add your advisors

Highlights

Optimal monitoring design for uncertainty quantification during geologic CO_2 sequestration: A machine learning approach

Misael M. Morales,Mohamed Mehana,Bailian Chen

- Filtering-based data assimilation method is developed to perform monitoring design.
- Machine learning reduced-order model is used to reduce computational cost of data assimilation process.
- Monitoring well placement optimization is performed to reduce uncertainty and minimize leakage risks. prediction

the uncertainty in

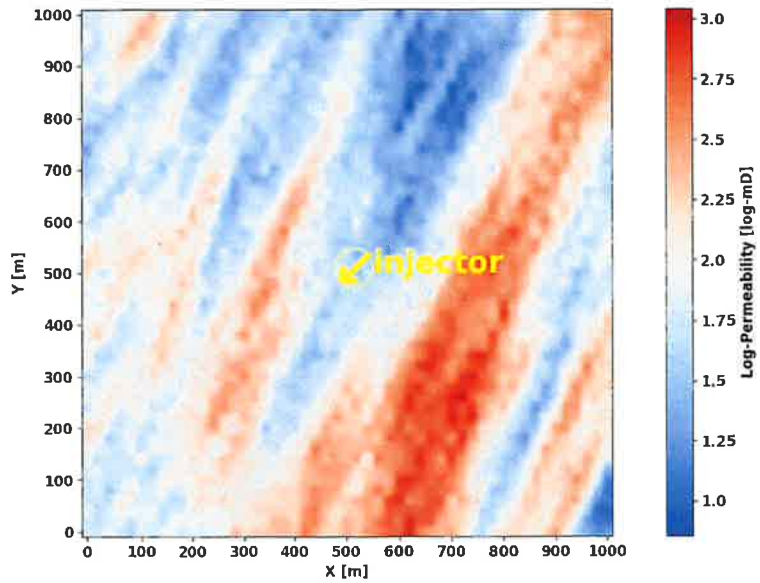


Figure 5: Log-permeability distribution of the base model. The darkest blue color corresponds to the lowest permeability, while the darkest red color corresponds to the highest. The yellow circle with an arrow indicates the CO_2 injection well.

3. Results

In this section, we apply our optimal monitoring design workflow using the ANN ROMs and filter-based uncertainty quantification approach to obtain the optimal monitoring well placement and monitoring measurement data type for two synthetic GCS examples.

3.1. Workflow Validation

We validate the workflow for optimal GCS monitoring design using a simple example. Fig. 5 shows the log-permeability distribution for the base model with a CO_2 injection well at the center, noted with a yellow circle and arrow. All the monitoring data in this study are collected in the aquifer zone, similar to monitoring at the above zone monitoring interval (AZMI) in the work of Sun et al. [47]. The monitoring frequency is once per month for the duration of 5 years injection, resulting in 60 monitoring data points. The objective function, M_c, is the cumulative CO_2 leakage through legacy wells at the end of 5 years. In the model, we set up three material zones corresponding to the three adjacent formations, namely the storage reservoir, caprock, and aquifer. The cumulative CO_2 saturation in each zone can be output from the FEHM simulation results, and the cumulative leakage is computed by summing the CO_2 mass in the aquifer and caprock layers. Our approach for monitoring design involves quantifying the uncertainty reduction by monitoring pressure, CO_2 saturation, are or temperature at each potential monitoring well location.

reference and full name for FEHM

The data assimilation error tolerance, τ from Eq. (6), for pressure is set equal to 0.002 MPa, while for CO_2 saturation it is 0.05, and for temperature it is 0.002°C. Note that the choice of τ is site ~~and case~~ specific and is based on engineering judgement that takes into consideration the measurement and modeling error.

Two case studies are considered in this study: (1) GCS project with 3 potential leakage pathways (i.e. legacy wells), and (2) GCS project with 6 potential leakage pathways. The uncertain parameters are the permeability multiplier, k_R for the storage reservoir, and the ℓ permeability values for the ℓ potential leakage pathways, where $\ell = 3$ and $\ell = 6$, respectively. The total number of uncertain parameters, u^ℓ are 4 and 7, respectively. The lower and upper bounds for the uncertain parameters are shown in Table 1. For each case study, we run 500 training simulations generated by LHS with u^ℓ uncertain parameters. Each HFS requires approximately 22 minutes. We perform parallelization on an 8-node cluster, and the total simulation time is approximately 23 hours to finish all 500 training realizations. Fig. 6 shows the base model for Case 1 and Case 2, respectively.

