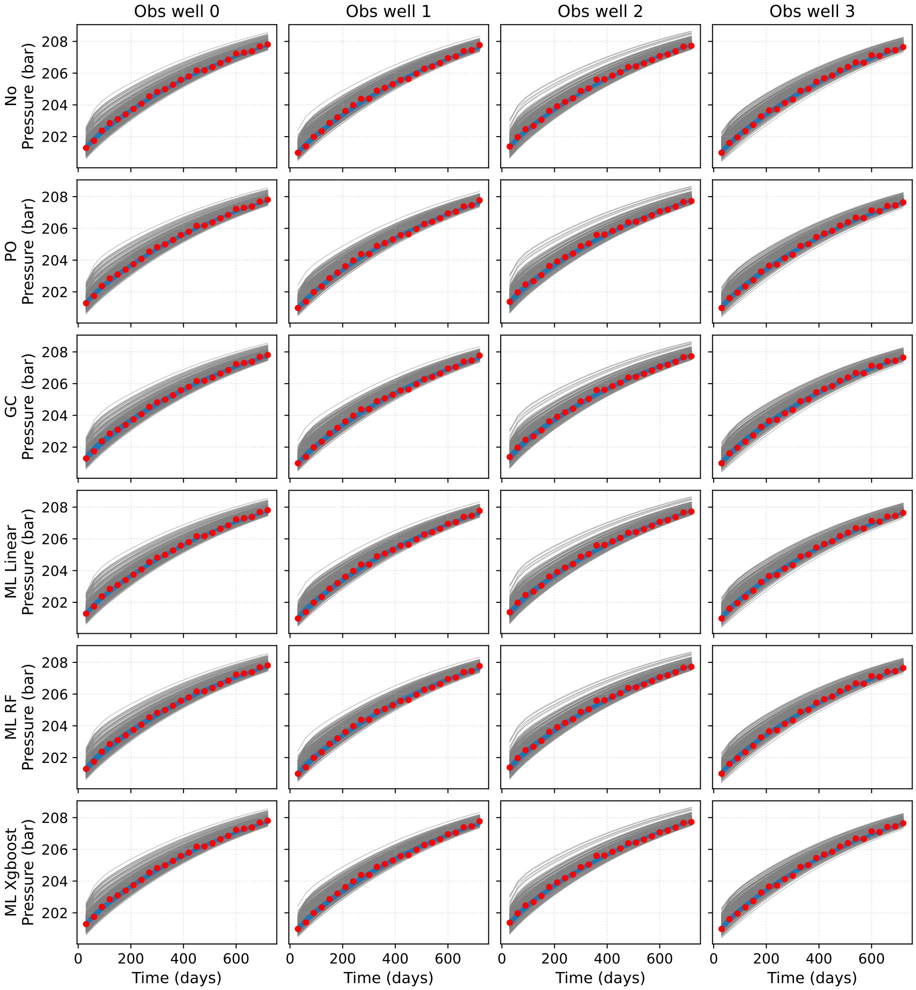


Figure A.4: Data match results for $N_e = 500$.

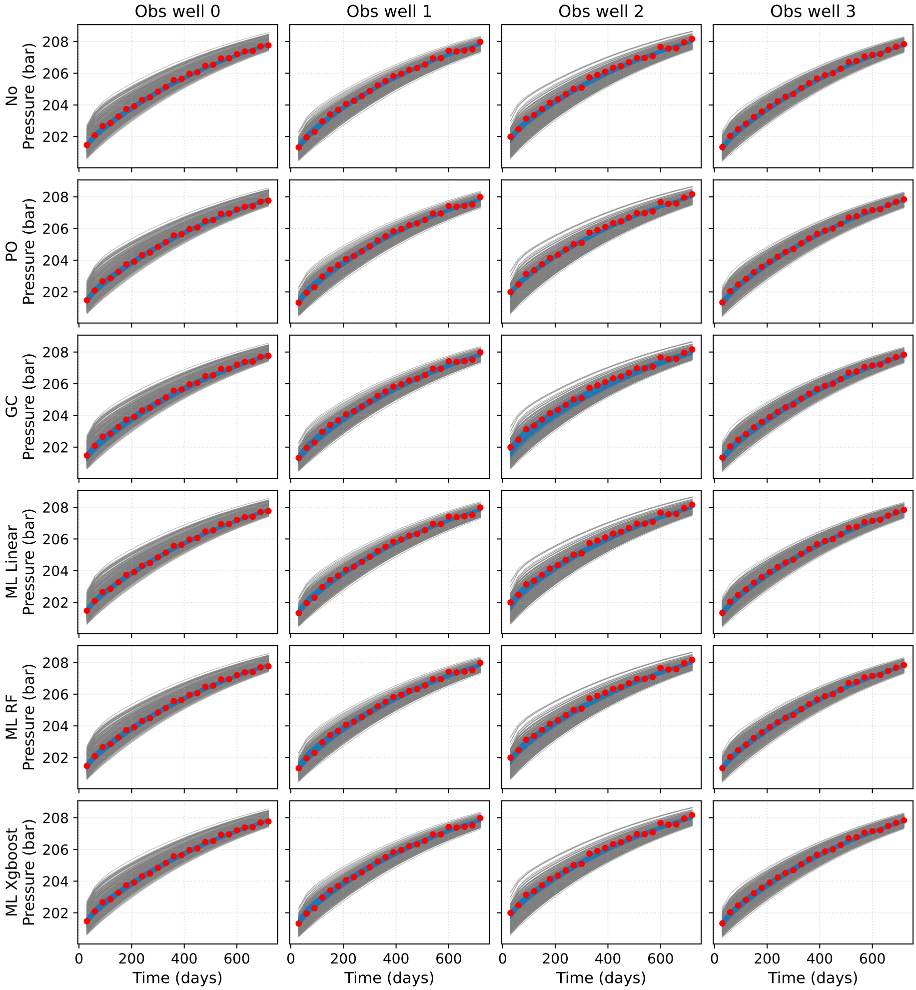


Figure A.5: Data match results for $N_e = 1000$.

B

LOCALIZATION MAPS FOR VARYING ENSEMBLE SIZES

This appendix presents the localization maps for all ensemble sizes considered in the study from Chapter 6 ($N_e = 50, 100, 200, 500, 1000$). Each figure shows the spatial pattern of the localization coefficients for a representative observation well and assimilation time step, illustrating how the localization adapts to the channelized reservoir structure for different ensemble sizes.

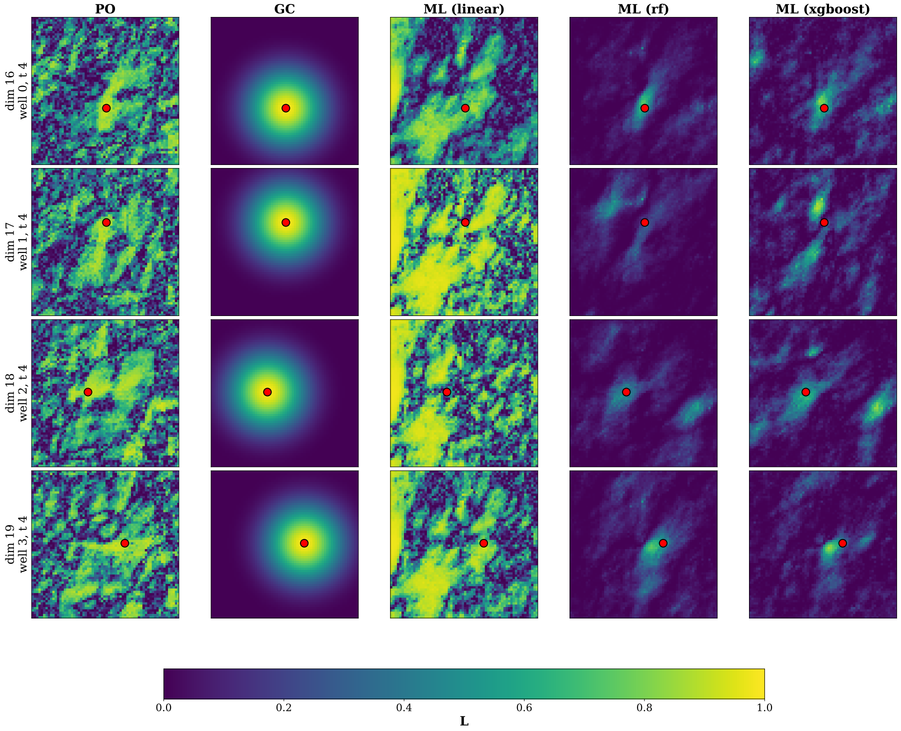


Figure B.1: Localization map for $N_e = 50$.

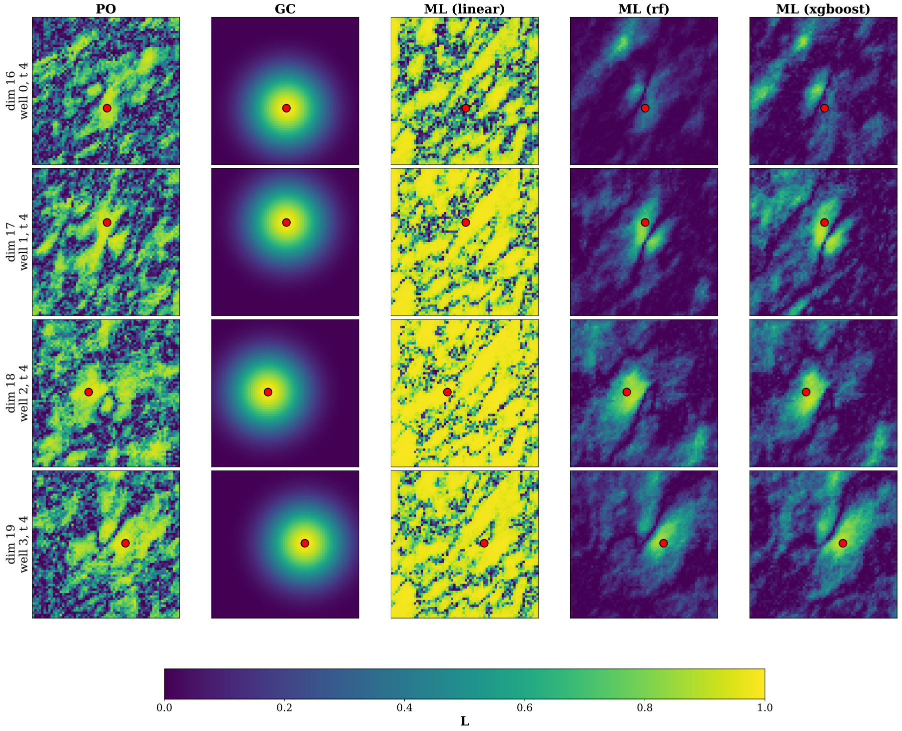


Figure B.2: Localization map for $N_e = 100$.

