Spatial Data Assimilation in Geologic CO2 Sequestration

Bailian Chen, Misael M. Morales, Zhiwei Ma, Qinjun Kang, Rajesh Pawar

Los Alamos National Laboratory

Abstract

Uncertainty in geologic formation usually leads to large uncertainty in the predictions of risk-related system properties and/or risk metrics (e.g., CO2 plumes and CO2/brine leakage rates) at a geologic CO2 storage site. Different type of data, e.g., point measurements from monitoring wells and spatial data from 4D seismic surveys, can be leveraged or assimilated to reduce the risk predictions. In this work, we develop a novel framework for spatial data assimilation and risk forecasting. Under the U.S. Department of Energy’s National Risk Assessment Partnership (NRAP), we have developed a framework using an ensemble-based data assimilation approach called ES-MDA-GEO for spatial data assimilation and forecasting. In particular, we took CO2 saturation maps (can be interpreted from 4D seismic surveys) as inputs for spatial data assimilation. Three seismic surveys at years 1, 3, and 5 were considered in this study. Accordingly, three saturation maps were generated for data assimilation. The impact of level of data noise was also investigated in this work. Our results show increased similarity between the updated reservoir models and the “ground-truth” model with the increased number of seismic surveys. Predictive accuracy in CO2 saturation plume increases with the increased number of seismic surveys as well. We also observed that with the increase of the level of data noise from 1% to 10%, the difference between the updated models and the ground truth does not increase significantly. Similar observation was made for the prediction of CO2 plume distribution at the end of post CO2 injection period by increasing the data noise. Our study results also indicate that comparatively spatial data carry more value than point measurements from monitoring wells to reduce the uncertainty in the risk predictions.

Introduction

Data assimilation is also called history matching or inverse modeling in different communities. It has been widely applied to assimilate history data to calibrate model for predictions in petroleum industry, weather forecast, and hydrology community (Oliver et al, 2008; Pu and Kalnay, 2019; Ghorbanidehno, et al., 2020), and has been used extensively in 4D seismic modeling and inversion (Luo et al, 2016, 2020; Oliver, 2022). Data assimilation is also increasingly applied in geologic CO2 sequestration (GCS) community to calibrate the uncertain reservoir parameters (e.g., permeability and porosity) and reduce the uncertainty in predictions (e.g., CO2 plume and CO2/brine leakage risk) (Li et al, 2015). Some of the most relevant and recent research work is introduced below. In general, GCS community deals with higher levels of uncertainty compared to petroleum industry given the lack of appropriate characterization data or financial incentives to collect it.

Under the context of U.S. DOE’s National Risk Assessment Partnership (NRAP), Chen et al. (2020) revealed how uncertainty in predicted risks can be reduced by performing monitoring data assimilation. They developed a workflow based on the ensemble smoother with multiple data assimilation with geometric inflation factors (ES-MDA-GEO) algorithm (Emerick and Reynolds, 2013; Rafiee and Reynolds, 2017) to assimilate the monitoring data into reservoir models and to calibrate models. The updated models were used to predict future risks and reduction in their uncertainties. The effectiveness of this proposed workflow for monitoring data assimilation was demonstrated with multiple examples including a field scale hypothetical case on Rock Springs Uplift storage site in Wyoming, USA. Thereafter, Chen et al. (2022a) developed a novel framework based on iterative risk assessment using data assimilation to effectively quantify the uncertainty reduction in the predicted risk quantities. Their findings indicated that the application of ES-MDA-GEO based data assimilation in conjunction with the NRAP's Open-Source Integrated Assessment Model (NRAP-Open-IAM) was effective in reducing uncertainty in risk-related predictions for GCS.

Later, researchers from U.S. DOE’s Science-informed Machine Learning for Accelerating Real-Time Decisions in Subsurface Applications (SMART) Initiative developed multiple machine learning accelerated data assimilation approach for history matching and uncertainty quantification in CO2 sequestration. Tang et al. (2021) leveraged physics in porous media flow behavior and machine learning technique to develop a rapid data assimilation framework. The backend data assimilation approach is ES-MDA, and a deep learning-based proxy model was developed to replace the full-physics simulations which is required in the inverse modeling. Chen et al. (2022b) developed a deep learning accelerated data assimilation approach in GCS. The major difference of this work from the work of Tang et al. (2021) is that Chen et al. (2022b) applied a feature coarsening technique to reduce the model dimension during the training and prediction, that is, the training and prediction processes were performed at the coarse scale. Thereafter, the resolution was further recovered to the fine scale by a piecewise cubic interpolation. This proposed workflow can easily handle data assimilation for large-scale GCS site. Note that the data used for data assimilation in both studies are point measurements from monitoring and injection wells.

At CO2 storage sites the data from monitoring wells are usually very limited given the extremely limited number of monitoring wells. Thus, it is crucial to identify other types of data which can be used for data assimilation and uncertainty reduction quantification in risk predictions. One such type of data that can be used to improve the accuracy of risk predictions is spatial measurements inferred from 4-dimensional seismic surveys (time-dependent, repeat 3-dimensional surveys). In this study, we extend the workflow developed for monitoring data assimilation to spatial data assimilation of seismic data. Instead of directly using the seismic data in our framework, we use multiple CO2 plume interpretations from 4D seismic surveys as inputs for spatial data assimilation.

To the best of our knowledge, this is the first study to perform spatial data assimilation to reduce the prediction of risk quantities at geologic CO2 storage sites. Our framework is integrated into NRAP open-source Integrated Assessment Model (NRAP-Open-IAM) to support the deployment of carbon capture and storage to meet the net-zero emission target by 2050 in the United States.

