Accurate and Timely Forecasts of Geologic Carbon Storage using Machine Learning Methods

Dan Lu
Computational Sciences and Engineering Division
Oak Ridge National Laboratory
Oak Ridge, TN 37830
`lud1@ornl.gov`

Scott Painter
Environmental Sciences Division
Oak Ridge National Laboratory
`paintersl@ornl.gov`

Nicholas Azzolina
Energy & Environmental Research Center
University of North Dakota
`nazzolina@undeerc.org`

Matthew Burton-Kelly
Energy & Environmental Research Center
University of North Dakota
`mburtonkelly@undeerc.org`

Abstract

Carbon capture and storage is one strategy to reduce greenhouse gas emissions. One approach to storing the captured CO_2 is to inject it into deep saline aquifers. However, dynamics of the injected CO_2 plume is uncertain and the potential for leakage back to the atmosphere must be assessed. Thus, accurate and timely forecasts of CO_2 storage via real-time measurements integration becomes very crucial. This study proposes a learning-based, inverse-free prediction method that can accurately and rapidly forecast CO_2 movement and distribution with uncertainty quantification based on limited simulation and observation data. The machine learning techniques include dimension reduction, multivariate data analysis, and Bayesian learning. The outcome is expected to provide CO_2 storage site operators with an effective tool for real-time decision making.

1   Introduction

Carbon capture and storage (CCS) has been proposed as a strategy to reduce greenhouse gas emissions so as to address the global climate crisis [15, 1]. The Intergovernmental Panel on Climate Change (IPCC) estimated that capturing carbon at a modern conventional power plant could reduce CO_2 emissions to the atmosphere by approximately 80-90% compared to a plant that doesn't have the technology to remove carbon [11]. Once the carbon has been captured, it must be stored. An approach of interest is to inject CO_2 into deep saline aquifers [8]. However, understanding and data are significantly lacking in these aquifers[5]. More importantly, leakage of CO_2 into overlying resource-bearing strata, protected groundwater aquifers, and back into atmosphere could also be a problem for saline aquifer storage [4]. Uncertainty associated with predicting subsurface response to CO_2 injection and storage is a key barrier to developers seeking to secure financing, permits, and social license to inject CO_2 into deep underground [13, 6]. Providing CO_2 storage site operators with forecasting tools for real-time decision making is essential to address these barriers to CCS project development and management. Delivering on this need requires transformational changes in how we predict geologic carbon storage and update those predictions using real-time measurements.

The traditional workflow for predicting CO_2 distribution in a geological reservoir relies heavily on inverse modeling (history matching, model calibration) to constrain uncertain parameters in complex reservoir simulation models [2, 14]. This inversion-based prediction approach has limitations for rapid integration of streaming data and providing real-time decision support due to the following

reasons: (1) Model inversion is computationally expensive and can require thousands of expensive reservoir model simulations, which need to be performed iteratively rather than concurrently and thus cannot take full advantage of contemporary parallel computing resources. (2) Model inversion can be numerically ill-posed resulting in poor predictions when the number of parameters is greater than the measurements, which is usually the case in CO_2 storage simulation. (3) Model inversion needs to be repeated when incorporating new measurements.

To address these challenges, our research aims to develop machine learning (ML) techniques with potential to provide step-change improvements in forecasts relative to the conventional history matching-based forecasts, thus enhancing the timeliness and accuracy of information provided to the operator. This paper describes our methods and analyzes their performance in predicting CO_2 distribution at a commercial scale reservoir. Our project is part of a large initiative called SMART (Science-informed Machine Learning for Accelerating Real Time Decisions in Subsurface Applications) funded by U.S. Department of Energy with the goal to enable better decisions in CO_2 sequestration given scarce resources in a highly uncertain subsurface.