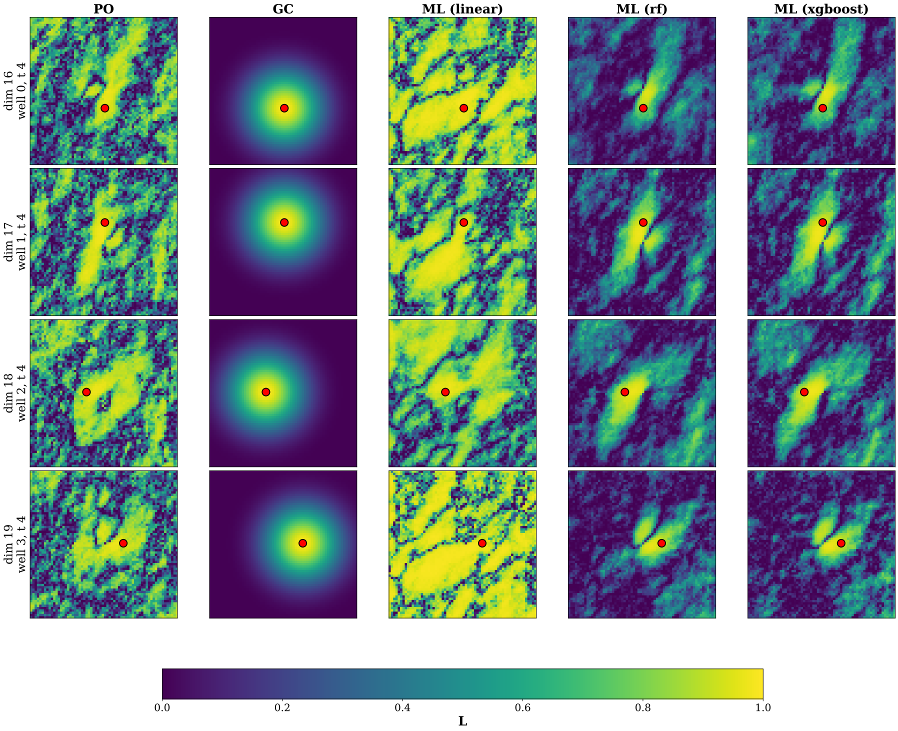


Figure B.3: Localization map for $N_e = 200$.

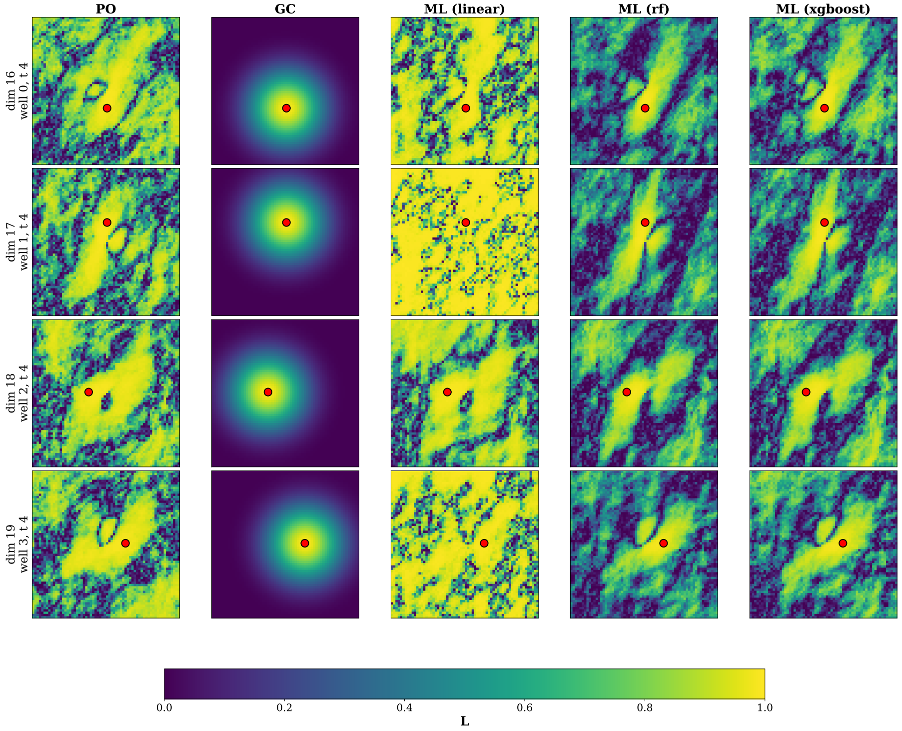


Figure B.4: Localization map for $N_e = 500$.

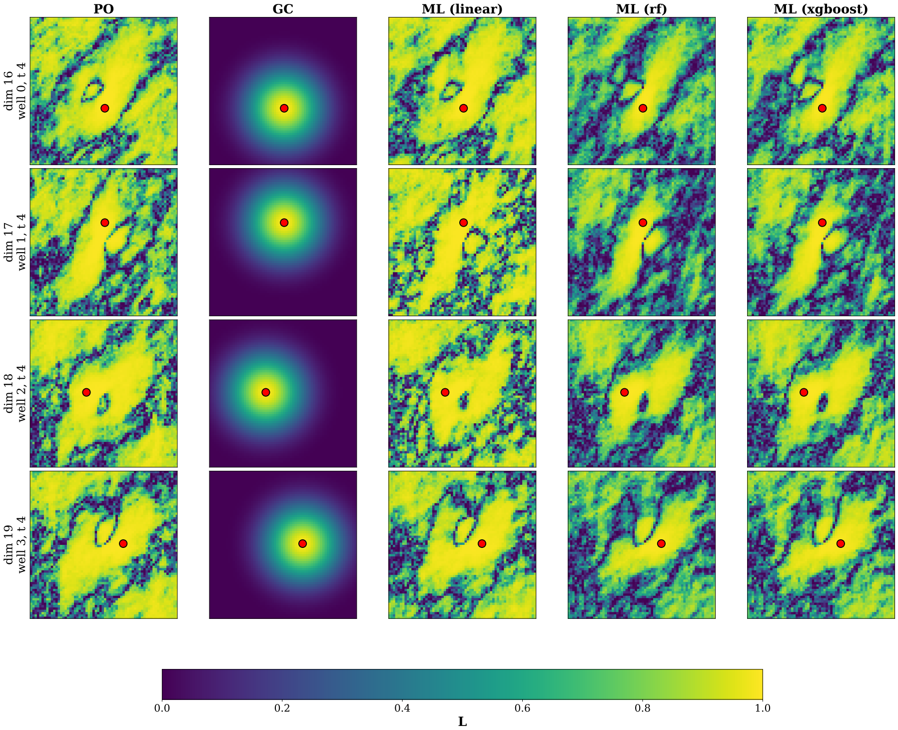


Figure B.5: Localization map for $N_e = 1000$.

NOMENCLATURE

DATA ASSIMILATION & STATISTICS

α, α_i ESMDA inflation factors for iteration i, dimensionless

$\mathbf{C}_{dd}$ Observation error covariance matrix

$\mathbf{C}^f_{dd}$ Auto-covariance matrix of forecasted data

$\mathbf{C}^f_{zd}$ Cross-covariance matrix between parameters and forecast data

$\mathbf{C}_{zz}$ Prior model covariance matrix for log-permeability field

$C_{\mathbf{corr}}(x, y)$ Spatial correlation function

$C_{\mathbf{corr}}(h)$ Spatial correlation function

$\boldsymbol{d}_{\mathbf{obs}}$ Vector of observed data (vertical displacements)

$\boldsymbol{d}_j$ Vector of perturbed observations for j-th ensemble member

d_n Individual observation value at point n

$\mathbf{d}$ Generic data vector

$\mathbf{d}^f$ Predicted data vector from the forward model

$\boldsymbol{\epsilon}_j$ Random perturbations from observation error distribution

ϵ_n Relative measurement error for observation point n

$\mathbf{G}()$ Forward model operator

$\boldsymbol{\eta}_j$ Vector of standard normal random variables

$H(X)$ Shannon entropy, nats

$H_{\mathbf{max}}$ Maximum entropy, nats

$H_{\mathbf{prior}}$ Prior entropy, nats

$H_{\mathbf{post}}$ Posterior entropy, nats

ΔH Entropy reduction, nats; Relative entropy reduction, %

I Mutual information, nats

$\boldsymbol{k}^a_j$ Analyzed permeability vector for j-th ensemble member

$\mathbf{L}$ Localization matrix

$L_{\mathbf{chol}}$ Cholesky factor of the prior covariance matrix $\mathbf{C}_{zz}$

L_x, L_y Correlation lengths in x and y directions, m

μ Mean of the logarithm of permeability, dimensionless

$\mu_{\mathbf{prior}}$ Mean of a distribution

N_a Number of assimilation cycles/steps in ESMDA

N_e Number of ensemble members

N_s Super-ensemble size

N_z Number of model parameters

$p(x_i)$ Probability of the i-th outcome in a distribution

$\boldsymbol{r}_j$ Vector of standard normal random variables for observation perturbation

$\rho_{\mathbf{corr},ij}$ Sample correlation coefficient between parameter i and observation j

$\tilde{\rho}_{ij}^{ML}$ ML-enhanced correlation coefficient

$\sigma_{\mathbf{err},n}$ Standard deviation of measurement error for observation point n

$\sigma_{\mathbf{obs}}$ Standard deviation of observation noise

$\theta_{\mathbf{rot}}$ Rotation angle for correlation function, radians

$\theta_{\mathbf{anis}}$ Angle of anisotropy for correlation function, radians

r_i Vector of standard normal random variables

$\mathbf{z}$ Model parameter vector (log-permeability field)

$\boldsymbol{z}_j^a$ Analyzed model parameter vector for j-th ensemble member

$\boldsymbol{z}_j^f$ Forecasted model parameter vector for j-th ensemble member

$J(\boldsymbol{z})$ Objective function

RESERVOIR & FLOW PHYSICS

$A_{\mathbf{pore}}$ Total pore compressibility

B Body forces, N/m^3

B_{CO_2} Parameter for Langmuir isotherm relation, 1/kPa

C Land's constant, dimensionless

c_f Rock compressibility, kPa^{-1}

$c_{\mathbf{rock}}$ Rock compressibility, 1/bar

$D_{\mathbf{diff}}$ Diffusion coefficient, cm^2/s

$D_{\mathbf{eff}}$ Effective diffusion coefficient, cm^2/s

$g_{\mathbf{grav}}$ Acceleration due to gravity, m/s^2

h Normalized distance for correlation function, dimensionless

$h_{\mathbf{enth},j}$ Enthalpy of phase j, kJ/kg

$H_{\mathbf{aq}}$ Aquifer thickness, m

H_i Henry's constant at current pressure (p) and temperature (T), dimensionless

H_i^* Henry's constant at reference pressure (p^*) and temperature (T), dimensionless

J Leverett J-function, dimensionless

k **or** k_h Average horizontal permeability, mD

$k(x, y)$ Permeability field, mD

$\mathbf{K}$ Permeability tensor, mD

k_v Average vertical permeability, mD

k_{rl} Relative permeability, dimensionless

k_{rj} Relative permeability of phase j, dimensionless

k_{rg}, k_{rw} End-point relative permeabilities for gas and water

L Length of the aquifer, m

L_w Horizontal well length, m

M Mobility ratio, dimensionless

$\mathbf{m}$ Model parameters; Conditioned permeability field

$\mu_{\mathbf{visc},j}$ Viscosity of phase j, Pa s

n_c Number of components

n_p Number of phases

n_g, n_w Brooks-Corey exponents for gas and water

N_j Total moles of mineral j, $\mathrm{gmol}/\mathrm{m}^3$

N_{gv} Characteristic time ratio for fluid to flow in the transverse direction due to gravity, dimensionless

p Pore pressure, Pa

p^* Reference pressure (Henry's law), Pa

p_j Pressure of phase j, bar

Δp Pressure change, bar

q_j Source/sink term for phase j

$\rho_{\mathbf{fluid}}$ Density, kg/m^3

ρ_m Mineral molar density, gmol/m^3

$\rho_{\mathbf{phase},j}$ Density of phase j, kg/m^3

S Saturation

s_j Saturation of phase j, fraction

S_{gt} Trapped gas saturation, dimensionless

S_{gi} Initial gas saturation, dimensionless

S_{gr} Residual gas saturation, dimensionless

S_{gc}, S_{wc} Critical gas and connate water saturations

$S_{g_{max}}$ Maximum gas saturation (for Land's trapping)