Methodology

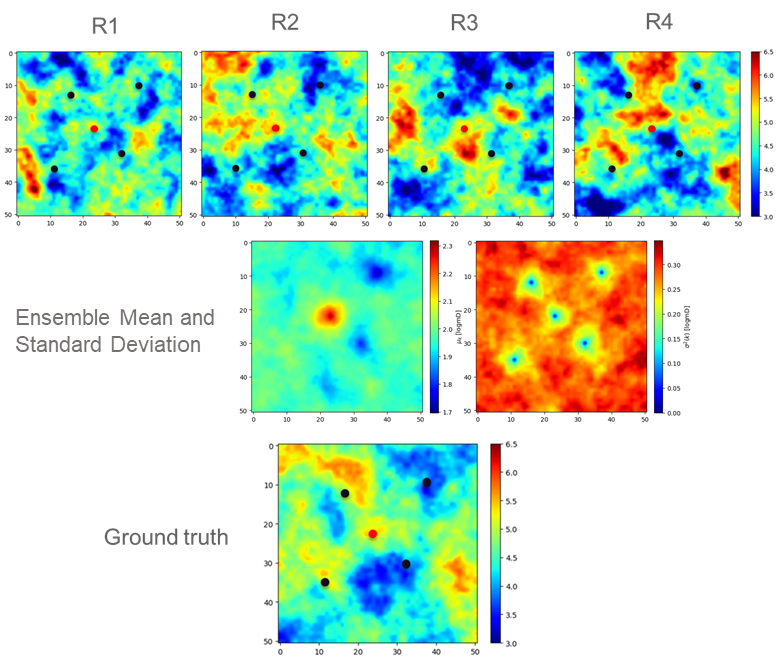
Ensemble-based data assimilation approaches, including the Ensemble Kalman Filter (EnKF), Ensemble Smoother (ES), and Ensemble Smoother with Multiple Data Assimilation (ES-MDA), are the most commonly used techniques for assimilating historical data and updating uncertain subsurface models in various applications, such as oil and gas production, and CO2 storage. ES-MDA, introduced by Emerick and Reynolds, 2013, for history matching in oil and gas production, demonstrated superior performance compared to EnKF for data assimilation. Since then, the ES-MDA algorithm has gained widespread usage for solving history matching or data assimilation problems in diverse communities, especially the oil and gas reservoir modeling community. For instance, several studies (Le et al., 2015; Silva et al., 2017; Soares et al., 2018; Evensen, 2018; Kim et al., 2020; Zhang et al., 2020; Guo et al., 2023) have successfully employed ES-MDA for history matching or data assimilation problems.

Although ES-MDA has been demonstrated to be a robust approach for data assimilation or history matching in subsurface applications, the main drawback of the original version of ES-MDA is that the inflation factors for each assimilation step and the total number of steps for data assimilation must be pre-determined before the start of data assimilation. Researchers proposed different methods to resolve these ES-MDA application issues including adaptive ES-MDA by Le et al. (2016) and ES-MDA with geometric inflation factors (ES-MDA-GEO) by Rafiee and Reynolds (2017). ES-MDA-GEO is recognized as the most effective and practical way to address the issues in the original ES-MDA algorithm. Rafiee and Reynolds (2017) presented a practical approach to determine the precise minimum inflation factor for each data assimilation step. These inflation factors can be obtained through the truncated singular value decomposition (TSVD) of the data sensitivity matrix. By enabling users to set a limit on the total number of data assimilation steps based on available computational resources, this technique offers adequate attenuation of variations in reservoir model realizations during each iteration. This effectively manages overshooting and undershooting, which can otherwise result in crude or imprecise evaluations of uncertain reservoir properties, such as permeability.

In this study, the most advanced and practical version of ES-MDA algorithms – ES-MDA-GEO – is leveraged to perform the spatial data assimilation. The difference from the work of Chen et al. (2020) is that the inferred CO2 saturation maps from 4D seismic surveys at different time intervals are used as history or observational data in this study, while in the previous work the point measurements from monitoring/injection wells were utilized as history data for data assimilation. The details about the implementation of this algorithm for data assimilation in GCS can be found in the work of (Chen et al., 2020).