We choose one simulation from the 500 training realizations in Case 1 to show when CO_2 leakage occurs. The values of the different parameters for the chosen model are shown in Table 2. The cumulative CO_2 leakage over the GCS project time is shown in Fig. 7. Figure 8 shows the leaked CO_2 saturation distribution at the top of the aquifer.

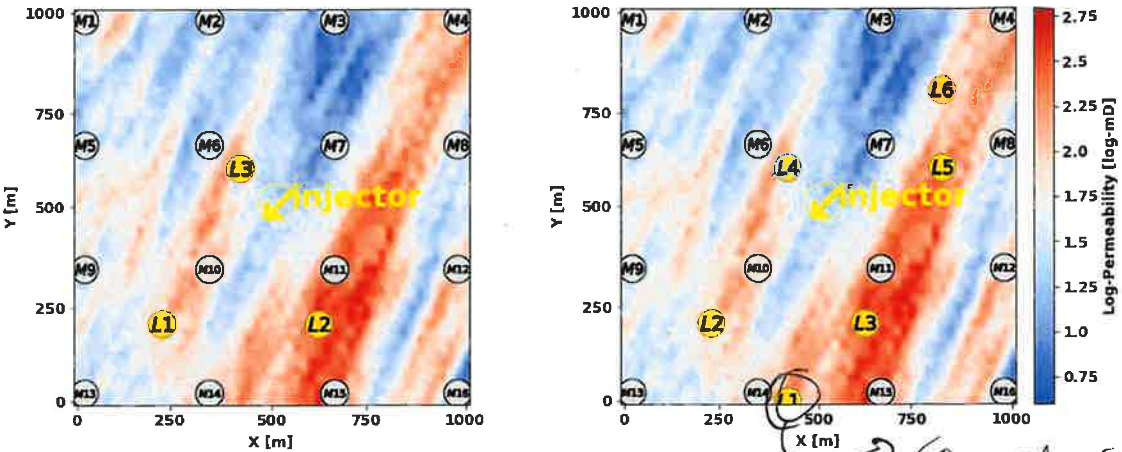


Figure 6: Log-permeability distribution of the base model for Case 1 (left) with 3 potential leaky pathways, and Case 2 (right) with 6 potential leaky pathways. The dark yellow circles labeled L_i represent the leakage pathways, light gray circles labels M_i are the possible monitoring well locations, and the yellow circle with an arrow is the CO_2 injection well.

Cannot see it

It can be seen that CO_2 leakage occurs after about 210 days of injection. We observe that CO_2 is leaking through the potential pathway L_3, which is 141.4 m away from the injector, while no leakage occurs at potential pathways L_1 and L_2 after 5 years of injection. For this specific example, it is important to note that the permeability of L_3, k_v^3 is higher than that of L_1 and L_2.

effective well

Table 2
The parameters for one chosen model from the 500 training realizations in Case 1.

Parameters	Value	Unit
CO_2 injection rate	3.17	kg/s
Thickness of caprock layer	30	m
Permeability of 1st potential leakage pathway	2.19×10^{-17}	m^2
Permeability of 2nd potential leakage pathway	3.37×10^{-17}	m^2
Permeability of 3rd potential leakage pathway	2.97×10^{-16}	m^2
Distance between injector and 1st potential leakage pathway	424.3	m
Distance between injector and 2nd potential leakage pathway	360.6	m
Distance between injector and 3rd potential leakage pathway	141.4	m
Permeability for aquifer layer	1×10^{-13}	m^2
Permeability for caprock layer	1×10^{-19}	m^2
Reservoir permeability multiplier	1.88	–

For each case, the 500 training realizations are used to train ROMs for the monitoring data and cumulative CO_2 leakage using the ANN architecture in Fig 1. Fig. 2 shows the quality of the ROMs tested by 10-fold cross-validation [116, 123]. The MSE and R^2 are 8.5×10^{-4} and 0.98, respectively. This proves that the fidelity of ROMs to the numerical simulations is high at the advantage of a much lower computational cost.

With the proposed workflow, the expected uncertainty reduction of the cumulative CO_2 leakage can be computed for each of the 16 possible monitoring well locations, for each monitoring measurement type. For each data set, 200 possible realizations of monitoring data are generated following Step 2 in Section 2.4. To obtain the expected uncertainty reduction using Eq. (3), the prior uncertainty $U[P(M_c)]$ and posterior uncertainty $U[P(M_c|D^j)]$ corresponding to each possible monitoring data realization D^j for each possible well location x^p should be computed. Higher uncertainty reduction of the objective function indicates greater VOI in the monitoring data obtained from the optimal well location and monitoring measurement type. Through these examples, we can see that our proposed workflow can be effectively used to determine optimal CO_2 monitoring design from a set of alternative monitoring designs.