T Temperature, °C

ΔT Temperature change, °C

$\mathbf{u}_{\mathbf{vel},j}$ Velocity of phase j, m/s

U Internal energy

U_j Specific internal energy of phase j, kJ/kg

U_r Specific internal energy of rock, kJ/m^3

$\overline{v_i}$ Partial molar volume at infinite dilution, L/mol

R_g Universal gas constant, J/(mol K)

x_{cj} Molar fraction of component c in phase j, fraction

$x_{\mathbf{CO_2}}$ Molar fraction of CO_2, fraction

$\boldsymbol{f}$ CO_2 molar fraction field (state variable)

z Depth, m

z_j Depth of j-th nucleus of strain

Z Global mole fraction, dimensionless

GEOMECHANICS

α_B Biot-Willis coefficient, dimensionless

$\alpha_{\mathbf{therm}}$ Linear thermal expansion coefficient, 1/°C

C_m Uniaxial compaction coefficient, 1/MPa

E Young's modulus, MPa

$\boldsymbol{\epsilon}$ Strain tensor, dimensionless

η Thermoelastic coefficient, $1/\mathrm{K}$

$\gamma_{\mathbf{weight},j}$ Specific weight of phase j, $\mathrm{N/m^3}$

$G_{\mathbf{disp},x}, G_{\mathbf{disp},y}, G_{\mathbf{disp},z}$ Geometrical terms for displacement calculation

$\mathbf{I}$ Identity tensor

$\mathbf{K}$ Stiffness tensor

$\kappa_{\mathbf{therm}}$ Total thermal conductivity, $\mathrm{kJ/m/day/K}$

$\kappa_{\mathbf{total}}$ Total thermal conductivity

ν Poisson's ratio, dimensionless

ϕ Rock porosity, fraction

$\boldsymbol{\phi}$ Porosity field, fraction

$\Phi_{\mathbf{disp}}$ Poroelastic displacement potential

r Distance vector between evaluation point and nucleus of strain

r' Distance vector to image nucleus of strain (semi-infinite domain)

R Distance from evaluation point to nucleus of strain, m

dv_j Infinitesimal volume element in the j-th nucleus of strain

$\boldsymbol{\sigma}$ Stress tensor, Pa

τ Tortuosity, dimensionless

u Total thermoporoelastic displacement vector, m

$u^{\mathbf{poro}}$ Poroelastic displacement vector

$u^{\mathbf{therm}}$ Thermoelastic displacement vector

V Volume of the nucleus of strain, m^3

MACHINE LEARNING & NEURAL NETWORKS

$\mathbf{a}$ Fourier coefficients vector

$D_{\mathbf{dim}}$ Dimensionality of the data space

f^{ML} Machine learning proxy model

$g_{\mathbf{ML}}$ ML forward model operator

$\mathbf{g}_{\mathbf{grad}}$ Gradient of the log-likelihood (posterior sampling)

γ Conditioning strength for posterior diffusion sampling

$\mathbf{H}_{\mathbf{obs}}$ Observation operator matrix

j Index for ensemble members

$K(\cdot,\cdot)$ Fourier integral kernel

$\mathscr{L}(\theta)$ Loss function for score network training

$\lambda_{\mathbf{reg}}$ Regularization hyperparameter

N Number of cells or grid points

$N_{\mathbf{int}}$ Number of interior ViT layers

$N_{\mathbf{layers}}$ Number of neural network layers

$\nu_{\mathbf{UNet}}(\cdot)$ T-UNet architecture function

$\boldsymbol{\omega}_{\mathbf{NN}}$ Trainable neural network weights

P Penalty operator (Fourier Neural Operator)

$p_t(\cdot)$ Probability density function at diffusion time t

$\mathbf{r}$ Random normal vector

$s_{\mathbf{score},\theta}$ Neural network approximation of the score function

$s_\theta(\mathbf{x}_t, t)$ Score network output at time t

$g(t)$ Diffusion coefficient in SDE/ODE formulations

$\sigma_{\mathbf{act}}$ Activation function in neural networks

$\sigma_{\mathbf{noise}}$ Noise scale parameter for VE-SDE; also standard deviation

σ_t Marginal standard deviation of the data distribution at time t

t Diffusion time variable

$T_{\mathbf{diff}}$ Maximum diffusion time

$\theta_{\mathbf{NN}}$ Parameters of neural networks

$\tilde{\mathbf{C}}_{zd}^{ML}$ Machine learning-enhanced cross-covariance matrix

$v(x)$ Input function

$\mathbf{W}_t, \bar{\mathbf{W}}_t$ Standard Wiener process (forward and reverse)

$\mathbf{x}_t$ State vector (e.g., permeability field) at diffusion time t

$\mathbf{x}_0$ Initial data sample (from the true data distribution)

$\hat{\mathbf{x}}_0(\mathbf{x}_t)$ Denoised estimate of $\mathbf{x}_0$ given $\mathbf{x}_t$

$\mathbf{y}$ Observation vector

OPERATORS

$\circ$ Schur (element-wise) product

$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ Score function: gradient of the log-probability density

$\mathbb{E}[\cdot]$ Expectation operator

$\mathscr{F}$ Fourier transform operator

$\mathscr{G}_{\mathbf{NN},\theta}$ Neural operator parameterized by θ_{NN}

$\mathscr{N}(\mu, \Sigma)$ Normal distribution with mean μ and covariance Σ

$\mathscr{U}(a, b)$ Continuous uniform distribution on $[a, b]$

BIBLIOGRAPHY

Aanonsen, S. I., Nævdal, G., Oliver, D. S., Reynolds, A. C., & Vallès, B. (2009). The ensemble Kalman filter in reservoir engineering–a review. *SPE Journal*, *14*(03), 393–412. https://doi.org/10.2118/117274-PA

Abbate, E., Anastasi, P., Di Curzio, D., Facchi, G., & Della Rossa, E. (2022). Ensemble data space inversion for fast CO_2 injection forecast evaluation. *Eur. Conf. Math. Geol. Reserv. , ECMOR*. https://doi.org/10.3997/2214-4609.202244022

Abdoulghafour, H., Sarmadivaleh, M., Hauge, L. P., Fernø, M., & Iglauer, S. (2020). Capillary pressure characteristics of CO_2-brine-sandstone systems. *International Journal of Greenhouse Gas Control*, *94*. https://doi.org/10.1016/j.ijggc.2019.102876

Ahusborde, E., Amaziane, B., De Hoop, S., El Ossmani, M., Flauraud, E., Hamon, F. P., Kern, M., Socié, A., Su, D., Mayer, K. U., Tóth, M., & Voskov, D. (2024). A benchmark study on reactive two-phase flow in porous media: Part II - results and discussion. *Computational Geosciences*, *28*(3), 395–412. https://doi.org/10.1007/s10596-024-10269-y

AL-Ameri, W. A., Abdulraheem, A., & Mahmoud, M. (2016). Long-term effects of CO_2 sequestration on rock mechanical properties. *Journal of Energy Resources Technology*, *138*(1), 012201. https://doi.org/10.1115/1.4032011

Alsalmi, H., & Elsheikh, A. H. (2024). Automated seismic semantic segmentation using attention U-Net. *Geophysics*, *89*(1), WA247–WA263. https://doi.org/10.1190/geo2023-0149.1

Arcucci, R., Zhu, J., Hu, S., & Guo, Y.-K. (2021). Deep data assimilation: Integrating deep learning with data assimilation. *Applied Sciences (Switzerland)*, *11*(3), 1–21. https://doi.org/10.3390/app11031114

Assumpção, M. (1998). Seismicity and stresses in the Brazilian passive margin. *Bulletin of the Seismological Society of America*, *88*(1), 160–169. https://doi.org/10.1785/bssa0880010160

Avansi, G., & Schiozer, D. (2015). UNISIM-I: Synthetic model for reservoir development and management applications.

Bakker, R. J. (2003). Package FLUIDS 1. Computer programs for analysis of fluid inclusion data and for modelling bulk fluid properties. *Chemical Geology*, *194*(1-3), 3–23. https://doi.org/10.1016/S0009-2541(02)00268-1

Balhoff, M. (2022). *An introduction to multiphase, multicomponent reservoir simulation*. Elsevier.

Barbosa, V. C. F., Oliveira Jr, V. C., Arelaro, A. D., & Borges, F. A. S. (2022). 3D displacement and stress fields of compacting reservoir: Alternative solutions. *Brazilian Journal of Geophysics*, *40*(5). https://doi.org/10.22564/brjg.v40i5.2139

Bennion, D. B., & Bachu, S. (2008). Drainage and imbibition relative permeability relationships for supercritical CO_2/brine and H2S/brine systems in intergranular sandstone, carbonate, shale, and anhydrite rocks. *SPE Reservoir Evaluation and Engineering*, *11*(3), 487–496. https://doi.org/10.2118/99326-pa

Bizzo, Y. F. (2017). *Evaluation of the effects of fluid and rock properties on geomechanical simulations of reservoirs from the Namorado field* [Master's thesis].

Bjørnarå, T. I., Aker, E., Cuisiat, F., & Skurtveit, E. (2010). Modeling CO_2 storage using coupled reservoir-geomechanical analysis. *COMSOL Conference*.

Bocquet, M., Brajard, J., Carrassi, A., & Bertino, L. (2020). Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization. *Foundations of Data Science*, *2*(1), 55–80. https://doi.org/10.3934/fods.2020004

Bohloli, B., Berndt, C., Bjørnarå, T. I., Bouquet, S., Dujardin, J.-R., Eiken, O., Estublier, A., Fourno, A., Frauenfelder, R., Frey, J., Karstens, J., Lee, Y.-K., Marin-Moreno, H., Meland, H., Mondol, N. H., Park, J., Park, Y.-C., Sparrevik, P. M., Vincent, C., . . . Xue, Z. (2022). Monitoring CO_2 storage sites onshore and offshore using InSAR data and strain sensing fibre optics cables. https://doi.org/10.2139/ssrn.4286039

Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2024). Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(6), 7499–7519. https://doi.org/10.1109/TNNLS.2022.3229161

Boso, F., & Tartakovsky, D. M. (2018). Information-theoretic approach to bidirectional scaling. *Water Resources Research*, *54*(7), 4916–4928. https://doi.org/10.1029/2017WR021993

Bouqueta, S., Frey, J., Malinouskaya, I., Estublier, A., & Fourno, A. (2022). Evaluation of surface movement observability and optimization of the monitoring plan through conceptual and coupled flow-geomechanics models. Examples of carbonate and sandstone reservoirs in CCS context. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4301904

Braim, M., Bowman, S., & Waggott, B. (2023). Stress evolution during CO_2 storage- a case for long term monitoring. *First Break*, *41*(3), 43–47. https://doi.org/10.3997/1365-2397.fb2023019

Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2020). Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model. *Journal of Computational Science*, *44*, 101171. https://doi.org/10.1016/j.jocs.2020.101171

Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2021). Combining data assimilation and machine learning to infer unresolved scale parametrization. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2194), 20200086. https://doi.org/10.1098/rsta.2020.0086

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buizza, C., Quilodrán Casas, C., Nadler, P., Mack, J., Marrone, S., Titus, Z., Le Cornec, C., Heylen, E., Dur, T., Baca Ruiz, L., Heaney, C., Díaz Lopez, J. A., Kumar, K. S., & Arcucci, R. (2022). Data Learning: Integrating Data Assimilation and Machine Learning. *Journal of Computational Science*, *58*, 101525. https://doi.org/10.1016/j.jocs.2021.101525