2   Method

We propose a Learning-based Inversion-free Prediction (LIP) framework that produces real-time prediction with uncertainty quantification via integrating observation streams with parallel forward simulations. The key idea of LIP framework is to circumvent the challenging inverse modeling by precomputing an ensemble of unconstrained forward simulations and then using ML methods to learn the relationship between simulated observation and prediction variables. Once the ML model has learned the relationship, it can be used to continually update predictions of future system behavior based on streaming and multiple sources of observations to enable rapid data assimilation and real-time decision support. Specifically, LIP consists of four steps:

1. Generate prior samples of observation and prediction variables by running forward models based on the prior distribution of model parameters;

2. Dimension reduction of the simulated observations and predictions;

3. Establish a statistical relationship between observation and prediction in reduced dimension;

4. Bayesian inference of the prediction based on the statistical model with observation data.

Steps 1-3 correspond to the training stage, where the observation-prediction relationship in the reduced dimension is learned from unconstrained forward simulations. Step 4 corresponds to the prediction stage, where the posterior distribution of the prediction is deduced from the observed data after back transformation to its original high-dimensional space.

In LIP, dimension reduction is important for effective learning of the observation-prediction relationship. In carbon storage simulation, the prediction variables and observation variables are usually spatial images and time series which present spatial and temporal correlations. Reducing these variables' dimensions and learning their relationship in the reduced dimension not only simplify the learning task and thus improve the computational efficiency, but also remove the multicollinearity [7] and therefore enhance the model fitting reliability. Principal component analysis (PCA) usually works effectively for carbon storage simulation data and PCA is a bijective operation so the original high-dimensional variable can be recovered uniquely by undoing the projection. In the reduced dimension, we first use canonical correlation analysis (CCA) [16] to linearize the relationship. If the observation and prediction variables are nearly linearly correlated (e.g., with a correlation coefficient greater than 0.9) in the reduced canonical space, a linear model can be used to simulate their relationship. Next, by assuming a Gaussian likelihood as commonly done in the CCS community [14], a Gaussian regression can be used to infer the prediction and its uncertainty. This analysis is computationally and data efficient, making it particularly suitable for real-time carbon storage forecasting where observation and simulation data are very limited and timely forecasts are needed for decision. If the linear assumption is not satisfied, we can use advanced ML models such as Bayesian neural networks [9] for prediction. More details of the LIP method are presented in Appendix A.

3 Application

A model for a clastic continental shelf [3] was considered as the geological model because such depositional environments provide significant capacity for CO_2 storage. The three-dimensional (3D) simulation domain comprises 30 layers with 211 by 211 cells in each layer, i.e., 1,335,630 cells in total. Each cell has a size of 500 by 500 by 10 feet. One-hundred geomodels were generated to capture porosity and permeability uncertainty. For each realization of the geomodel, CO_2 injection was simulated for 10 years to produce 100 prior samples of CO_2 pressure fields in the 3D domain at 32 time steps (monthly data in first 2 years and annual data in last 8 years). Four injection wells—located regularly at the grid cells [71, 71, 3–30], [141, 71, 3–30], [71, 141, 3–30], and [141, 141, 3–30]—inject CO_2 into the reservoir with a mass injection target of 2 million metric tons/year. Each forward simulation takes several days on an Intel Core CPU, making it really difficult, if possible, to enable conventional inversion-based history matching and forecasting.

In this study, we use the LIP method to predict the spatial distribution of CO_2 pressure in layer 3 (the first layer of injection) after 10 years of injection based on pressure measurements in the four injection wells. We performed five case studies, depending on the duration of the observation period and thus the look-ahead period. Specifically, we forecast pressure distribution in year 10 from the perspective of year 5, 6, 7, 8 and 9, in each case using only the data available up to that time, which corresponds to varying the look-ahead period from 5 years to 1 year. In that latter case, we have 9 years of observations (31 time steps×4 wells=124 observation variables) to predict the CO_2 pressure map at 211 by 211 grid cells (44521 variables in total) in year 10. The other four cases have correspondingly fewer observations. These five case studies are designed to evaluate LIP's accuracy, efficiency, and capacity to incorporate streaming observations to improve prediction. These are challenging applications because of the large uncertain domain, the small number of prior samples and the limited observations. To evaluate LIP performance, we take observations from one realization as the synthetic observation and use the other 99 samples to learn the observation-prediction relationship. We made three choices of the synthetic "truth" corresponding to low, moderate, and high porosity. In all the three synthetic datasets, LIP demonstrated robust prediction performance in terms of accuracy and uncertainty quantification, producing CO_2 pressure plume similar to the reference with reduced uncertainty. The results of synthetic dataset from the moderate porosity are discussed below in the main text and the other two datasets are presented in Appendix A.