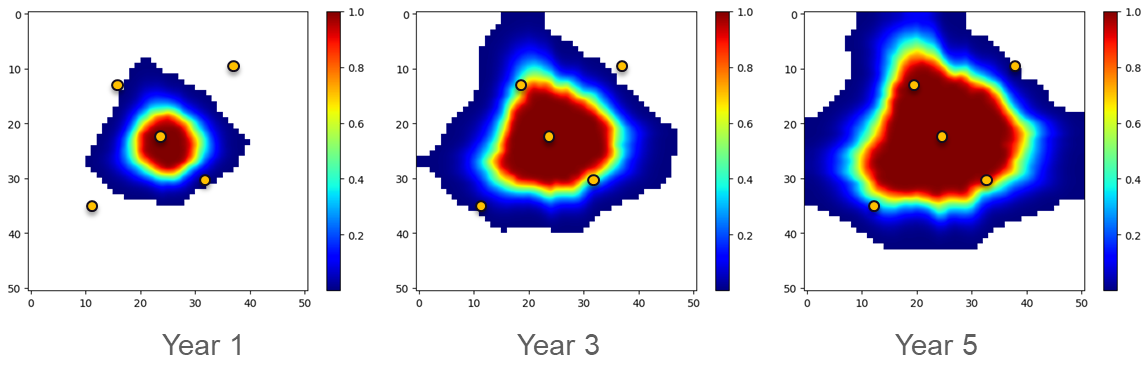
Case Study and Results

A set of synthetic models (Chen et al, 2020) were used to demonstrate the workflow developed for spatial data assimilation in geologic CO2 sequestration. In total, 101 reservoir realizations were used. One of the 101 models was selected as the ground truth model. The remaining 100 models were treated as prior models. The model consists of a mesh of 51 by 51 nodes in the x and y directions, respectively, and 11 layers in the z direction. This model is 4 km in the horizontal direction. The reservoir is 100 m in thickness, and 1 km deep below surface level. Figure 1 shows examples of horizontal log-permeability distributions for the top layer in four model realizations (prior models) and the ground truth model. Porosity is constant at 0.15. Our assumption was that the remaining ten layers possess an equivalent permeability distribution to that of the uppermost layer. 3D multi-phase simulations are performed by the Finite Element Heat and Mass Transfer code (FEHM). The sides of the reservoir are Dirichlet boundaries allowing outflow of CO2 and water pressures above hydrostatic. No inflow is allowed at the boundaries. The model is initialized to a geothermal gradient of 0.03C/m with a temperature of 20C along the top. Pressures are initialized with a 9.81 MPa/m with a pressure of 0.2 MPa along the top. In this study, we consider a five-year CO2 injection period with an injection rate of 0.1 million tons per year, followed by a monitoring period of 10 years.



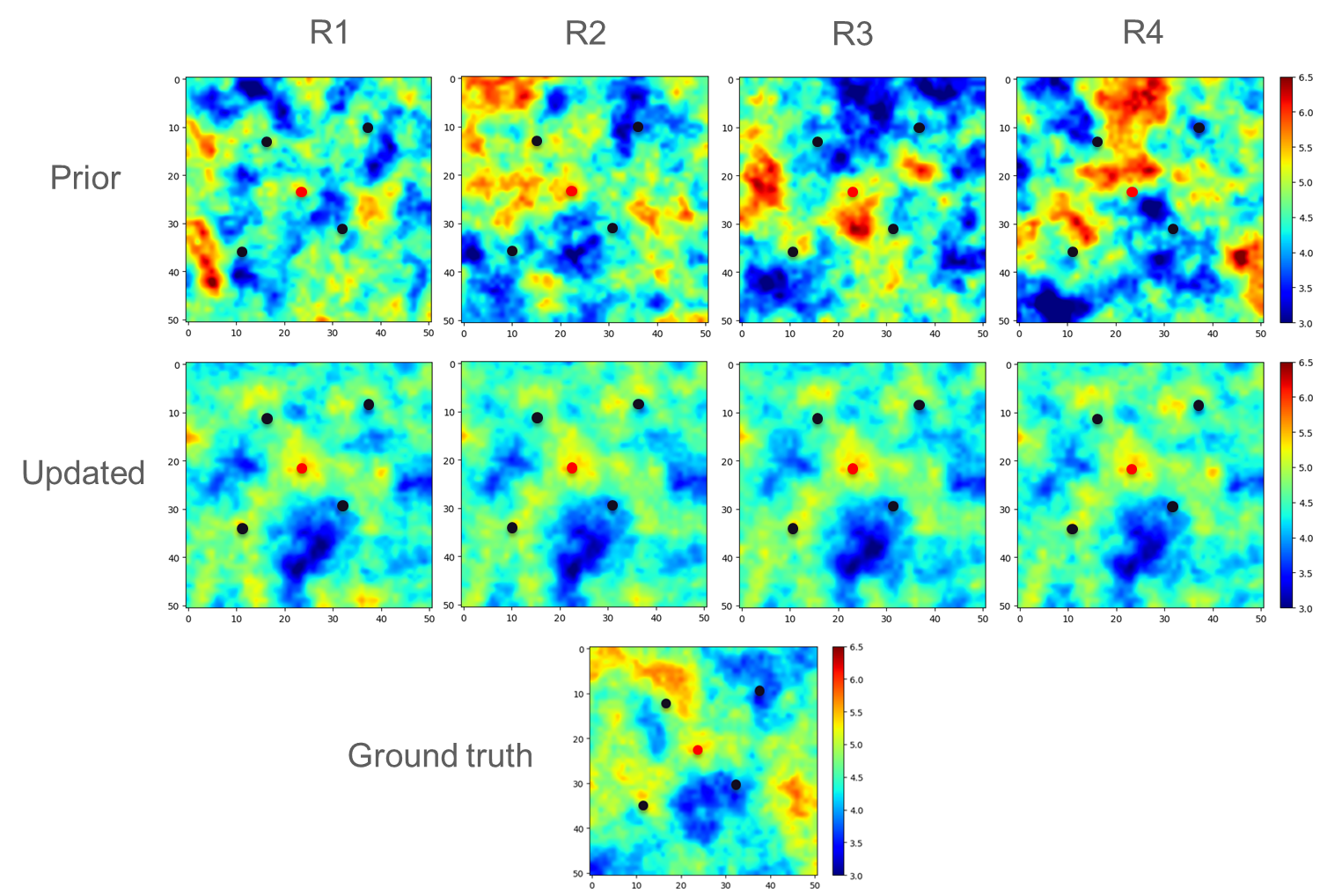
**Figure 1: Log-permeability distribution for the uppermost layer of the ground truth model and four prior models (R1, R2, R3 and R4). The red dot indicates the location of the injection well. The four black dots are the locations of four monitoring wells.**

In the work of Chen et al. (2020), the data (i.e., time series pressure and saturation) collected from the injection well and monitoring wells were used to demonstrate the effectiveness of monitoring data to reduce the uncertainty in the predictions of risk quantities (e.g., CO2 plume area). In this work, we only focus on the assimilation of spatial data (CO2 saturation maps) to reduce the uncertainty in the predictions of risk metrics. The spatial data we considered are the CO2 saturation maps which can be interpreted from 4D seismic surveys. Figure 2 shows the CO2 saturation maps at years 1, 3 and 5. These CO2 saturation maps are sequentially assimilated into the reservoir models using ES-MDA-GEO approach to update permeability distributions. Specifically, we performed three different data assimilation steps using the entire saturation maps at each step. Only the CO2 saturation map at year 1 was utilized in the first data assimilation step, while CO2 saturation maps at years 1 and 3 were utilized in the second data assimilation step and the maps at years 1, 3 and 5 were utilized in the third data assimilation step.