Bump, A. P., Bakhshian, S., Ni, H., Hovorka, S. D., Olariu, M. I., Dunlap, D., Hosseini, S. A., & Meckel, T. A. (2023). Composite confining systems: Rethinking geologic seals for permanent CO_2 sequestration. *International Journal of Greenhouse Gas Control*, *126*, 103908. https://doi.org/10.1016/j.ijggc.2023.103908

Burchette, T. P. (2012). Carbonate rocks and petroleum reservoirs: A geological perspective from the industry. *Geological Society, London, Special Publications*, *370*(1), 17–37. https://doi.org/10.1144/SP370.14

Burnside, N., & Naylor, M. (2014). Review and implications of relative permeability of CO_2/brine systems and residual trapping of CO_2. *International Journal of Greenhouse Gas Control*, *23*, 1–11. https://doi.org/10.1016/j.ijggc.2014.01.013

Busch, A., Alles, S., Gensterblum, Y., Prinz, D., Dewhurst, D. N., Raven, M. D., Stanjek, H., & Krooss, B. M. (2008). Carbon dioxide storage potential of shales. *International Journal of Greenhouse Gas Control*, *2*(3), 297–308. https://doi.org/10.1016/j.ijggc.2008.03.003

Buscheck, T. A., Sun, Y., Chen, M., Hao, Y., Wolery, T. J., Bourcier, W. L., Court, B., Celia, M. A., Julio Friedmann, S., & Aines, R. D. (2012). Active CO_2 reservoir management for carbon storage: Analysis of operational strategies to relieve pressure buildup and improve injectivity. *International Journal of Greenhouse Gas Control*, *6*, 230–245. https://doi.org/10.1016/j.ijggc.2011.11.007

Candela, T., Chitu, A. G., Peters, E., Pluymaekers, M., Hegen, D., Koster, K., & Fokker, P. A. (2022a). Subsidence induced by gas extraction: A data assimilation framework to constrain the driving rock compaction process at depth. *Frontiers in Earth Science*, *10*, 713273. https://doi.org/10.3389/feart.2022.713273

Candela, T., Pluymaekers, M., Ampuero, J.-P., van Wees, J.-D., Buijze, L., Wassing, B., Osinga, S., Grobbe, N., & Muntendam-Bos, A. G. (2022b). Controls on the spatiotemporal patterns of induced seismicity in Groningen constrained by physics-based modelling with Ensemble-Smoother data assimilation. *Geophysical Journal International*, *229*(2), 1282–1308. https://doi.org/10.1093/gji/ggab497

Carlson, F. M. (1981). Simulation of relative permeability hysteresis to the nonwetting phase. *SPE Annual Technical Conference and Exhibition*, SPE–10157–MS. https://doi.org/10.2118/10157-MS

Chaturantabut, S., & Sorensen, D. C. (2010). Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, *32*(5), 2737–2764. https://doi.org/10.1137/090766498

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, Y., & Oliver, D. S. (2012). Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences*, *44*(1), 1–26. https://doi.org/10.1007/s11004-011-9376-z

Chen, Y., & Voskov, D. (2020). Optimization of CO_2 injection using multi-scale reconstruction of composition transport. *Computational Geosciences*, *24*(2), 819–835. https://doi.org/10.1007/s10596-019-09841-8

Cheng, S., Quilodrán-Casas, C., Ouala, S., Farchi, A., Liu, C., Tandeo, P., Fablet, R., Lucor, D., Iooss, B., Brajard, J., Xiao, D., Janjic, T., Ding, W., Guo, Y., Carrassi, A., Bocquet, M.,

& Arcucci, R. (2023). Machine Learning With Data Assimilation and Uncertainty Quantification for Dynamical Systems: A Review. *IEEE/CAA Journal of Automatica Sinica*, *10*(6), 1361–1387. https://doi.org/10.1109/JAS.2023.123537

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., & Ye, J. C. (2022). Diffusion posterior sampling for general noisy inverse problems. https://doi.org/10.48550/ARXIV.2209.14687

Computer Modelling Group Ltd. (2023). CMG GEM Compositional & Unconventional Simulator, (version 2023.30). https://www.cmgl.ca/

Courtier, P., Thépaut, J.-N., & Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, *120*(519), 1367–1387. https://doi.org/10.1002/qj.49712051912

Crain, D. M., Benson, S. M., Saltzer, S. D., & Durlofsky, L. J. (2024). An integrated framework for optimal monitoring and history matching in CO_2 storage projects. *Computational Geosciences*, *28*(2), 211–225. https://doi.org/10.1007/s10596-023-10216-3

Davis, E., Wright, C., Demetrius, S., Choi, J., & Craley, G. (2000). Precise tiltmeter subsidence monitoring enhances reservoir management. *SPE/AAPG Western Regional Meeting*, SPE–62577–MS. https://doi.org/10.2118/62577-MS

Davis, E., Krug, R., Debastiani, D., Oliveira, C., Pereira, L. C., & Rachid De Araujo, M. A. (2023). High precision seabed deformation monitoring. *Offshore Technology Conference*, D011S013R001. https://doi.org/10.4043/32207-MS

De Brito, D. U., & Durlofsky, L. J. (2021). Field development optimization using a sequence of surrogate treatments. *Computational Geosciences*, *25*(1), 35–65. https://doi.org/10.1007/s10596-020-09985-y

Delottier, H., Doherty, J., & Brunner, P. (2023). Data space inversion for efficient uncertainty quantification using an integrated surface and sub-surface hydrologic model. *Geoscientific Model Development*, *16*(14), 4213–4231. https://doi.org/10.5194/gmd-16-4213-2023

Delshad, M., Kong, X., Tavakoli, R., Hosseini, S. A., & Wheeler, M. F. (2013). Modeling and simulation of carbon sequestration at Cranfield incorporating new physical models. *International Journal of Greenhouse Gas Control*, *18*, 463–473. https://doi.org/10.1016/j.ijggc.2013.03.019

Denis Voskov, D., Saifullin, I., Wapperom, M., Tian, X., Palha, A., Orozco, L., & Novikov, A. (2025, July). *Open Delft Advanced Research Terra Simulator (open-DARTS)*. Zenodo. https://doi.org/10.5281/ZENODO.8046982

Deutsch, C., & Tran, T. (2002). FLUVSIM: A program for object-based stochastic modeling of fluvial depositional systems. *Computers & Geosciences*, *28*(4), 525–535. https://doi.org/10.1016/S0098-3004(01)00075-9

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat GANs on image synthesis. https://doi.org/10.48550/ARXIV.2105.05233

Di Federico, G., & Durlofsky, L. J. (2024). Latent diffusion models for parameterization and data assimilation of facies-based geomodels. https://doi.org/10.48550/ARXIV.2406.14815

Dong, Z., Cheng, J., Liu, T., Si, G., & Ma, B. (2021). A novel surface parametric method and its application to aerodynamic design optimization of axial compressors. *Processes*, *9*(7), 1230. https://doi.org/10.3390/pr9071230

Duan, Z., & Sun, R. (2003). An improved model calculating CO_2 solubility in pure water and aqueous NaCl solutions from 273 to 533 K and from 0 to 2000 bar. *Chemical Geology*, *193*(3-4), 257–271. https://doi.org/10.1016/S0009-2541(02)00263-2

DuMux Development Team. (2023). DuMux simulator. https://dumux.org

Eidsvik, J., Mukerji, T., & Bhattacharjya, D. (2015). *Value of information in the earth sciences: Integrating spatial modeling and decision analysis*. Cambridge University Press.

Elenius, M., Voskov, D., & Tchelepi, H. (2015). Interactions between gravity currents and convective dissolution. *Advances in Water Resources*, *83*, 77–88. https://doi.org/10.1016/j.advwatres.2015.05.006

Emerick, A. A., & Reynolds, A. C. (2011). History matching a field case using the ensemble Kalman filter with covariance localization. *SPE Reservoir Evaluation & Engineering*, *14*(04), 443–452. https://doi.org/10.2118/141216-PA

Emerick, A. A., & Reynolds, A. C. (2013a). Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. *Computational Geosciences*, *17*(2), 325–350. https://doi.org/10.1007/s10596-012-9333-z

Emerick, A. A., & Reynolds, A. C. (2013b). Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, *55*, 3–15. https://doi.org/10.1016/j.cageo.2012.03.011

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, *99*(C5), 10143–10162. https://doi.org/10.1029/94JC00572

Evensen, G. (2009). *Data assimilation: The ensemble Kalman filter*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03711-5

Evensen, G., Raanes, P. N., Stordal, A. S., & Hove, J. (2019). Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. *Frontiers in Applied Mathematics and Statistics*, *5*, 47. https://doi.org/10.3389/fams.2019.00047

Evensen, G., Vossepoel, F. C., & Van Leeuwen, P. J. (2022). *Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem*. Springer International Publishing. https://doi.org/10.1007/978-3-030-96709-3

Evensen, G., Vossepoel, F. C., & Van Leeuwen, P. J. (2024). Iterative ensemble smoothers for data assimilation in coupled nonlinear multiscale models. *Monthly Weather Review*, *152*(6), 1277–1301. https://doi.org/10.1175/MWR-D-23-0239.1

Fan, M., Liu, Y., Lu, D., Wang, H., & Zhang, G. (2025). A novel conditional generative model for efficient ensemble forecasts of state variables in large-scale geological carbon storage. *Journal of Hydrology*, *648*, 132323. https://doi.org/10.1016/j.jhydrol.2024.132323

Farajzadeh, R., Zitha, P. L. J., & Bruining, J. (2009). Enhanced mass transfer of CO_2 into water: Experiment and modeling. *Industrial & Engineering Chemistry Research*, *48*(13), 6423–6431. https://doi.org/10.1021/ie801521u

Farchi, A., Bocquet, M., Laloyaux, P., Bonavita, M., & Malartic, Q. (2021a). A comparison of combined data assimilation and machine learning methods for offline and online model error correction. *Journal of Computational Science*, *55*, 101468. https://doi.org/10.1016/j.jocs.2021.101468

Farchi, A., Laloyaux, P., Bonavita, M., & Bocquet, M. (2021b). Using machine learning to correct model error in data assimilation and forecast applications. *Quarterly Journal of the Royal Meteorological Society*, *147*(739), 3067–3084. https://doi.org/10.1002/qj.4116

Farshidi, S. F., Fan, Y., Durlofsky, L. J., & Tchelepi, H. A. (2013). Chemical reaction modeling in a compositional reservoir-simulation framework. *SPE Reservoir Simulation Symposium*, SPE–163677–MS. https://doi.org/10.2118/163677-MS

Fenghour, A., Wakeham, W. A., & Vesovic, V. (1998). The viscosity of carbon dioxide. *Journal of Physical and Chemical Reference Data*, *27*(1), 31–44. https://doi.org/10.1063/1.556013

Gahlot, A. P., Orozco, R., Yin, Z., Bruer, G., & Herrmann, F. J. (2025). An uncertainty-aware digital shadow for underground multimodal CO_2 storage monitoring. *Geophysical Journal International*, *242*(1), ggaf176. https://doi.org/10.1093/gji/ggaf176

García-Pintado, J., & Paul, A. (2018). Evaluation of iterative Kalman smoother schemes for multi-decadal past climate analysis with comprehensive Earth system models. *Geoscientific Model Development*, *11*(12), 5051–5084. https://doi.org/10.5194/gmd-11-5051-2018

Garipov, T. T., & Hui, M. (2018, June). *Simulation of coupled geomechanics and multiphase flow in naturally fractured reservoirs* (Vol. 52nd U.S. Rock Mechanics/Geomechanics Symposium).