Figure 1–3 summarize results using 9 years of observations. The scree plots of the PCA in Figure 1 indicate that the dimensions of both observation and prediction variables are greatly reduced by keeping the first few components with a little information loss. The first ten principle components of the observation variable $\mathbf{d}$ capture 99% of its variation and the ten principle components of the prediction variable $\mathbf{h}$ capture 98% of its variation. We then establish the statistical relationship of $\mathbf{d}$ and $\mathbf{h}$ in their reduced ten dimensions. The scatter plot of Figure 2 indicates that after applying CCA, observation variables $\mathbf{d}^c$ and prediction variables $\mathbf{h}^c$ in the reduced canonical space have strong linear correlation with coefficient of 0.99 and 0.92 for the first two components, respectively. This suggests that a linear regression model can be established to simulate the relationship of $\mathbf{d}^c$ and $\mathbf{h}^c$. In this study, both observation and prediction variables are the same type of quantity (i.e., CO_2 pressure) with smooth variation, thus it is not surprising that they show strong linear correlation here. The final prediction results are summarized in Figure 3. The figure indicates that the posterior mean from LIP accurately predicts the pressure front and movement where the estimated pressure map is similar to the reference with a coefficient of determination up to 0.98. In comparison to the priors, the averaged absolute error of the posterior mean is 8.4 psi, which is more than three times smaller than that of the prior mean of 26.58 psi. After effectively incorporating the observations, LIP also greatly reduces the predictive uncertainty, giving a trustable forecasting. The resulted accurate and credible prediction of the CO_2 pressure distribution in the reservoir is critical for risk assessment and site operators' decision.

Figure 4 demonstrates that LIP can effectively assimilate observation data to gradually improve predictions. As we see, after incorporating more years of measurements for forecasting, the posterior mean of the estimated pressure map gets closer and closer to the reference in Figure 3. Note that assimilation of these additional observations in LIP does not require extra reservoir simulations. LIP incorporates new observation data by performing the analysis in Steps 2-4 of Section 2 based on the corresponding observation variable samples from the prior sample set. This statistical analysis is very fast which promises integrating real-time measurements for timely forecasting in field operation.
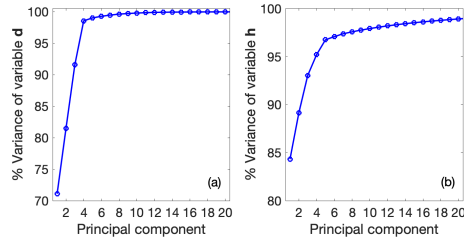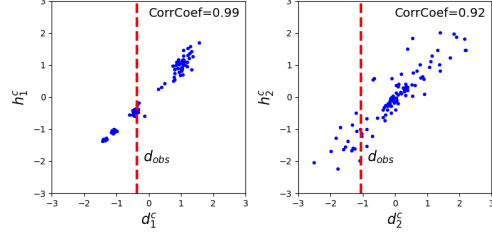
Figure 1: Scree plots of **d** and **h** in PCA.

Figure 2: Scatter plots of the first two components of $\mathbf{d}^c$ and $\mathbf{h}^c$, and observation $\mathbf{d}_{obs}$ after CCA.