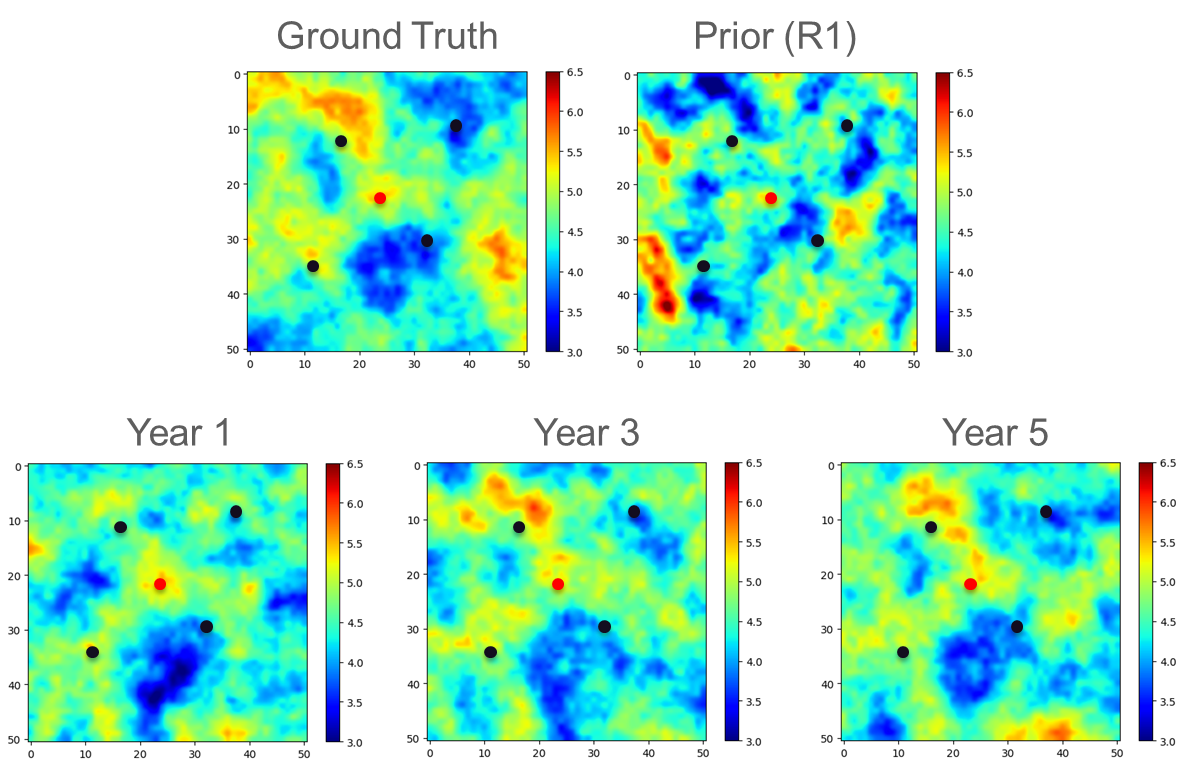


**Figure 2: CO2 saturation maps at years 1, 3, and 5 for the Ground Truth model.**

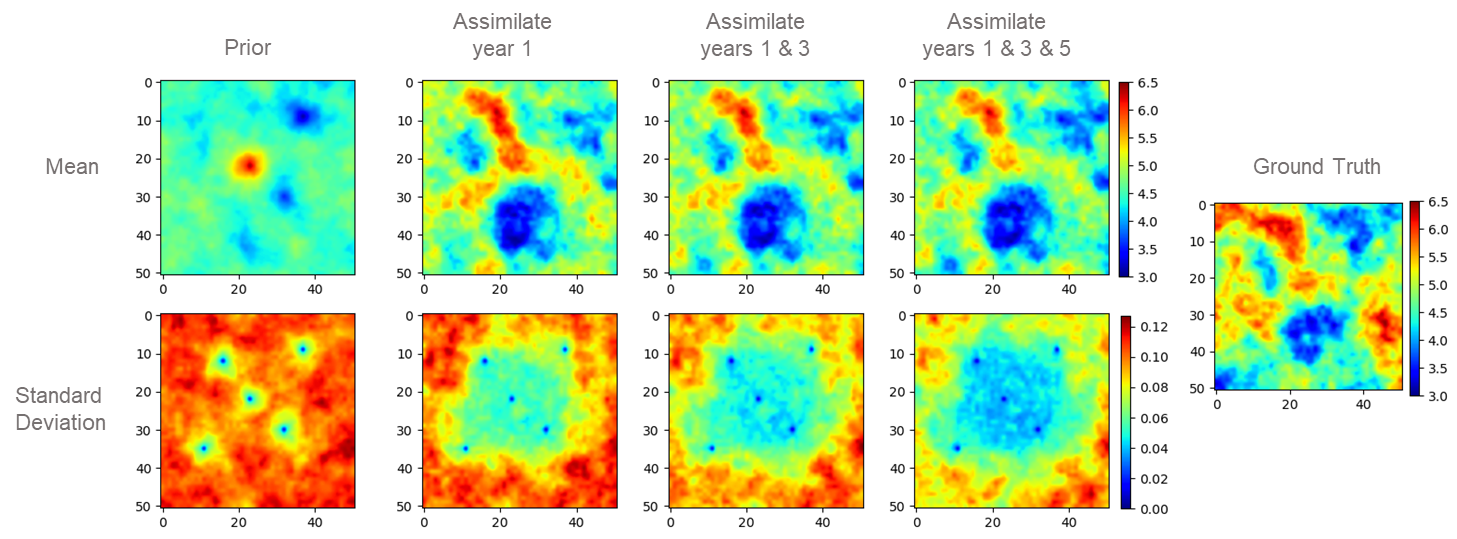
Figure 3 shows the model updating for the first four model realizations after the assimilation of first year CO2 saturation map. As observed from the figure, the permeability fields have been significantly updated compared to fields in the prior models, and the updated models appear closer to the ground truth model, especially the permeability distribution within the CO2 plume area at year 1. Figures 4 and 5 display the evolution of updated permeability field with increasing assimilation of CO2 saturation maps, and the ensemble uncertainty metrics, respectively. As can be observed from Figure 4, with the increased use of CO2 saturation maps (or number of seismic surveys) through the data assimilation process, the updated permeability distribution gets closer to the ground truth model. Figure 5 compares the mean and standard deviation of the ensemble with each assimilation step, showing a convergence to the ground truth model.