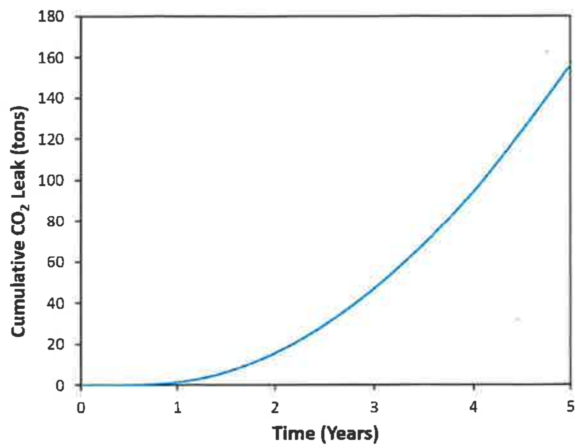
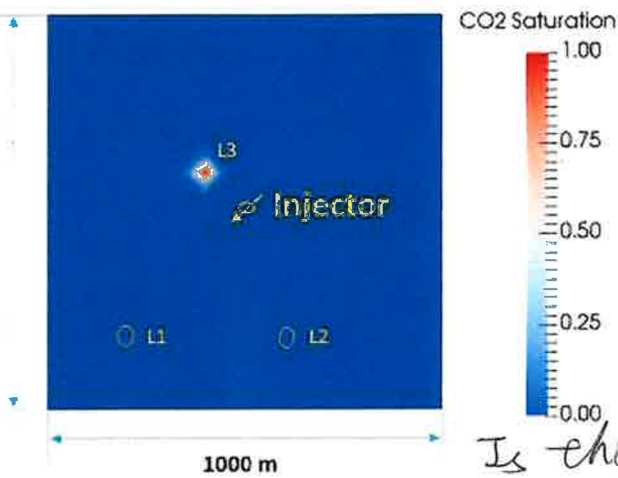


Figure 7: Cumulative CO_2 leakage over time computed for one chosen training realization in Case 1.



~~Does ... have location 6?~~

(i.e., M6)

Is this a new plot or adpt from ~~...~~ my original paper?

Figure 8: Plan view (top of the aquifer) of CO_2 leakage at the end of 5 years of injection based on one chosen training realization in Case 1. Yellow circles indicate the potential leakage pathways. Units for CO_2 saturation is fraction.

evidence?

If this is used in my original paper, you need to cite it.

We observe that monitoring for pressure provides the highest uncertainty reduction in general, followed by CO_2 saturation and lastly pressure. The spatial distribution of uncertainty reduction in CO_2 leakage is shown in Fig. 9 for every possible well location x^p in the 4×4 subgrid, and point-wise comparison of the uncertainty reduction at each monitoring well location for each measurement type is shown in Fig. 10. One can observe that placing a monitoring well at location 6 and assimilation the pressure measurements provides the highest uncertainty reduction possible in the monitoring design for both Case 1 and Case 2. For Case 1, the optimal monitoring design given by $(\textit{pressure}, x^6)$ yields an uncertainty reduction in the cumulative leakage of CO_2 of approximately $29.42{\times}10^3$ metric tons (29.24 kt), approximately a 73% reduction. Meanwhile, the optimal design for CO_2 saturation and temperature monitoring yield an uncertainty reduction of approximately 19.34 kt and 17.71 kt, respectively. Similarly, for Case 2, the optimal monitoring design given by $(\textit{pressure}, x^6)$ yields an uncertainty reduction of 26.29 kt of cumulative CO_2 leakage, a 62% reduction, while the optimal design for CO_2 saturation and temperature monitoring yield an uncertainty reduction of approximately 16.94 kt and 16.29 kt, respectively.

It's better to make it a little different.