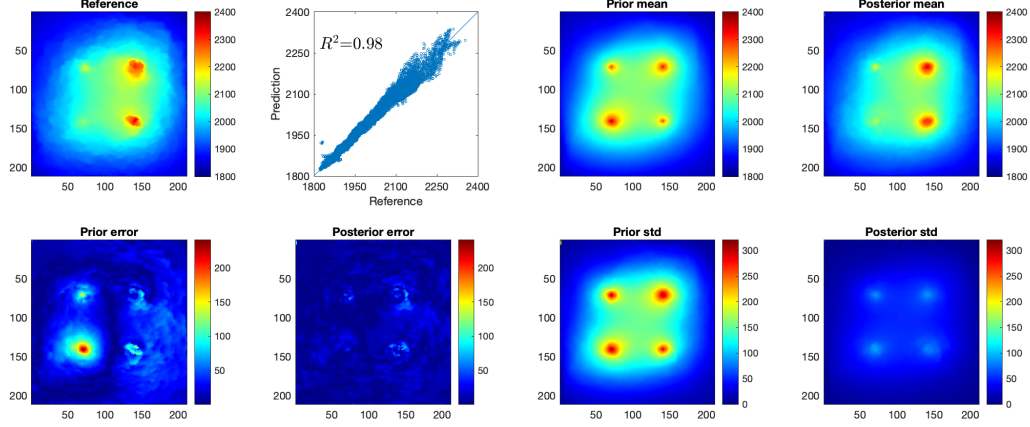**Figure 3:** Evaluation of LIP-predicted CO_2 pressure after 10-years of injection based on 9 years of observations. *Top, left-right*: reference CO_2 pressure distribution (psi) in year 10; cross-plot of reference and LIP-predicted pressure distribution; mean pressure distribution (psi) from the prior samples; and LIP-estimated posterior mean after incorporating 9 years of observations; *Bottom*: absolute prediction error and the standard deviation (std) from the prior samples and LIP-generated posterior samples.
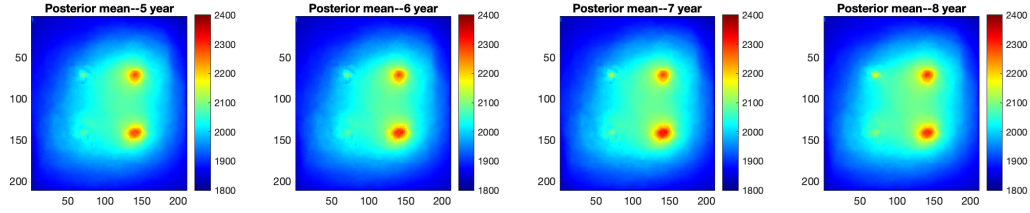

Figure 4: LIP estimated posterior mean of the CO_2 pressure distribution in year 10 after considering 5-8 years of observations.

Furthermore, the additional data is not necessarily from the same monitoring well with a longer period of observations, it can also come from other wells and can be different types of measurements.

4 Impacts and Future work

LIP has potential to fundamentally change how real-time decisions are made about CO_2 storage operations. Bypassing the traditional workflow of history matching and then forward simulations, LIP makes direct forecasting by learning observation-prediction relationship and provides continually updating forecasts of CO_2 distributions from streaming observations, thus providing operators with earlier warning of off-normal behavior and more time to implement mitigation measures. In this study, we demonstrated the robust performance of LIP in accurate and timely forecasting of CO_2 pressure maps using a limited simulation and observation data. Additionally, we showed that LIP can

quickly and effectively assimilate the observation streams for prediction improvement. The resulted accurate and credible forecasts of the CO_2 pressure distribution in the reservoir is crucial for carbon storage risk assessment and site decision making.

In the future, we will apply LIP to actual measurements from the field, and deploy it to CO_2 storage operators for real-time decisions. In deployment, it is important to ensure the actual measurements lie inside of the prior samples of the observation variables to effectively use the field measurements to update the priors to the posteriors. If not, we may increase the prior sample size, enlarge the prior uncertainty bound, use a different prior, or consider multiple models to increase the prior sample coverage.

References

[1] Alcalde, J., Flude, S., and Wilkinson, M. (2018). stimating geological co2 storage security to deliver on climate mitigation. *Nature Communication*, 9(2201).

[2] Bianco, A., Cominelli, A., Dovera, L., Naevdal, G., and Valles, B. (2007). History matching and production forecast uncertainty by means of the ensemble kalman filter: A real field application. All Days. SPE-107161-MS.