**Figure 3: Model update after the assimilation of first year saturation map.**

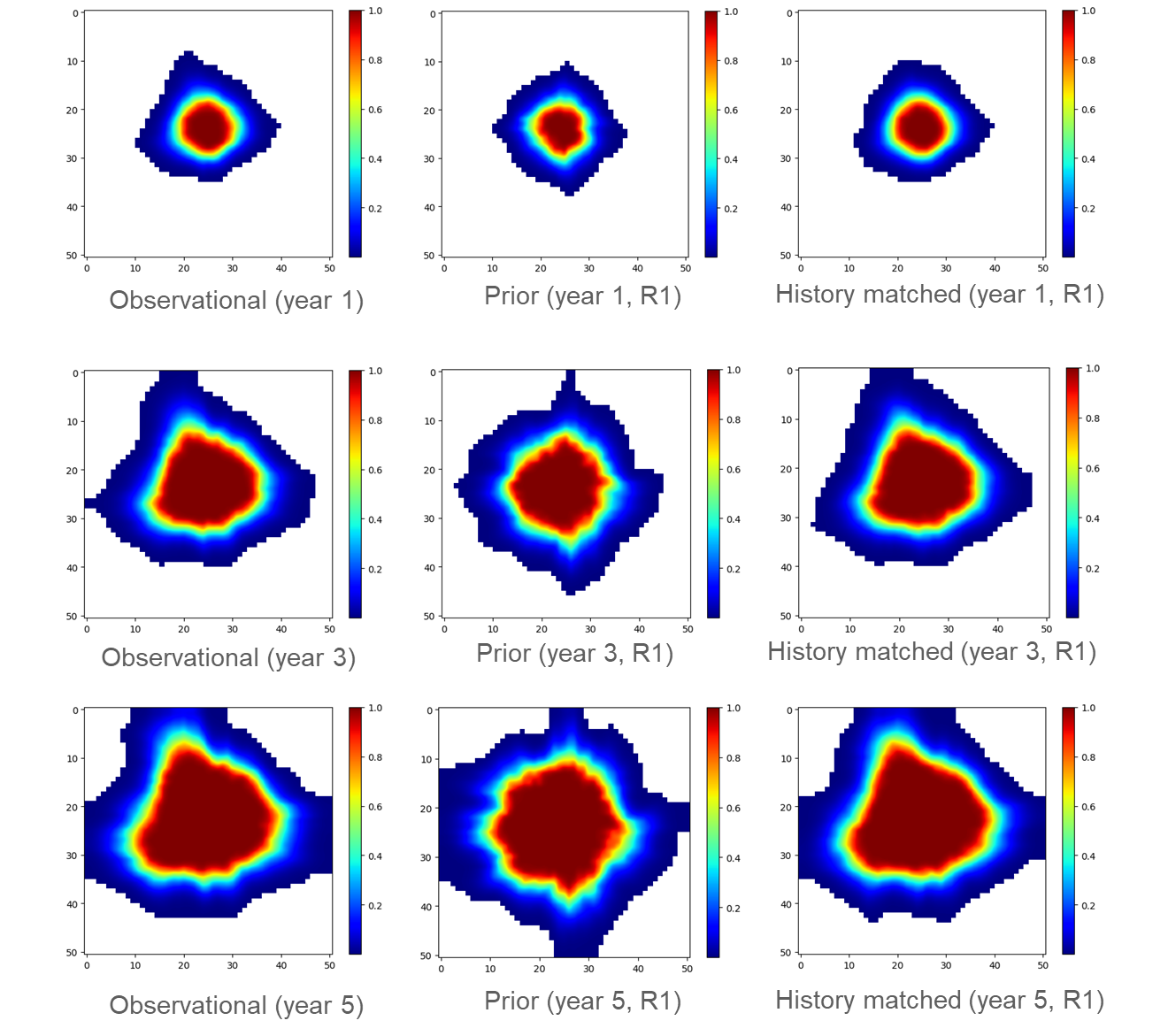


**Figure 4: Evolution of permeability field update.**



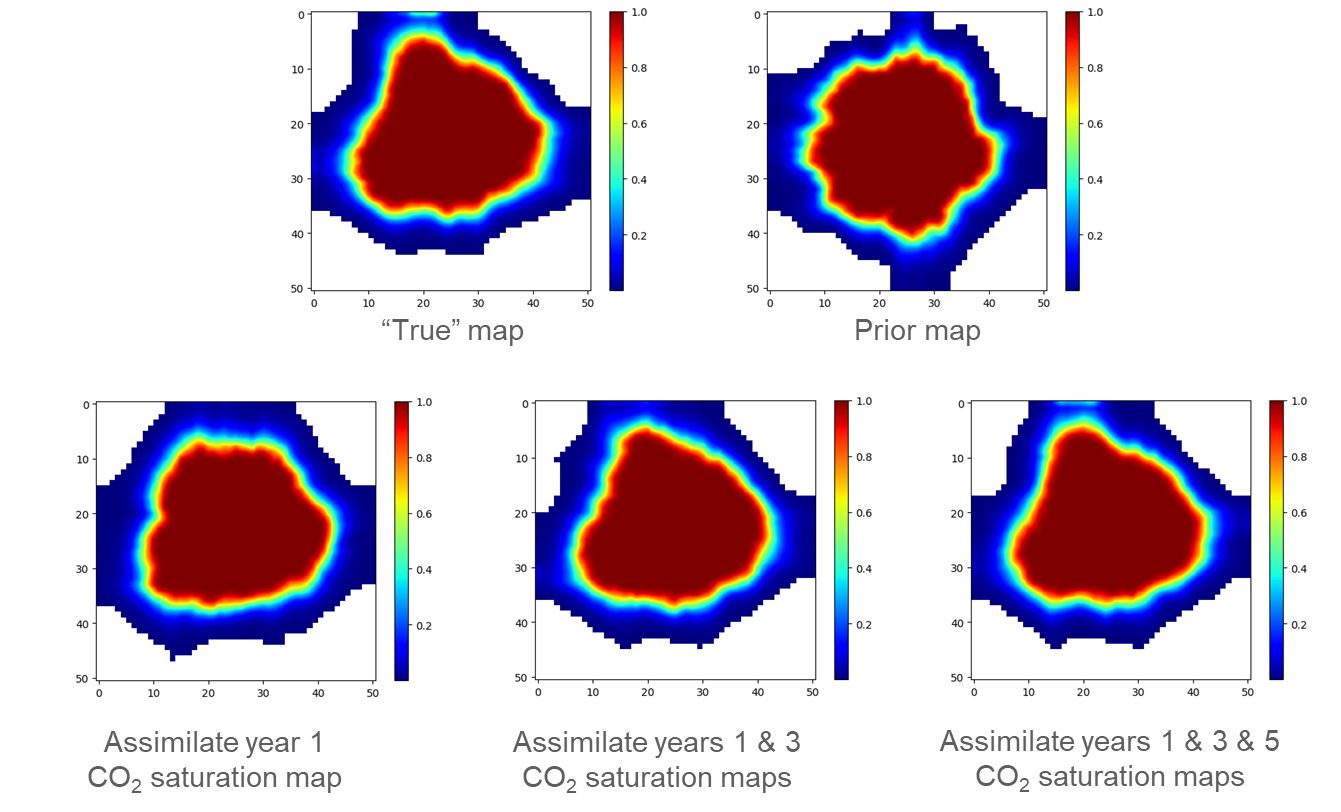
**Figure 5: Uncertainty metrics of permeability field ensemble in each update.**

The history matched CO2 saturation maps are shown in 6. In this figure, the first column of CO2 saturation maps is the observational data used in the spatial data assimilation, while the second and third columns show the predictions of CO2 saturation distributions before and after history matching (or data assimilation). As observed from this figure, the CO2 saturation distributions have been well matched with the observed CO2 saturation distributions.

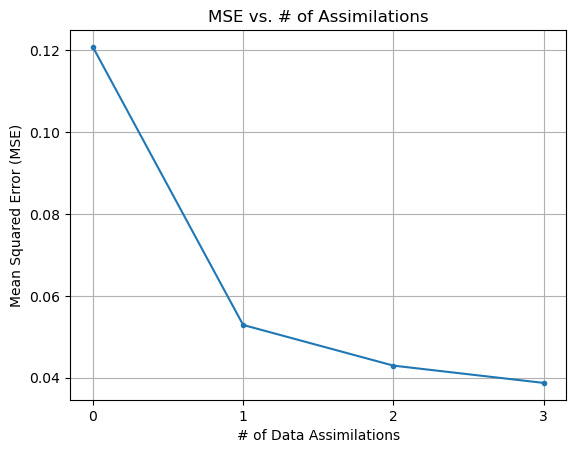


**Figure 6: History matching of CO2 saturation maps. First column of maps shows the observational saturation maps from seismic surveys; the second column of CO2 saturation maps show the saturation prediction based on prior models; the last column of maps shows the history matched CO2 saturation maps.**