The histograms for the prior and posterior distributions of the objective function obtained from the data realizations 1 and 100 for Case 1 and 2, respectively, are show in Fig. 11. The prior distribution is generated using LHS from the set of uncertain input parameters, k_V^ℓ and k_R, with a uniform distribution and calculating the cumulative CO_2 leakage using the ROMs. The posterior distribution for two random realizations, namely realization 1 and 100, are shown. These two are selected given that they had a relatively high amount of cumulative CO_2 leakage. Recall that the total

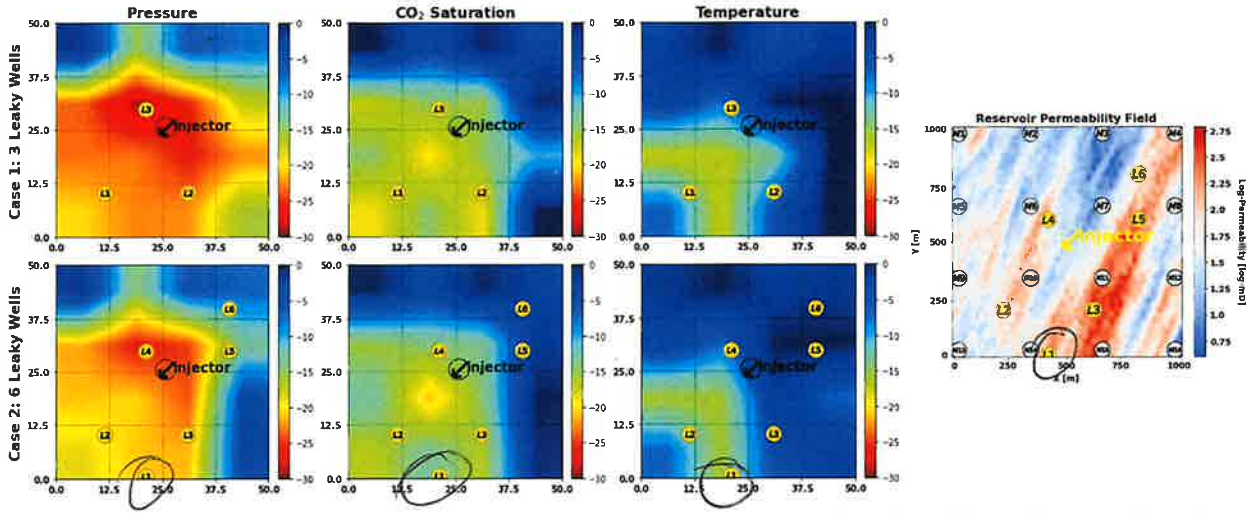


Figure 9: Left: Plan view (top of the aquifer) of the uncertainty reduction obtained by all possible monitoring well locations. Top row represents Case 1 with 3 leakage pathways, and the bottom row represents Case 2 with 6 leakage pathways. Each column represents monitoring data for pressure, CO_2 saturation, and temperature, respectively. Right: Plan view of the reservoir permeability field with possible leakage pathways (yellow) and all possible monitoring well locations (gray).

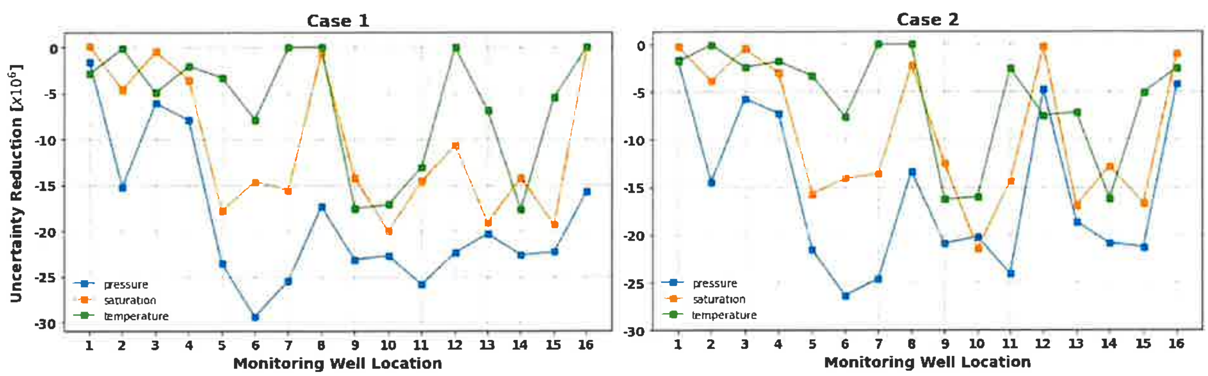


Figure 10: The point-wise calculated uncertainty reduction at each possible monitoring well location for each measurement type. Case 1 is shown on the left, and Case 2 on the right.

uncertainty reduction, U_R is given by the difference between the expected posterior uncertainty (the expected value of the ensemble of realizations) and the prior uncertainty distribution. The variances of the posterior distributions calculated show significant reduction in uncertainty of cumulative CO_2 leakage compared to the priors. The optimal monitoring design ($pressure, x^6$) yields a reduction in cumulative CO_2 leakage uncertainty of approximately 29.24 kt (73%) and 26.29 kt (62%) for Case 1 and Case 2, respectively.