[3] Bosshart, N. W., Azzolina, N. A., Ayash, S. C., Peck, W. D., Gorecki, C. D., Ge, J., Jiang, T., and Dotzenrod, N. W. (2018). Quantifying the effects of depositional environment on deep saline formation co2 storage efficiency and rate. *International Journal of Greenhouse Gas Control*, 69:8–19.

[4] Brandt, A. R., Heath, G. A., Kort, E. A., O'Sullivan, F., Pétron, G., Jordaan, S. M., Tans, P., Wilcox, J., Gopstein, A. M., Arent, D., Wofsy, S., Brown, N. J., Bradley, R., Stucky, G. D., Eardley, D., and Harriss, R. (2014). Methane leaks from north american natural gas systems. *Science*, 343(6172):733–735.

[5] Celia, M. A., Bachu, S., Nordbotten, J. M., and Bandilla, K. W. (2015). Status of co2 storage in deep saline aquifers with emphasis on modeling approaches and practical simulations. *Water Resources Research*, 51(9):6846–6892.

[6] Chen, B., Harp, D. R., Lu, Z., and Pawar, R. J. (2020). Reducing uncertainty in geologic co2 sequestration risk assessment by assimilating monitoring data. *International Journal of Greenhouse Gas Control*, 94:102926.

[7] Daoud, J. I. (2017). Multicollinearity and regression analysis. *Journal of Physics: Conference Series*, 949:012009.

[8] Ji, X. and Zhu, C. (2015). Chapter 10 - co2 storage in deep saline aquifers. In Shi, F. and Morreale, B., editors, *Novel Materials for Carbon Dioxide Mitigation Technology*, pages 299–332. Elsevier, Amsterdam.

[9] Lu, D., Liu, S., and Ricciuto, D. (2019). An efficient bayesian method for advancing the application of deep learning in earth science. In *2019 International Conference on Data Mining Workshops (ICDMW)*, pages 270–278.

[10] Lu, D., Ricciuto, D., Walker, A., Safta, C., and Munger, W. (2017). Bayesian calibration of terrestrial ecosystem models: a study of advanced markov chain monte carlo methods. *Biogeosciences*, 14(18):4295–4314.

[11] Metz, B., Davidson, O., De Coninck, H., Loos, M., and Meyer, L. (2005). *IPCC special report on carbon dioxide capture and storage*. Cambridge: Cambridge University Press.

[12] Nakaten, B. and Kempka, T. (2014). Workflow for fast and efficient integration of petrel-based fault models into coupled hydro-mechanical tough2-mp - flac3d simulations of co2 storage. *Energy Procedia*, 63:3576–3581. 12th International Conference on Greenhouse Gas Control Technologies, GHGT-12.

[13] Namhata, A., Oladyshkin, S., Dilmore, R., Zhang, L., and Nakles, D. (2016). Probabilistic assessment of above zone pressure predictions at a geologic carbon storage site. *Scientific reports*, 6.

[14] Oliver, D. S. and Chen, Y. (2011). Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15:185–221.

[15] Pacala, S. and Socolow, R. (2004). Stabilization wedges: Solving the climate problem for the next 50 years with current technologies. *Science*, 305(5686):968–972.

[16] Yang, X., Liu, W., Liu, W., and Tao, D. (2021). A survey on canonical correlation analysis. *IEEE Transactions on Knowledge and Data Engineering*, 33(6):2349–2368.

A   Appendix

The key of LIP is to establish an observation-prediction relationship from their prior samples in a reduced dimension to be able to estimate posterior prediction distributions for given observations. LIP consists of the following four steps.

A.1   Generation of prior samples

The prior samples were generated by researchers at the University of North Dakota (UND) [3]. The uncertainty in geological properties such as porosity and permeability was considered. The UND team used the marginal probability distribution for porosity and the joint probability distribution for porosity-permeability to create 100 geomodels using Schlumberger's Petrel [12]. The 100 geomodels capture low, moderate and high porosity, where the porosity samples were generated from a beta distribution centered on 0.05, 0.17 and 0.27 for the low, moderate, and high porosity, respectively. Then, for each geomodel, a reservoir simulation was performed for 10 years using CMG-GEM (a reservoir simulator for compositional, chemical and unconventional reservoir modelling) to produce 100 prior samples of CO_2 saturation and pressure fields in the 3D domain at 32 time steps (monthly data in first 2 years and annual data in last 8 years).