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Gazzola, L., Ferronato, M., Frigo, M., Janna, C., Teatini, P., Zoccarato, C., Antonelli, M., Corradi, A., Dacome, M. C., & Mantica, S. (2021). A novel methodological approach for land subsidence prediction through data assimilation techniques. *Computational Geosciences*, *25*(5), 1731–1750. https://doi.org/10.1007/s10596-021-10062-1

Geertsma, J. (1973). Land subsidence above compacting oil and gas reservoirs. *Journal of Petroleum Technology*, *25*(06), 734–744. https://doi.org/10.2118/3730-PA

Geneva, N., & Zabaras, N. (2022). Transformers for modeling physical systems. *Neural Networks*, *146*, 272–289. https://doi.org/10.1016/j.neunet.2021.11.022

GEOSX Development Team. (2023). GEOSX Simulator [Accessed 8 March 2023]. https://geosx.org

Gercek, H. (2007). Poisson's ratio values for rocks. *International Journal of Rock Mechanics and Mining Sciences*, *44*(1), 1–13. https://doi.org/10.1016/j.ijrmms.2006.04.011

Gershenzon, N. I., Soltanian, M., Ritzi, R. W., & Dominic, D. F. (2014). Influence of small scale heterogeneity on CO_2 trapping processes in deep saline aquifers. *Energy Procedia*, *59*, 166–173. https://doi.org/10.1016/j.egypro.2014.10.363

Gharbi, O., Bijeljic, B., Boek, E., & Blunt, M. J. (2013). Changes in pore structure and connectivity induced by CO_2 injection in carbonates: A combined pore-scale approach. *Energy Procedia*, *37*, 5367–5378. https://doi.org/10.1016/j.egypro.2013.06.455

Gholami, R., Raza, A., & Iglauer, S. (2021). Leakage risk assessment of a CO_2 storage site: A review. *Earth-Science Reviews*, *223*, 103849. https://doi.org/10.1016/j.earscirev.2021.103849

Glegola, M., Ditmar, P., Hanea, R. G., Eiken, O., Vossepoel, F. C., Arts, R., & Klees, R. (2012). History matching time-lapse surface-gravity and well-pressure data with ensemble smoother for estimating gas field aquifer support—A 3D numerical study. *SPE Journal*, *17*(04), 966–980. https://doi.org/10.2118/161483-PA

Goodfellow, I., Bengio, Y., & Courville, A. (2020). *Deep learning*. MIT Press. https://www.deeplearningbook.org

Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). Revisiting deep learning models for tabular data. https://doi.org/10.48550/ARXIV.2106.11959

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data? https://doi.org/10.48550/ARXIV.2207.08815

Gunter, W. D., Bachu, S., & Benson, S. (2004). The role of hydrogeological and geochemical trapping in sedimentary basins for secure geological storage of carbon dioxide. *Geological Society, London, Special Publications*, *233*(1), 129–145. https://doi.org/10.1144/GSL.SP.2004.233.01.09

Guyant, E., Han, W. S., Kim, K.-Y., Park, M.-H., & Kim, B.-Y. (2015). Salt precipitation and CO_2/brine flow distribution under different injection well completions. *International Journal of Greenhouse Gas Control*, *37*, 299–310. https://doi.org/10.1016/j.ijggc.2015.03.020

Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, *129*(11), 2776–2790. https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2

Han, W. S., McPherson, B. J., Lichtner, P. C., & Wang, F. P. (2010). Evaluation of trapping mechanisms in geologic CO_2 sequestration: Case study of SACROC northern platform, a 35-year CO_2 injection site. *American Journal of Science*, *310*(4), 282–324. https://doi.org/10.2475/04.2010.03

Harvey, S., O'Brien, S., Minisini, S., Oates, S., & Braim, M. (2021). Quest CCS facility: Microseismic system monitoring and observations. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3817042

Henbest, S., Mills, A., & Ottey, P. (1978). Two tiltmeters and an integrating seismometer for the monitoring of volcanic activity, and the results of some trials on Mount Etna. *Journal of Volcanology and Geothermal Research*, *4*(1-2), 133–149. https://doi.org/10.1016/0377-0273(78)90033-1

Hesthaven, J. S., Rozza, G., & Stamm, B. (2016). *Certified reduced basis methods for parametrized partial differential equations*. Springer International Publishing. https://doi.org/10.1007/978-3-319-22470-1

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. https://doi.org/10.48550/ARXIV.2006.11239

Hosa, A., Esentia, M., Stewart, J., & Haszeldine, S. (2011). Injection of CO_2 into saline formations: Benchmarking worldwide projects. *Chemical Engineering Research and Design*, *89*(9), 1855–1864. https://doi.org/10.1016/j.cherd.2011.04.003

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *129*(1), 123–137. https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*(12), 4489–4532. https://doi.org/10.1175/MWR-D-15-0440.1

Hovorka, S. (2009, May). *Optimization of geological environments for carbon dioxide disposal in saline aquifers in the United States (Part One)* (tech. rep.). https://doi.org/10.2172/990445

IEA. (2022). *Carbon capture, utilisation and storage - fuels & technologies* (Technical Report). International Energy Agency.

Im, K., Elsworth, D., Guglielmi, Y., & Mattioli, G. S. (2017). Geodetic imaging of thermal deformation in geothermal reservoirs - production, depletion and fault reactivation. *Journal of Volcanology and Geothermal Research*, *338*, 79–91. https://doi.org/10.1016/j.jvolgeores.2017.03.021

Jenkins, J. (2015). *A case study of the petra nova carbon capture project* (Case Study Report). The Paulson Institute. Chicago, IL, USA.

Ketelaar, G., Bähr, H., Liu, S., Piening, H., Van Der Veen, W., Hanssen, R., Van Leijen, F., Van Der Marel, H., & Samiei-Esfahany, S. (2020). Integrated monitoring of subsidence due to hydrocarbon production: Consolidating the foundation. *Proceedings of the International Association of Hydrological Sciences*, *382*, 117–123. https://doi.org/10.5194/piahs-382-117-2020

Khait, M., & Voskov, D. V. (2017). Operator-based linearization for general purpose reservoir simulation. *Journal of Petroleum Science and Engineering*, *157*, 990–998. https://doi.org/10.1016/j.petrol.2017.08.009

Kim, J., Tchelepi, H. A., & Juanes, R. (2011). Stability, accuracy, and efficiency of sequential methods for coupled flow and geomechanics. *SPE Journal*, *16*(02), 249–262. https://doi.org/10.2118/119084-PA

Kim, K., Vilarrasa, V., & Makhnenko, R. Y. (2018). CO_2 injection effect on geomechanical and flow properties of calcite-rich reservoirs. *Fluids*, *3*(3), 66. https://doi.org/10.3390/fluids3030066

Kim, M., Kwon, S., Ji, M., Shin, H., & Min, B. (2023). Multi-lateral horizontal well with dual-tubing system to improve CO_2 storage security and reduce CCS cost. *Applied Energy*, *330*, 120368. https://doi.org/10.1016/j.apenergy.2022.120368

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

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. https://doi.org/10.48550/ARXIV.1412.6980

Kingma, D. P., & Welling, M. (2022, December). Auto-encoding variational Bayes. https://doi.org/10.48550/arXiv.1312.6114

Korondi, P. Z., Marchi, M., Parussini, L., & Poloni, C. (2021). Multi-fidelity design optimisation strategy under uncertainty with limited computational budget. *Optimization and Engineering*, *22*(2), 1039–1064. https://doi.org/10.1007/s11081-020-09510-1

Kuk, M., Kuk, E., Janiga, D., Wojnarowski, P., & Stopa, J. (2020). Optimization wells placement policy for enhanced CO_2 storage capacity in mature oil reservoirs. *Energies*, *13*(16), 4054. https://doi.org/10.3390/en13164054

Lacerda, A., Luo, X., & Emerick, A. A. (2019). Comparison of adaptive localization methods for updating petrophysical properties and areal parameters. *Journal of Petroleum Science and Engineering*, *183*, 106362. https://doi.org/10.1016/j.petrol.2019.106362

Lacerda, J. M., Emerick, A. A., & Pires, A. P. (2021). Using a machine learning proxy for localization in ensemble data assimilation. *Computational Geosciences*, *25*(3), 931–944. https://doi.org/10.1007/s10596-020-10031-0

Land, C. S. (1968). Calculation of imbibition relative permeability for two- and three-phase flow from rock properties. *Society of Petroleum Engineers Journal*, *8*(02), 149–156. https://doi.org/10.2118/1942-PA

Landry, C. J., Prodanović, M., & Eichhubl, P. (2016). Direct simulation of supercritical gas flow in complex nanoporous media and prediction of apparent permeability. *International Journal of Coal Geology*, *159*, 120–134. https://doi.org/10.1016/j.coal.2016.03.015

Le, D. H., & Reynolds, A. C. (2013). Estimation of mutual information and conditional entropy for surveillance optimization. *SPE Reservoir Simulation Symposium*, SPE–163638–MS. https://doi.org/10.2118/163638-MS

Lei, T., & Luo, K. H. (2021). Pore-scale simulation of miscible viscous fingering with dissolution reaction in porous media. *Physics of Fluids*, *33*(3), 034134. https://doi.org/10.1063/5.0045051

Li, D., Beckner, B., & Kumar, A. (2001). A new efficient averaging technique for scaleup of multimillion-cell geologic models. *SPE Reservoir Evaluation & Engineering*, *4*(04), 297–307. https://doi.org/10.2118/72599-PA

Li, G., Jin, D., Yu, Q., & Qi, M. (2023). IB-TransUNet: Combining information bottleneck and Transformer for medical image segmentation. *Journal of King Saud University - Computer and Information Sciences*, *35*(3), 249–258. https://doi.org/10.1016/j.jksuci.2023.02.012

Li, Q., & Liu, G. (2016). Risk assessment of the geological storage of CO_2: A review. In V. Vishal & T. Singh (Eds.), *Geologic Carbon Sequestration* (pp. 249–284). Springer International Publishing. https://doi.org/10.1007/978-3-319-27019-7_13

Li, Y. J., Kokkinaki, A., Darve, E. F., & Kitanidis, P. K. (2017). Smoothing-based compressed state Kalman filter for joint state-parameter estimation: Applications in reservoir characterization and CO_2 storage monitoring. *Water Resources Research*, *53*(8), 7190–7207. https://doi.org/10.1002/2016WR020168

Li, Y.-K., & Nghiem, L. X. (1986). Phase equilibria of oil, gas and water/brine mixtures from a cubic equation of state and henry's law. *The Canadian Journal of Chemical Engineering*, *64*(3), 486–496. https://doi.org/10.1002/cjce.5450640319

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2021, May). Fourier neural operator for parametric partial differential equations. https://doi.org/10.48550/arXiv.2010.08895

Lima, C., Nascimento, E., & Assumpção, M. (1997). Stress orientations in Brazilian sedimentary basins from breakout analysis: Implications for force models in the South American plate. *Geophysical Journal International*, *130*(1), 112–124. https://doi.org/10.1111/j.1365-246X.1997.tb00991.x

Lima, R. O., Do Nascimento Guimarães, L. J., & Pereira, L. C. (2021). Evaluating geomechanical effects related to the production of a Brazilian reservoir. *Journal of Petroleum Exploration and Production Technology*, *11*(6), 2661–2678. https://doi.org/10.1007/s13202-021-01190-6

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11966–11976. https://doi.org/10.1109/CVPR52688.2022.01167

Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, *3*(3), 218–229. https://doi.org/10.1038/s42256-021-00302-5

Lyu, X., & Voskov, D. (2023). Advanced modeling of enhanced CO_2 dissolution trapping in saline aquifers. *International Journal of Greenhouse Gas Control*, *127*, 103907. https://doi.org/10.1016/j.ijggc.2023.103907

Machado, M. V. B. (2023, July). *Numerical petroleum reservoir modeling: Integrated simulation practice*. Petrobras.