3.2. Discussion

GCS monitoring operations require detailed data processing and interpretation in order to accurately quantify and potentially minimize leakage risks. Associated costs of performing monitoring operations requires evaluating the potential value of monitoring measurement type, and optimal monitoring well location, before the actual monitoring strategy takes place in the field. The workflow proposed can be used to select an optimal monitoring design that is robust under multiple potential leakage scenarios. Previous efforts in characterizing the geologic uncertainty in GCS projects show the importance and effect of these parameters [46, 85, 109], selecting the optimal measurement type [49, 106], or selecting the optimal monitoring well location [39, 47]. Our work provides a framework to unify these three concepts under a single optimal monitoring design framework.

This is not actual wells. They are just potential locations for monitoring wells. What you need to analyze is the relationship among the legacy well (e.g. L1, L2 ...) perm/location, injection well location, reservoir perm, and the "potential monitoring well location".

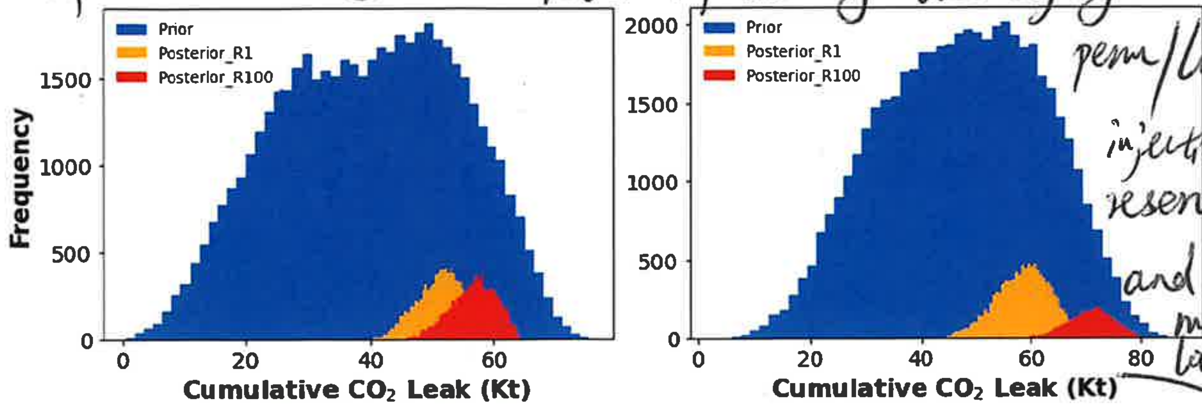


Figure 11: The histograms for the prior (blue) and posterior distributions obtained at the optimal monitoring design from the data realizations 1 (orange) and 100 (red) for Case 1 (left) and Case 2 (right), respectively.

This part of discussion is more case specific. In some cases, you will find that effective wellbore perm for legacy well is very important.

The slight difference in uncertainty reduction between Case 1 and Case 2, despite the fact that we have more leakage pathways, can be attributed to the fact that the main source of uncertainty is the geologic parameters, namely k_v^ℓ and k_R, ~~rather than the number of leakage pathways~~. These parameters ultimately control the leaking of CO_2, and can range ~~between very small ($\sim 0.001mD$) to medium ($\sim 10mD$)~~. Thus, the cumulative CO_2 leakage does not directly correlate with having ~~more or less~~ leakage pathways, but with the permeability of the leakage pathways, k_V^ℓ, and the reservoir permeability multiplier, k_R. ~~For instance,~~ wells 6-7 and 10-11 lie closest to the injection well at approximately 176.8 m, yet well 6 yields the highest ~~uncertainty~~ reduction since it is closest to a high permeability streak, similar to well 11. On the other hand, wells 7 and 10 lie on low permeability streaks, making the plumes travel relatively slower. Therefore, distance from injector and leakage pathways is important, but mostly controlled by the permeability heterogeneity in the reservoir.