A.2   Dimension reduction

The prediction variable (denote as $\mathbf{h}$) here is a spatial image and the observation variables (denote as $\mathbf{d}$) are four time series, which have spatial and temporal correlations, respectively. When the variable dimensions are highly correlated with each other, multicollinearity occurs. Multicollinearity results in numerical issues during model fitting and degrades predictive performance of the statistical model. Dimension reduction identifies degrees of freedom that capture the majority of the variance in the data. Therefore, performing statistical analysis in the reduced dimension removes the multicollinearity and facilitates the model fitting. Additionally, dimension reduction reduces the variables and thus reduces the required number of samples, which improves the computational efficiency and enhances the model reliability.

We use principal component analysis (PCA) for dimension reduction. PCA is a multivariate analysis technique that applies an orthogonal transformation to convert a set of samples of possibly correlated variables into a set of values of uncorrelated variables, called principal components. Typically, the first a few components of the PCA decomposition explain most of the variance of data. By keeping only those a few dimensions, we thus achieve a dimension reduction. Since our observation variables are from multiple sources (i.e., four injection wells), we use a mixed PCA to pool data together and generate a reduced dimensional projection of the combined data to consider difference in magnitude. First, a standard PCA is performed on each of the data source to obtain the largest singular values. Next, each data source is normalized according to its first singular value; this accounts for any difference in scales amongst the data sources. Last, the normalized data inputs are concatenated and the standard PCA is applied to this final matrix. After dimension reduction, we obtain observation variables $\mathbf{d}^f$ and prediction variables $\mathbf{h}^f$, respectively, in the reduced dimension. PCA is a bijective operation, so the original high-dimensional variable can be recovered uniquely by undoing the projection.

A.3 Establishing the statistical relationship

The relationship between $\mathbf{d}^f$ and $\mathbf{h}^f$ in the reduced dimension can be nonlinear, which challenges the statistical model learning. We first use canonical correlation analysis (CCA) to linearize the relationship to simplify the model fitting. CCA is a multivariate analysis method that can be applied to transform the relationships between pairs of vector variables into a set of independent linearized relationships between pairs of scalar variables [16]. The resulting linear combinations are denoted as $\mathbf{d}^c$ and $\mathbf{h}^c$, and called the canonical variates of $\mathbf{d}^f$ and $\mathbf{h}^f$. The canonical transformation is found through the eigen-decomposition of the sample covariance matrix and this CCA transformation is reversible. If $\mathbf{d}^c$ and $\mathbf{h}^c$ in the canonical space are nearly linearly correlated (e.g., with a correlation coefficient greater than 0.9), a linear model can be used to simulate their relationship. If after CCA, the relationship of $\mathbf{d}^c$ and $\mathbf{h}^c$ is still not quite linear, we can use advanced ML models such as neural networks for regression.

A.4 Bayesian inference of the prediction

We use Bayesian inference to estimate predictions. But unlike the traditional workflow which uses Bayesian methods to quantify uncertainties of model parameters first and then infers prediction uncertainties [10], we use Bayesian methods to calculate the posterior distribution of the predictions directly. Based on Bayes' rule, the posterior distribution of a prediction variable $\mathbf{h}$ for some observed data $\mathbf{d}_{obs}$ is

$$p(\mathbf{h}|\mathbf{d}_{obs}) \propto L(\mathbf{h}|\mathbf{d}_{obs})p(\mathbf{h}), \tag{1}$$