Machado, M. V. B., Delshad, M., & Sepehrnoori, K. (2023a). Potential benefits of horizontal wells for CO_2 injection to enhance storage security and reduce leakage risks. *Applied Sciences*, *13*(23), 12830. https://doi.org/10.3390/app132312830

Machado, M. V. B., Delshad, M., & Sepehrnoori, K. (2023b). Injectivity assessment for CCS field-scale projects with considerations of salt deposition, mineral dissolution, fines migration, hydrate formation, and non-Darcy flow. *Fuel*, *353*, 129148. https://doi.org/10.1016/j.fuel.2023.129148

Machado, M. V. B., Khanal, A., & Delshad, M. (2024). Unveiling the essential parameters driving mineral reactions during CO_2 storage in carbonate aquifers through proxy models. *Applied Sciences*, *14*(4), 1465. https://doi.org/10.3390/app14041465

March, R., Doster, F., & Geiger, S. (2018). Assessment of CO_2 storage potential in naturally fractured reservoirs with dual-porosity models. *Water Resources Research*, *54*(3), 1650–1668. https://doi.org/10.1002/2017WR022159

Mariethoz, G., & Caers, J. (2014, October). *Multiple-point geostatistics: Stochastic modeling with training images* (1st ed.). Wiley. https://doi.org/10.1002/9781118662953

Molina, O., Vilarrasa, V., & Zeidouni, M. (2017). Geologic carbon storage for shale gas recovery. *Energy Procedia*, *114*, 5748–5760. https://doi.org/10.1016/j.egypro.2017.03.1713

Montegrossi, G., Cantucci, B., Piochi, M., Fusi, L., Misnan, M. S., Rashidi, M. R. A., Abu Bakar, Z. A., Tuan Harith, Z. Z., Bahri, N. H. S., & Hashim, N. (2022). CO_2 reaction-diffusion experiments in shales and carbonates. *Minerals*, *13*(1), 56. https://doi.org/10.3390/min13010056

Morris, J., Hao, Y., Foxall, W., & McNab, W. (2011). In Salah CO_2 storage JIP: Hydromechanical simulations of surface uplift due to CO_2 injection at In Salah. *Energy Procedia*, *4*, 3269–3275. https://doi.org/10.1016/j.egypro.2011.02.246

Mücke, N. T., Bohté, S. M., & Oosterlee, C. W. (2021). Reduced order modeling for parameterized time-dependent PDEs using spatially and memory aware deep learning. *Journal of Computational Science*, *53*, 101408. https://doi.org/10.1016/j.jocs.2021.101408

Nagel, N. (2001). Compaction and subsidence issues within the petroleum industry: From Wilmington to Ekofisk and beyond. *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy*, *26*(1-2), 3–14. https://doi.org/10.1016/S1464-1895(01)00015-1

Nagy, D., Papp, G., & Benedek, J. (2000). The gravitational potential and its derivatives for the prism. *Journal of Geodesy*, *74*(7-8), 552–560. https://doi.org/10.1007/s001900000116

Neufeld, J. A., Hesse, M. A., Riaz, A., Hallworth, M. A., Tchelepi, H. A., & Huppert, H. E. (2010). Convective dissolution of carbon dioxide in saline aquifers. *Geophysical Research Letters*, *37*(22), 2010GL044728. https://doi.org/10.1029/2010GL044728

Nghiem, L., Shrivastava, V., Kohse, B., Hassam, M., & Yang, C. (2009). Simulation of trapping processes for CO_2 storage in saline aquifers. *Canadian International Petroleum Conference*, PETSOC–2009–156. https://doi.org/10.2118/2009-156

Novikov, A. (2024). *A finite volume framework for accurate modeling of fault reactivation in poroelastic rocks* [Doctoral dissertation, Delft University of Technology]. https://doi.org/10.4233/UUID:1F4731B3-CBDD-4317-98F7-8520B7446684

Novikov, A., Voskov, D., Khait, M., Hajibeygi, H., & Jansen, J. D. (2022). A scalable collocated finite volume scheme for simulation of induced fault slip. *Journal of Computational Physics*, *469*, 111598. https://doi.org/10.1016/j.jcp.2022.111598

Ohlberger, M., & Rave, S. (2015). Reduced basis methods: Success, limitations and future challenges. *Proceedings of Algoritmy 2016*, 1–12.

Oliver, D. S., & Chen, Y. (2011). Recent progress on reservoir history matching: A review. *Computational Geosciences*, *15*(1), 185–221. https://doi.org/10.1007/s10596-010-9194-2

Oliver, D. S., Reynolds, A. C., & Liu, N. (2008, May). *Inverse theory for petroleum reservoir characterization and history matching* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511535642

open-DARTS Development Team. (2024). Open-DARTS: Open Delft Advanced Research Terra Simulator. https://gitlab.com/open-darts/open-darts

Paullo Muñoz, L. F., & Roehl, D. (2017). An analytical solution for displacements due to reservoir compaction under arbitrary pressure changes. *Applied Mathematical Modelling*, *52*, 145–159. https://doi.org/10.1016/j.apm.2017.06.023

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, *12*(null), 2825–2830.

Peters, E., Arts, R. J., Brouwer, G. K., Geel, C. R., Cullick, S., Lorentzen, R. J., Chen, Y., Dunlop, K. N. B., Vossepoel, F. C., Xu, R., Sarma, P., Alhutali, A. H., & Reynolds,

A. C. (2010). Results of the Brugge benchmark study for flooding optimization and history matching. *SPE Reservoir Evaluation & Engineering, 13*(03), 391–405. https://doi.org/10.2118/119094-PA

Pintea, S. L., Sharma, S., Vossepoel, F. C., Van Gemert, J. C., Loog, M., & Verschuur, D. J. (2022). Seismic inversion with deep learning: A proposal for litho-type classification. *Computational Geosciences, 26*(2), 351–364. https://doi.org/10.1007/s10596-021-10118-2

Portier, S., & Rochelle, C. (2005). Modelling CO_2 solubility in pure water and NaCl-type waters from 0 to 300 °C and from 1 to 300 bar. *Chemical Geology, 217*(3-4), 187–199. https://doi.org/10.1016/j.chemgeo.2004.12.007

Pour, K. M., Voskov, D., & Bruhn, D. (2023). Nonlinear solver based on trust region approximation for CO_2 utilization and storage in subsurface reservoir. *Geoenergy Science and Engineering, 225*, 211698. https://doi.org/10.1016/j.geoen.2023.211698

Pyrcz, M., & Deutsch, C. V. (2014). *Geostatistical reservoir modeling* (Second edition). New York, New York : Oxford University Press.

Pyrcz, M., Boisvert, J., & Deutsch, C. (2009). ALLUVSIM: A program for event-based stochastic modeling of fluvial depositional systems. *Computers & Geosciences, 35*(8), 1671–1685. https://doi.org/10.1016/j.cageo.2008.09.012

Qi, R., Laforce, T., & Blunt, M. (2009). Design of carbon dioxide storage in aquifers. *International Journal of Greenhouse Gas Control, 3*(2), 195–205. https://doi.org/10.1016/j.ijggc.2008.08.004

Quarteroni, A., Manzoni, A., & Negri, F. (2016). *Reduced basis methods for partial differential equations* (Vol. 92). Springer International Publishing. https://doi.org/10.1007/978-3-319-15431-2

R. H. Brooks And A. T. Corey. (1964). Hydraulic properties of porous media and their relation to drainage design. *Transactions of the ASAE, 7*(1), 0026–0028. https://doi.org/10.13031/2013.40684

Rackley, S. A. (2017). Overview of carbon capture and storage. In *Carbon Capture and Storage* (pp. 23–36). Elsevier. https://doi.org/10.1016/B978-0-12-812041-5.00002-7

Raziperchikolaee, S., Cotter, Z., & Gupta, N. (2021). Assessing mechanical response of CO_2 storage into a depleted carbonate reef using a site-scale geomechanical model calibrated with field tests and InSAR monitoring data. *Journal of Natural Gas Science and Engineering, 86*, 103744. https://doi.org/10.1016/j.jngse.2020.103744

Remy, N., Boucher, A., & Wu, J. (2009). *Applied geostatistics with SGeMS: A user's guide*. Cambridge University Press.

Rezk, M. G., Foroozesh, J., Abdulrahman, A., & Gholinezhad, J. (2022). CO_2 diffusion and dispersion in porous media: Review of advances in experimental measurements and mathematical models. *Energy & Fuels, 36*(1), 133–155. https://doi.org/10.1021/acs.energyfuels.1c03552

Ringrose, P. (2020). *How to store CO_2 underground: Insights from early-mover CCS projects*. Springer International Publishing. https://doi.org/10.1007/978-3-030-33113-9

Ringrose, P., Mathieson, A., Wright, I., Selama, F., Hansen, O., Bissell, R., Saoula, N., & Midgley, J. (2013). The In Salah CO_2 storage project: Lessons learned and knowledge transfer. *Energy Procedia*, *37*, 6226–6236. https://doi.org/10.1016/j.egypro.2013.06.551

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241, Vol. 9351). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Rucci, A., Ferretti, A., Monti Guarnieri, A., & Rocca, F. (2012). Sentinel 1 SAR interferometry applications: The outlook for sub millimeter measurements. *Remote Sensing of Environment*, *120*, 156–163. https://doi.org/10.1016/j.rse.2011.09.030

Rukhin, A. L. (2007). Estimating common vector parameters in interlaboratory studies. *Journal of Multivariate Analysis*, *98*(3), 435–454. https://doi.org/10.1016/j.jmva.2006.09.005

Rutqvist, J. (2012). The geomechanics of CO_2 storage in deep sedimentary formations. *Geotechnical and Geological Engineering*, *30*(3), 525–551. https://doi.org/10.1007/s10706-011-9491-0

S. Oliver, D., He, N., & C. Reynolds, A. (1996). Conditioning permeability fields to pressure data. *ECMOR V - 5th European Conference on the Mathematics of Oil Recovery*. https://doi.org/10.3997/2214-4609.201406884

Saifullin, I., Seabra, G. S., Pluymakers, A., Vossepoel, F. C., & Voskov, D. (2024). Integrating geomechanical proxy models with data assimilation for energy transition applications. *Geomechanics for Energy and the Environment*, *40*, 100618. https://doi.org/10.1016/j.gete.2024.100618

Schlumberger. (2023). ECLIPSE Reservoir Simulator [https://www.slb.com/reservoir-characterization-and-simulation/software/eclipse, [Accessed 30 August 2023]].