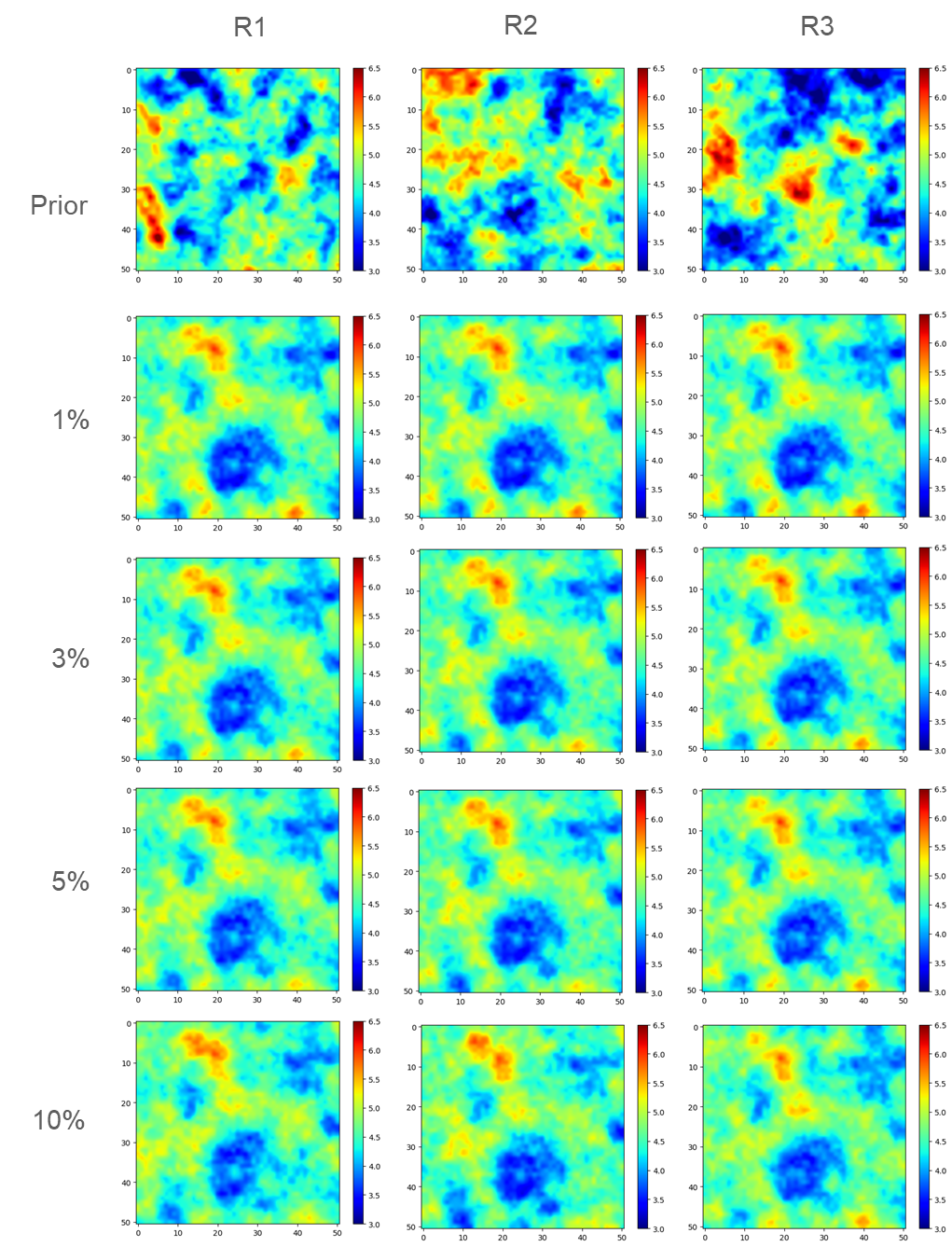
After the spatial data assimilation process, we predict the CO2 plume distribution using the updated reservoir models. The CO2 saturation maps projected at the conclusion of the CO2 storage endeavor (i.e., after 15 years of CO2 injection) are presented in 7. There are five different saturation maps in the figure. The two plots at the top show the “true” saturation map based on the ground truth model and predicted saturation map after 15 years based on the prior model (without data assimilation). The three figures at the bottom of 7 display the predictions of CO2 plume at the end of year 15 for the first realization after three different data assimilation steps, assimilation of year 1 CO2 saturation map, assimilation of years 1 and 3 saturation maps, and assimilation of years 1, 3, and 5 CO2 saturation maps. As can be seen from this figure, the prediction based on prior model is significantly different from the ground truth map. After the assimilation of year 1 spatial data (i.e., CO2 saturation map), the predicted CO2 plume is much closer to the plume predicted using the ground truth model compared to the one predicted using the prior model (without data assimilation). With the increased assimilation of spatial data (i.e., CO2 saturation maps or seismic surveys), the similarity between the predicted plume and the true map increases. Figure 8 shows the evolution of the mean squared error (MSE) with each assimilation step. The average MSE of the ensemble compared to the ground truth model after assimilating years 1, 3, and 5 is 0.0387.



**Figure 7: CO2 plume prediction for R1 at the end of project (year 15).**



**Figure 8: Evolution of the average mean squared error of ensemble per assimilation step.**

Given that there is uncertainty in the CO2 saturation interpretation from seismic data, the saturation maps used in spatial data assimilation will inherently have uncertainty. We investigated the impact of the level of data uncertainty (or noise) in the CO2 saturation maps used in spatial data assimilation. Specifically, we varied the level of data noise from 1% to 10% in the saturation response maps for the ground truth model. 9 shows the comparison of the updated permeability for the first three model realizations under different level of data noise after fully assimilating the noisy data for 1 year, 1 and 3 years, and 1, 3, and 5 years (n=3 assimilation steps). Overall, there is some impact of data noise on spatial data assimilation and the model updating, but the impact is not significant even if we increase the noise level to 10%, which implies that for the level of uncertainty associated with seismic data interpretation used in this study, the uncertainty in the updated permeability distribution is not significant. 