where $p(\mathbf{h})$ is the prior distribution and $L(\mathbf{h}|\mathbf{d}_{obs})$ is the likelihood function. PCA and CCA enable reducing a set of high-dimensional variables $(\mathbf{d}, \mathbf{h})$ to a set of low-dimensional and linearly correlated variables $(\mathbf{d}^c, \mathbf{h}^c)$. We first estimate the posterior distribution $p(\mathbf{h}^c|\mathbf{d}_{obs}^c)$ and then transform back $\mathbf{h}^c$ to its original space $\mathbf{h}$. In the canonical space, $p(\mathbf{h}^c|\mathbf{d}_{obs}^c)$ can be estimated by

$$p(\mathbf{h}^c|\mathbf{d}_{obs}^c) \propto L(\mathbf{h}^c|\mathbf{d}_{obs}^c)p(\mathbf{h}^c). \tag{2}$$

We use a linear model G to simulate the relationship between $\mathbf{d}^c$ and $\mathbf{h}^c$, i.e., $\mathbf{d}^c = G\mathbf{h}^c$. By assuming a Gaussian likelihood, as commonly done in the CCS community [14], $L(\mathbf{h}^c|\mathbf{d}_{obs}^c)$ can be formulated as

$$L(\mathbf{h}^c|\mathbf{d}_{obs}^c) = \exp\Big(-\frac{1}{2}\big(G\mathbf{h}^c - \mathbf{d}_{obs}^c\big)^T \mathbf{C}_{\mathbf{d}^c}^{-1}\big(G\mathbf{h}^c - \mathbf{d}_{obs}^c\big)\Big). \tag{3}$$

Through normal score transformation based on the sample mean $\bar{\mathbf{h}}_{prior}^c$ and the sample covariance $\mathbf{C}_{\mathbf{h}^c}$ calculated from the prior samples of $\mathbf{h}^c$, we obtain a Gaussian prior of $\mathbf{h}^c$ in the transformed space. Since the prior and the likelihood of $\mathbf{h}^c$ are Gaussian, its posterior is also Gaussian and the posterior mean $\tilde{\mathbf{h}}^c$ and posterior covariance $\tilde{\mathbf{C}}_{\mathbf{h}^c}$ can be analytically estimated by

$$\tilde{\mathbf{h}}^c = \bar{\mathbf{h}}_{prior}^c + \mathbf{C}_{\mathbf{h}^c}G^T\big(G\mathbf{C}_{\mathbf{h}^c}G^T + \mathbf{C}_{\mathbf{d}^c}\big)^{-1}\big(\mathbf{d}_{obs}^c - G\bar{\mathbf{h}}_{prior}^c\big), \tag{4}$$

$$\tilde{\mathbf{C}}_{\mathbf{h}^c} = \big(G^T\mathbf{C}_{\mathbf{d}^c}^{-1}G + \mathbf{C}_{\mathbf{h}^c}^{-1}\big)^{-1}, \tag{5}$$

where $\mathbf{C}_{\mathbf{d}^c}$ is the covariance matrix of the observation error. In this work, we are considering a synthetic case where the observed data is one realization from the prior samples, so $\mathbf{C}_{\mathbf{d}^c}$ here is calculated as the covariance of the residuals from the linear model fitting.

An advantage of the Gaussian process regression is that a Gaussian distribution is uniquely defined by its mean and covariance and sampling a Gaussian distribution is straightforward. Thus, based on Eq. (4) and (5), we can generate posterior samples of $\mathbf{h}^c$ directly. By undoing the normal score transformation followed by the back transformation of CCA, we obtain posterior samples of $\mathbf{h}^f$. Next, after back transformation of PCA, we obtain the posterior samples of prediction quantity $\mathbf{h}$ in its original space. Based on these $\mathbf{h}$ samples, we then estimate posterior prediction distribution.