Schütz, M. K., Lopes, N., Cenci, A., Ligabue, R., Dullius, J., Einloft, S., & Ketzer, J. M. (2011). Effect of time on the carbonation reaction of saline aquifers with controlled pH. *Energy Procedia*, *4*, 4546–4551. https://doi.org/10.1016/j.egypro.2011.02.412

Seabra, G. S., De Hoop, S., Voskov, D., & Vossepoel, F. (2022). Improved understanding of naturally fractured reservoirs using data assimilation. *EAGE GET 2022*, 1–5. https://doi.org/10.3997/2214-4609.202221095

Seabra, G. S., Barbosa Machado, M. V., Delshad, M., Sepehrnoori, K., Voskov, D., & Vossepoel, F. C. (2024a). Unveiling valuable geomechanical monitoring insights: Exploring ground deformation in geological carbon storage. *Applied Sciences*, *14*(10), 4069. https://doi.org/10.3390/app14104069

Seabra, G. S., Mücke, N. T., Silva, V. L. S., Voskov, D., & Vossepoel, F. C. (2024b). AI enhanced data assimilation and uncertainty quantification applied to Geological Carbon Storage. *International Journal of Greenhouse Gas Control*, *136*, 104190. https://doi.org/10.1016/j.ijggc.2024.104190

Segall, P. (1992). Induced stresses due to fluid extraction from axisymmetric reservoirs. *pure and applied geophysics*, *139*(3-4), 535–560. https://doi.org/10.1007/BF00879950

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shchipanov, A., Kollbotn, L., Encinas, M., Fjelde, I., & Berenblyum, R. (2022). Periodic CO_2 injection for improved storage capacity and pressure management under intermittent CO_2 supply. *Energies*, *15*(2), 566. https://doi.org/10.3390/en15020566

Shi, J.-Q., Durucan, S., Korre, A., Ringrose, P., & Mathieson, A. (2019). History matching and pressure analysis with stress-dependent permeability using the In Salah CO_2 storage case study. *International Journal of Greenhouse Gas Control*, *91*, 102844. https://doi.org/10.1016/j.ijggc.2019.102844

Shi, J.-Q., Smith, J., Durucan, S., & Korre, A. (2013). A coupled reservoir simulation-geomechanical modelling study of the CO_2 injection-induced ground surface uplift observed at Krechba, in Salah. *Energy Procedia*, *37*, 3719–3726. https://doi.org/10.1016/j.egypro.2013.06.266

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, *81*, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Silva, V. L. S., Seabra, G. S., & Emerick, A. A. (2025a). Mitigating loss of variance in ensemble data assimilation: Machine learning-based and distance-free localization. https://doi.org/10.48550/ARXIV.2506.13362

Silva, V. L. S., Seabra, G. S., & Emerick, A. A. (2025b). Machine learning to enhance the covariance estimations of non-local model parameters in ensemble based-data assimilation. *SPE Reservoir Simulation Conference*, D031S012R004. https://doi.org/10.2118/223908-MS

Silva, V. L., Heaney, C. E., Nenov, N., & Pain, C. C. (2024). Generative model-based framework for parameter estimation and uncertainty quantification applied to a compartmental model in epidemiology. *Journal of Computational Science*, *83*, 102451. https://doi.org/10.1016/j.jocs.2024.102451

Singh, M., Chaudhuri, A., Chu, S. P., Stauffer, P. H., & Pawar, R. J. (2019). Analysis of evolving capillary transition, gravitational fingering, and dissolution trapping of CO_2 in deep saline aquifers during continuous injection of supercritical CO_2. *International Journal of Greenhouse Gas Control*, *82*, 281–297. https://doi.org/10.1016/j.ijggc.2019.01.014

Snippe, J., Berg, S., Ganga, K., Brussee, N., & Gdanski, R. (2020). Experimental and numerical investigation of wormholing during CO_2 storage and water alternating gas injection. *International Journal of Greenhouse Gas Control*, *94*, 102901. https://doi.org/10.1016/j.ijggc.2019.102901

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. *Proceedings of the 32nd International Conference on Machine Learning*, 2256–2265. Retrieved October 7, 2025, from https://proceedings.mlr.press/v37/sohl-dickstein15.html

Song, Y., & Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. https://doi.org/10.48550/ARXIV.1907.05600

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020). Score-based generative modeling through stochastic differential equations. https://doi.org/10.48550/ARXIV.2011.13456

Song, Y., Jun, S., Na, Y., Kim, K., Jang, Y., & Wang, J. (2023). Geomechanical challenges during geological CO_2 storage: A review. *Chemical Engineering Journal*, *456*, 140968. https://doi.org/10.1016/j.cej.2022.140968

Spiteri, E. J., Juanes, R., Blunt, M. J., & Orr, F. M. (2005). Relative permeability hysteresis: Trapping models and application to geological CO_2 sequestration. *SPE Annual Technical Conference and Exhibition*, SPE–96448–MS. https://doi.org/10.2118/96448-MS

Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, *34*(1), 1–21. https://doi.org/10.1023/A:1014009426274

Taccari, M. L., Nuttall, J., Chen, X., Wang, H., Minnema, B., & Jimack, P. K. (2022). Attention U-Net as a surrogate model for groundwater prediction. *Advances in Water Resources*, *163*, 104169. https://doi.org/10.1016/j.advwatres.2022.104169

Tadjer, A., & Bratvold, R. B. (2021). Managing uncertainty in geological CO_2 storage using Bayesian evidential learning. *Energies*, *14*(6), 1557. https://doi.org/10.3390/en14061557

Taheri, A., Torsæter, O., Lindeberg, E., Hadia, N. J., & Wessel-Berg, D. (2021). Effect of convective mixing process on storage of CO_2 in saline aquifers with layered permeability. *Advances in Chemical Research*, *3*(1). https://doi.org/10.21926/acr.2101012

Tang, H., & Durlofsky, L. J. (2022). Use of low-fidelity models with machine-learning error correction for well placement optimization. *Computational Geosciences*, *26*(5), 1189–1206. https://doi.org/10.1007/s10596-022-10153-7

Tang, H., Fu, P., Sherman, C. S., Zhang, J., Ju, X., Hamon, F., Azzolina, N. A., Burton-Kelly, M., & Morris, J. P. (2021). A deep learning-accelerated data assimilation and forecasting workflow for commercial-scale geologic carbon storage. *International Journal of Greenhouse Gas Control*, *112*, 103488. https://doi.org/10.1016/j.ijggc.2021.103488

Tang, M., Ju, X., & Durlofsky, L. J. (2022, July). Deep-learning-based coupled flow-geomechanics surrogate model for CO_2 sequestration. https://doi.org/10.1016/j.ijggc.2022.103692

Tarantola, A. (2005, January). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial; Applied Mathematics. https://doi.org/10.1137/1.9780898717921

Tarrahi, M., Jafarpour, B., & Ghassemi, A. (2015). Integration of microseismic monitoring data into coupled flow and geomechanical models with ensemble Kalman filter. *Water Resources Research*, *51*(7), 5177–5197. https://doi.org/10.1002/2014WR016264

Tempone, P., Fjær, E., & Landrø, M. (2010). Improved solution of displacements due to a compacting reservoir over a rigid basement. *Applied Mathematical Modelling*, *34*(11), 3352–3362. https://doi.org/10.1016/j.apm.2010.02.025

Teng, W., & Durlofsky, L. J. (2025). Likelihood-free inference and hierarchical data assimilation for geological carbon storage. *Advances in Water Resources*, *201*, 104961. https://doi.org/10.1016/j.advwatres.2025.104961

Thomas, F., Livio, F. A., Ferrario, F., Pizza, M., & Chalaturnyk, R. (2024). A review of subsidence monitoring techniques in offshore environments. *Sensors*, *24*(13), 4164. https://doi.org/10.3390/s24134164

Tian, X., Wapperom, M., Gunning, J., Jackson, S., Wilkins, A., Green, C., Ennis-King, J., & Voskov, D. (2024). A history matching study for the FluidFlower benchmark project. *Transport in Porous Media*, *151*(5), 1113–1139. https://doi.org/10.1007/s11242-023-02048-7

Tran, D., Settari, A., & Nghiem, L. (2002). New iterative coupling between a reservoir simulator and a geomechanics module. *SPE/ISRM Rock Mechanics Conference*, SPE–78192–MS. https://doi.org/10.2118/78192-MS

van Wees, J.-D., Pluymaekers, M., Osinga, S., Fokker, P., Van Thienen-Visser, K., Orlic, B., Wassing, B., Hegen, D., & Candela, T. (2019). 3-D mechanical analysis of complex reservoirs: A novel mesh-free approach. *Geophysical Journal International*, *219*(2), 1118–1130. https://doi.org/10.1093/gji/ggz352

Vasco, D. W., Rutqvist, J., Ferretti, A., Rucci, A., Bellotti, F., Dobson, P., Oldenburg, C., Garcia, J., Walters, M., & Hartline, C. (2013). Monitoring deformation at the Geysers Geothermal Field, California using C-band and X-band interferometric synthetic aperture radar. *Geophysical Research Letters*, *40*(11), 2567–2572. https://doi.org/10.1002/grl.50314

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett, Eds.). *30*. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Verma, Y., Vishal, V., & Ranjith, P. G. (2021). Sensitivity analysis of geomechanical constraints in CO_2 storage to screen potential sites in deep saline aquifers. *Frontiers in Climate*, *3*, 720959. https://doi.org/10.3389/fclim.2021.720959

Viana, F. A. C., Gogu, C., & Goel, T. (2021). Surrogate modeling: Tricks that endured the test of time and some recent developments. *Structural and Multidisciplinary Optimization*, *64*(5), 2881–2908. https://doi.org/10.1007/s00158-021-03001-2

Vilarrasa, V., Makhnenko, R. Y., & Rutqvist, J. (2019). Field and laboratory studies of geomechanical response to the injection of CO_2. In *Science of Carbon Storage in Deep Saline Formations* (pp. 209–236). Elsevier. https://doi.org/10.1016/B978-0-12-812752-0.00009-5

Vincent, P. (2011). A connection between score matching and denoising autoencoders. *Neural Computation*, *23*(7), 1661–1674. https://doi.org/10.1162/NECO_a_00142

Voskov, D., Saifullin, I., Novikov, A., Wapperom, M., Orozco, L., Seabra, G. S., Chen, Y., Khait, M., Lyu, X., Tian, X., De Hoop, S., & Palha, A. (2024). Open Delft Advanced Research Terra Simulator(open-DARTS). *Journal of Open Source Software*, *9*(99), 6737. https://doi.org/10.21105/joss.06737

Voskov, D. V. (2017). Operator-based linearization approach for modeling of multiphase multi-component flow in porous media. *Journal of Computational Physics*, *337*, 275–288. https://doi.org/10.1016/j.jcp.2017.02.041

Vossepoel, F. C., Evensen, G., & Van Leeuwen, P. J. (2025). Adaptive correlation- and distance-based localization for iterative ensemble smoothers in a coupled nonlinear multiscale model. *Monthly Weather Review*. https://doi.org/10.1175/MWR-D-24-0269.1

Wang, G. Q. (2013). Millimeter-accuracy GPS landslide monitoring using Precise Point Positioning with Single Receiver Phase Ambiguity (PPP-SRPA) resolution: A case study in Puerto Rico. *Journal of Geodetic Science*, *3*(1). https://doi.org/10.2478/jogs-2013-0001

Wang, S., Li, B. Z., Khabsa, M., Fang, H., & Ma, H. (2020a). Linformer: Self-attention with linear complexity. https://doi.org/10.48550/ARXIV.2006.04768

Wang, Y., Voskov, D., Khait, M., & Bruhn, D. (2020b). An efficient numerical simulator for geothermal simulation: A benchmark study. *Applied Energy*, *264*, 114693. https://doi.org/10.1016/j.apenergy.2020.114693

Wapperom, M., Lyu, X., Nichita, D. V., & Voskov, D. (2023). A unified thermal-reactive compositional simulation framework for modeling CO_2 sequestration at various scales. *SPE Reservoir Simulation Conference*, D021S010R005. https://doi.org/10.2118/212182-MS

Wapperom, M., Tian, X., Novikov, A., & Voskov, D. (2024). FluidFlower benchmark: Lessons learned from the perspective of subsurface simulation. *Transport in Porous Media*, *151*(5), 1033–1052. https://doi.org/10.1007/s11242-023-01984-8

Wen, G., Li, Z., Azizzadenesheli, K., Anandkumar, A., & Benson, S. M. (2022). U-FNO—an enhanced Fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, *163*, 104180. https://doi.org/10.1016/j.advwatres.2022.104180