**Figure 9: Impact of the level of data noise on permeability updating.**

Conclusions

In this work, a novel workflow for assimilation of spatial data (e.g., CO2 saturation distribution) was developed and demonstrated for an example CO2 sequestration operation. Our results indicate that the assimilation of spatial data, namely CO2 saturation maps derived from seismic surveys, enhances the resemblance between the predictions generated by updated reservoir models and the "ground-truth" model. Predictive accuracy in CO2 saturation plume increases with the increased number of spatial data. We also observed that for data noise in the range of 1% to 10%, the difference between the updated models and the ground truth does not significantly increase. Compared to our previous work (Chen et al, 2020) based on data assimilation of point measurements from monitoring wells, the spatial data appear to have more value of information than point measurements from monitoring wells to reduce the uncertainty in the prediction of CO2 plume area or distribution.

The future directions of this study can be summarized as follows. First, given the high efficiency of machine learning-based reduced order models (ROMs) for predictions, it would be beneficial to speed up the workflow by introducing ROMs into data assimilation procedures. Second, the feasibility and effectiveness of this proposed workflow can be investigated by using 3D models.

Acknowledgement

This work was supported by the US Department of Energy through the Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). This work was completed as part of the National Risk Assessment Partnership (NRAP) project. Support for this project came from the U.S. Department of Energy’s (DOE) Office of Fossil Energy and Carbon Management.

References

Chen, B., Harp, D. R., Lu, Z., & Pawar, R. J. (2020). Reducing uncertainty in geologic CO2 sequestration risk assessment by assimilating monitoring data. *International Journal of Greenhouse Gas Control*, *94*, 102926.

Chen, B., Harp, D. R., Zhang, Y., Oldenburg, C. M., & Pawar, R. J. (2022). Dynamic risk assessment for geologic CO2 sequestration. *Gondwana Research*.

Chen, B., Yan, B., Kang, Q., Harp, D., & Pawar, R. (2022). Deep Learning Accelerated Inverse Modeling and Forecasting for Large-Scale Geologic CO2 Sequestration. *Available at SSRN 4283252*.

Emerick, A. A., & Reynolds, A. C. (2013). Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, *55*, 3-15.

Evensen, G. (2018). Analysis of iterative ensemble smoothers for solving inverse problems. *Computational Geosciences*, *22*(3), 885-908.

Ghorbanidehno, H., Kokkinaki, A., Lee, J., & Darve, E. (2020). Recent developments in fast and scalable inverse modeling and data assimilation methods in hydrology. *Journal of Hydrology*, *591*, 125266.

Guo, Z., Li, Y., & Sankaran, S. (2023, January). A Reduced-Physics Framework for Forecasting in Unconventional Reservoirs. In *SPE Reservoir Characterisation and Simulation Conference and Exhibition*. OnePetro.

Kim, J., Kim, S., Park, C., & Lee, K. (2020). Construction of prior models for ES-MDA by a deep neural network with a stacked autoencoder for predicting reservoir production. *Journal of Petroleum Science and Engineering*, *187*, 106800.

Le, D. H., Younis, R., & Reynolds, A. C. (2015, February). A history matching procedure for non-Gaussian facies based on ES-MDA. In *SPE Reservoir Simulation Symposium*. OnePetro.

Le, D. H., Emerick, A. A., & Reynolds, A. C. (2016). An adaptive ensemble smoother with multiple data assimilation for assisted history matching. *Spe Journal*, *21*(06), 2195-2207.

Li, J. Y., Kokkinaki, A., Ghorbanidehno H., Darve, E. F., Kitanidis, P. K. (2015). The compressed state Kalman filter for nonlinear state estimation: Application to large-scale reservoir monitoring. *Water Resources Research, 51*. 9942-9963.

Luo, X., Bhakta, T., Jakobsen, M., & Naevdal, G. (2016). An ensemble 4D-seismic history-matching framework with sparse representation based on wavelet multiresolution analysis. *SPE J.* 22 (2017): 985-1010.

Luo, X., Lorentzen, R. J., & Bhakta, T. (2020). Accounting for model errors of rock physics models in 4D seismic history matching problems: A perspective of machine learning. *Journal of Petroleum Science and Engineering, 196,* 107961.

Oliver, D. (2022). Information content in 4D seismic data: Effect of correlated noise. *Journal of Petroleum Science and Engineering, 208*, 109728.

Oliver, D. S., Reynolds, A. C., & Liu, N. (2008). *Inverse theory for petroleum reservoir characterization and history matching*.

Pu, Z., & Kalnay, E. (2019). Numerical weather prediction basics: Models, numerical methods, and data assimilation. *Handbook of hydrometeorological ensemble forecasting*, 67-97.

Rafiee, J., & Reynolds, A. C. (2017). Theoretical and efficient practical procedures for the generation of inflation factors for ES-MDA. *Inverse Problems*, *33*(11), 115003.

Silva, V. L. S., Emerick, A. A., Couto, P., & Alves, J. L. D. (2017). History matching and production optimization under uncertainties–Application of closed-loop reservoir management. *Journal of Petroleum Science and Engineering*, *157*, 860-874.

Soares, R. V., Maschio, C., & Schiozer, D. J. (2018). Applying a localization technique to Kalman Gain and assessing the influence on the variability of models in history matching. *Journal of Petroleum Science and Engineering*, *169*, 110-125.

Tang, H., Fu, P., Sherman, C. S., Zhang, J., Ju, X., Hamon, F., ... & 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.

Zhang, Q., Jiang, S., Wu, X., Wang, Y., & Meng, Q. (2020). Development and Calibration of a Semianalytic Model for Shale Wells with Nonuniform Distribution of Induced Fractures Based on ES-MDA Method. *Energies*, *13*(14), 3718.Zyvolosky, G. FEHM: a control volume finite element code for simulating subsurface multi-phase multi-fluid heat and mass transfer. Los Alamos unclassified Report LA-UR-07-3359