A.5 Additional results

In Section 3 of the main text, we presented the prediction results of one synthetic dataset corresponding to the moderate porosity. Here we show the results of another two synthetic datasets which correspond to the low (Figure 5) and high (Figure 6) porosity, respectively. These two synthetic datasets, although presenting dramatically different target CO_2 pressure patterns, consistently demonstrate that the LIP method can accurately forecast the pressure distribution and movement, resulting in an estimated pressure field similar to the reference with a reduced uncertainty.
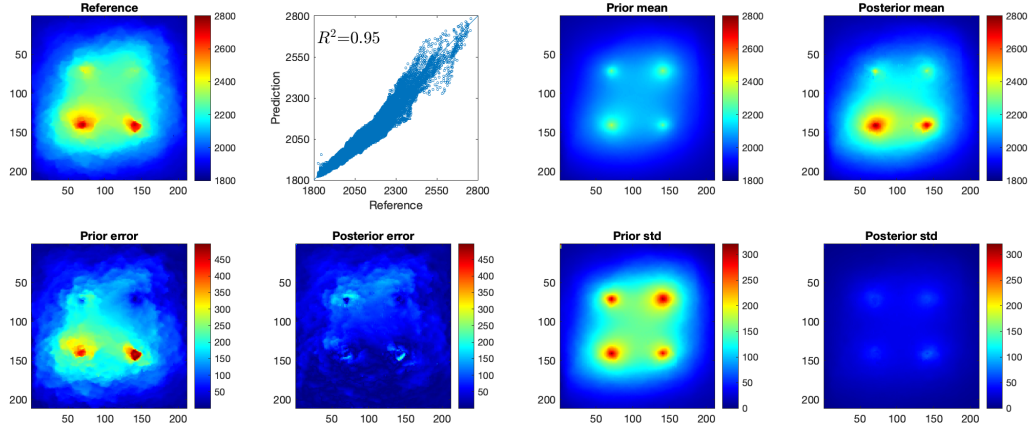
Figure 5: Evaluation of LIP-predicted CO_2 pressure after 10-years of injection based on 9 years of observations. *Top, left-right*: reference CO_2 pressure distribution (psi) in year 10; cross-plot of reference and LIP-predicted pressure distribution; mean pressure distribution (psi) from the prior samples; and LIP-estimated posterior mean after incorporating 9 years of observations; *Bottom*: absolute prediction error and the standard deviation (std) from the prior samples and LIP-generated posterior samples. This result corresponds to the synthetic dataset with low porosity.
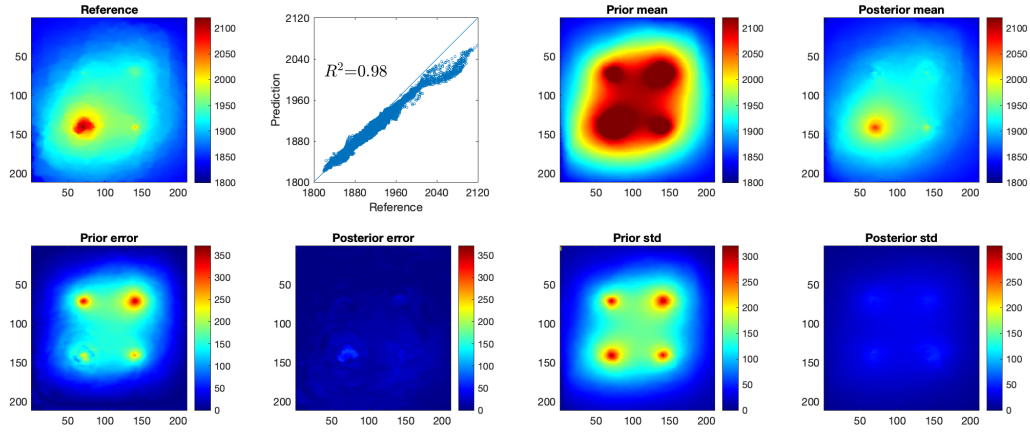

Figure 6: Evaluation of LIP-predicted CO_2 pressure after 10-years of injection based on 9 years of observations. *Top, left-right*: reference CO_2 pressure distribution (psi) in year 10; cross-plot of reference and LIP-predicted pressure distribution; mean pressure distribution (psi) from the prior samples; and LIP-estimated posterior mean after incorporating 9 years of observations; *Bottom*: absolute prediction error and the standard deviation (std) from the prior samples and LIP-generated posterior samples. This result corresponds to the synthetic dataset with high porosity.