Wen, G., Tang, M., & Benson, S. M. (2021). Towards a predictor for CO_2 plume migration using deep neural networks. *International Journal of Greenhouse Gas Control*, *105*, 103223. https://doi.org/10.1016/j.ijggc.2020.103223

Werthmüller, D., Serrao Seabra, G., & Vossepoel, F. (2025, April). Dageo: Data assimilation in geosciences. https://doi.org/10.31223/X55X55

White, J. A., & Foxall, W. (2016). Assessing induced seismicity risk at CO_2 storage projects: Recent progress and remaining challenges. *International Journal of Greenhouse Gas Control*, *49*, 413–424. https://doi.org/10.1016/j.ijggc.2016.03.021

Wikner, A., Pathak, J., Hunt, B. R., Szunyogh, I., Girvan, M., & Ott, E. (2021). Using data assimilation to train a hybrid forecast system that combines machine-learning and knowledge-based components. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *31*(5), 053114. https://doi.org/10.1063/5.0048050

Witte, P. A., Konuk, T., Skjetne, E., & Chandra, R. (2023). Fast CO_2 saturation simulations on large-scale geomodels with artificial intelligence-based Wavelet Neural Operators. *International Journal of Greenhouse Gas Control*, *126*, 103880. https://doi.org/10.1016/j.ijggc.2023.103880

Witter, J. B., Trainor-Guitton, W. J., & Siler, D. L. (2019). Uncertainty and risk evaluation during the exploration stage of geothermal development: A review. *Geothermics*, *78*, 233–242. https://doi.org/10.1016/j.geothermics.2018.12.011

Wouters, M., Hanssen, R. F., & Govers, R. (2021). Using PS-InSAR observations to detect aseismic fault slip in the seismically active Groningen gas field. *Earth and Space Science Open Archive*. https://doi.org/10.1002/essoar.10509685.2

Xu, P., Zhu, X., & Clifton, D. A. (2023). Multimodal learning with transformers: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(10), 12113–12132. https://doi.org/10.1109/TPAMI.2023.3275156

Xu, T., Yue, G., Wang, F., & Liu, N. (2014). Using natural CO_2 reservoir to constrain geochemical models for CO_2 geological sequestration. *Applied Geochemistry*, *43*, 22–34. https://doi.org/10.1016/j.apgeochem.2014.01.009

Zhang, C., Song, X., & Azevedo, L. (2021). U-net generative adversarial network for subsurface facies modeling. *Computational Geosciences*, *25*(1), 553–573. https://doi.org/10.1007/s10596-020-10027-w

Zhang, T., Zhang, W., Yang, R., Gao, H., & Cao, D. (2022). Analysis of available conditions for InSAR surface deformation monitoring in CCS projects. *Energies*, *15*(2), 672. https://doi.org/10.3390/en15020672

Zheng, F., Jha, B., & Jafarpour, B. (2023). Controlled CO_2 injection into storage reservoirs to minimize geomechanical risks under geologic uncertainty. *SPE Annual Technical Conference and Exhibition*, D021S020R005. https://doi.org/10.2118/215155-MS

Ziegel, E. R. (1995). *Review of GSLIB: Geostatistical software library and user's guide* (Vol. 37). https://doi.org/10.2307/1269177

Zoback, M. D., & Gorelick, S. M. (2012). Earthquake triggering and large-scale geologic storage of carbon dioxide. *Proceedings of the National Academy of Sciences*, *109*(26), 10164–10168. https://doi.org/10.1073/pnas.1202473109

CURRICULUM VITÆ

Gabriel SERRAO SEABRA

1990	Born in Nova Friburgo, Rio de Janeiro, Brazil.

EDUCATION

2008–2011	B.Sc. in Mechanical Engineering (Oil & Gas emphasis) State University of Rio de Janeiro (UERJ/IPRJ), Brazil – *Magna Cum Laude*
2015–2016	M.Sc. in Mechanical Engineering Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil
2022–Present	Ph.D. Department of Geoscience and Engineering Delft University of Technology (TU Delft), The Netherlands *Thesis:* Improving Uncertainty Quantification of Geological Carbon Storage with Data Assimilation *Promotors:* Prof. dr. ir. F.C. Vossepoel and Dr. D.Voskov

AWARDS

2022	OTC Emerging Leaders Award (Offshore Technology Conference – First Brazilian awardee)

ACKNOWLEDGEMENTS

This thesis marks the end of one journey and the beginning of many others. I would like to express my deepest gratitude to those who walked alongside me and made this work possible.

To my wife, Rina: thank you for coming all the way from Brazil to embrace this journey with me. Your love and partnership gave me strength every day. You left important things behind so we could build, brick by brick, our future together. Your unwavering support, patience, and faith in me made this thesis possible. This achievement is ours.

To my family and friends: thank you for the encouragement across distance and time zones, for every message, call, and visit, and for the constant reminders of what truly matters. To my parents and grandparents, who taught me resilience and curiosity, and to my close friends who kept me grounded and brought joy along the way—your support and belief carried me through challenging moments.

To my colleagues at TU Delft: I am grateful for your camaraderie, insightful discussions, and the collaborative spirit that made research both rigorous and enjoyable. Thank you for sharing ideas, data, coffee, code, and laughter. Working with such a talented and generous community has been a privilege.

I owe special thanks to my supervisors, Prof. Femke C. Vossepoel and Prof. Denis Voskov. Your guidance, trust, and understanding were invaluable. You gave me the freedom to explore, the clarity to focus, and the rigor to refine. Thank you for your steady support, high standards, and for challenging me to become a better researcher.

I also thank Petroleo Brasileiro S.A. (Petrobras) for believing in this work and sponsoring this project. Your support enabled the research presented in this thesis and its contributions to subsurface characterization and carbon storage.

Finally, to everyone who contributed in ways big and small—reviewing drafts, debating equations, debugging code, or simply listening—thank you. I am deeply grateful.

List of Publications

Peer-Reviewed Book Chapter

1. **Gabriel Serrão Seabra**, Ilshat Saifullin, Denis Voskov, Femke C. Vossepoel, *Geomechanical Proxy Models and Data Assimilation: Evaluating Monitoring Strategies in Geological Carbon Sequestration.* In: *Subsurface Data Assimilation: Theory and Applications*, edited by Xiaodong Luo, Olwijn Leeuwenburgh, and Alexandre Anoze Emerick. Elsevier, ISBN-13: 978-0443415449, ASIN: B0F1XF378L, expected publication date: 1 Aug. 2026.

Peer-Reviewed Journal Articles

5. **Gabriel Serrão Seabra**, Nikolaj T. Mücke, Vinícius L. S. Silva, Denis Voskov, Femke C. Vossepoel, *AI-Enhanced Data Assimilation and Uncertainty Quantification Applied to Geological Carbon Storage, International Journal of Greenhouse Gas Control* 136 (2024) 104190. doi:10.1016/j.ijggc.2024.104190

4. **Gabriel Serrão Seabra**, Marcos V. B. Machado, Mojdeh Delshad, Kamy Sepehrnoori, Denis Voskov, Femke C. Vossepoel, *Unveiling Valuable Geomechanical Monitoring Insights: Exploring Ground Deformation in Geological Carbon Storage, Applied Sciences* 14 (10), 4069 (2024). doi:10.3390/app14104069

3. Ilshat Saifullin, **Gabriel Serrão Seabra**, Anne Pluymakers, Femke C. Vossepoel, *Integrating Geomechanical Proxy Models with Data Assimilation for Energy Transition Applications, Geomechanics for Energy and the Environment* 40 (2024) 100618. doi:10.1016/j.gete.2024.100618 *(shared first authorship)*

2. João P. Pereira Nunes, **Gabriel Serrão Seabra**, Luis C. de Sousa Jr., *A Review of CO_2-Injection Projects in the Brazilian Pre-Salt – Storage Capacity and Geomechanical Constraints, International Journal of Greenhouse Gas Control* 137 (2024) 104232. doi:10.1016/j.ijggc.2024.104232

1. Denis Voskov, Ilshat Saifullin, Aleksei Novikov, Michiel Wapperom, Luisa Orozco, **Gabriel Serrão Seabra**, Yuan Chen, Mark Khait, Xiaocong Lyu, Xiaoming Tian, Stephan de Hoop, Artur Palha, *open Delft Advanced Research Terra Simulator (open-DARTS), Journal of Open Source Software* 9 (99), 6737 (2024). doi:10.21105/joss.06737

Manuscripts Under Review and Preprints

3. **Gabriel Serrão Seabra**, Nikolaj T. Mücke, Vinícius L. S. Silva, Alexandre A. Emerick, Denis Voskov, Femke C. Vossepoel, *Integrating Score-Based Diffusion Models with Machine Learning-Enhanced Localization for Advanced Data Assimilation in Geological Carbon Storage, International Journal of Greenhouse Gas Control* – preprint (2025).

2. Vinicius L. S. Silva, **Gabriel Serrão Seabra**, Alexandre A. Emerick, *Mitigating Loss of Variance in Ensemble Data Assimilation: Machine Learning-Based Distance-Free Localization, Journal of Computational Physics* – preprint (2025).

1. Dieter Werthmüller, **Gabriel Serrão Seabra**, Femke C. Vossepoel, *dageo: Data Assimilation in Geosciences*, EarthArXiv pre-print (2025), doi:10.5281/zenodo.14869506

CONFERENCE PUBLICATIONS AND PRESENTATIONS

9. **Gabriel Serrão Seabra**, *Leveraging Diffusion Posterior Sampling for Data Assimilation in Geological Carbon Storage Projects*, 20th International EnKF Workshop, Bergen, 2025.

8. **Gabriel Serrão Seabra**, *Machine Learning to Enhance the Covariance Estimations of Non-Local Model Parameters in Ensemble-Based Data Assimilation*, SPE Reservoir Simulation Conference, Galveston, 2025.

7. **Gabriel Serrão Seabra**, Nikolaj T. Mücke, Denis Voskov, Femke C. Vossepoel, *Leveraging Diffusion Posterior Sampling for Data Assimilation in Geological Carbon Storage Projects*, ENUMATH 2025, Heidelberg.

6. **Gabriel Serrão Seabra**, Nikolaj T. Mücke, Vinícius L. S. Silva, Denis Voskov, Femke C. Vossepoel, *Advancing Data Assimilation and Uncertainty Quantification for CO_2 Sequestration through AI-Hybrid Methods, ECMOR 2024*, Bergen.

5. **Gabriel Serrão Seabra**, Nikolaj T. Mücke, Vinícius L. S. Silva, Denis Voskov, Femke C. Vossepoel, *Enhancing Data Assimilation and Uncertainty Analysis in CCUS Projects through AI Integration, ECCOMAS Congress 2024*, Lisbon.

4. **Gabriel Serrão Seabra**, *Use of Gradient and Ensemble-Based Data Assimilation Methods for Quantifying Reservoir Geomechanics Uncertainties*, SIAM, Bergen, 2023.

3. **Gabriel Serrão Seabra**, Stephan de Hoop, Denis Voskov, Femke C. Vossepoel, *Understanding of Naturally Fractured Geothermal Reservoirs Using Data Assimilation*, 48th Stanford Workshop on Geothermal Reservoir Engineering, Stanford, 2023.

2. **Gabriel Serrão Seabra**, Stephan de Hoop, Denis Voskov, Femke C. Vossepoel, *Improved Understanding of Naturally Fractured Reservoirs Using Data Assimilation*, EAGE GET Conference, Rotterdam, 2022.

1. Ilshat Saifullin, **Gabriel Serrão Seabra**, Denis Voskov, *Geomechanical Response in Energy Transition Applications: Assessing Its Role with Data Assimilation*, IS Energy Geotechnics 2023, Delft, The Netherlands.