This is not on low perm streak

Furthermore, it is evident that monitoring for pressure data yields the highest uncertainty reduction out of the three possible measurement types. This is due to the fact that the pressure ~~plumes~~ front travel the fastest along a given subsurface formation, followed by saturation plumes and temperature plumes, in that order [32]. Temperature plumes tend to travel the least, given the thermal equilibrium of deep underground formations and the gradual enthalpy change in the reaction of CO_2 in saline aquifers [124]. In some monitoring locations, such as wells 7 and 8, monitoring ~~for temperature~~ provides little to no uncertainty reduction whatsoever, while pressure monitoring yields about 17-25 kt reduction. The total uncertainty reduction given by the sum of all three measurements correlated primarily with the pressure monitoring data, and assimilating multiple measurements types simultaneously does not necessarily add linearly, and tends to be dominated by a single measurement. Refer to Chen et al. [110] for more information on assimilating multiple measurement types simultaneously.

uncertainty

in the cumulative CO_2 leakage

It is also important to note that the monitoring wells are only drilled through the aquifer, and not the caprock and reservoir, thus only collecting monitoring measurements at the aquifer zone. This is way, despite x^6 being further away from the injector than the closest leakage pathway (L_3 in Case 1 and L_4 in Case 2), it still provides the most information about CO_2 leakage into the aquifer zone. Also, recall the aquifer is geologically homogeneous, and the pressure, CO_2 saturation, and temperature response does not necessarily exactly correlate with the reservoir response.

Regarding the monitoring well location grid, location x^6 yields the highest uncertainty reduction when using pressure measurements for both Case 1 and Case 2. However, the 16 possible monitoring well locations exist on a coarse 4×4 subgrid. Monitoring location x^6 is closest to the injection well and a leakage pathway in both Case 1 and Case 2, and therefore provides the best uncertainty reduction. However, if a finer monitoring well location grid were used, say 16×16 or 32×32 (which is still coarser than the simulation grid at 51×51), a location between the injector well and L_3 in Case 1 and L_4 in Case 2 would most likely yield the best uncertainty reduction. This is due to the fact that a monitoring well in that location would be able to detect the CO_2 plume before it even reaches and leakage pathways, and would still be in the region of high permeability in the reservoir, thus yielding the best uncertainty reduction in cumulative CO_2 leakage.

please mention that we are going to conduct case study with finer monitoring well location grid in our next step.

paired with one injector

Moreover, given that we are using a coarse monitoring well location grid, several locations such as M_{1-4}, M_8, M_{12}, and M_{14-15} yield little to no uncertainty reduction. We stipulate that this occurs because the CO_2 plume reaches the leakage pathways before it reaches these monitoring wells, and therefore provide little information to the data assimilation to reduce the uncertainty in leakage. This also leads to the idea that most of the CO_2 leakage is occurring in the pathways closest to the injection well, and thus the closer monitoring locations provide a better uncertainty reduction than those in the boundaries or further away. This relates to the fact that despite Case 2 having more leakage pathways, the uncertainty reduction in kt of CO_2 is not much different from that of Case 1.

Even though the examples used in our study to demonstrate how monitoring data from a shallow aquifer can be used, the proposed workflow can be extended and applied to monitoring data collected at any location and time within the GCS project. The potential value of such monitoring data can be evaluated by the presented workflow. Furthermore, placing several monitoring wells ~~can~~ provide a slight advantage compared to a single injector-monitor pair, but is impractical in field applications. Moreover, using several monitoring measurement types simultaneously provides little to no advantage compared to pressure monitoring. Refer to Chen et al. [110] for further details.

(e.g., monitoring from storage reservoir and/or AZMI)

At a CO_2 storage field operation, an optimal monitoring schedule and location based on the VOI described in this work can be used to collect the best possible monitoring data. The monitoring data can be assimilated to calibrate the uncertain model parameters using traditional data assimilation methods such as EnKF [125] or ES-MDA [126]. The calibrated models can be used to improve the accuracy in prediction for future and long-term behavior of the injected CO_2 and associated

due to high cost in drilling ~~monitoring~~ monitoring wells.

4. Conclusions leakage risk assessment

In this study, a workflow based on a machine learning reduced-order modeling technique and uncertainty quantification method within an optimization loop is proposed for geologic CO_2 sequestration monitoring design. We use the uncertainty reduction in cumulative CO_2 leakage as the quantity of interest to measure the potential value of monitoring measurement data. The optimal monitoring design yields an uncertainty reduction of approximately 29.94 kt (73%) and 26.29 kt (62%) in CO_2 leakage for Case 1 and Case 2, respectively. The following conclusions have been drawn from this research:

1. The proposed workflow can generate reasonable values of uncertainty reduction in different risk ~~metrics~~ quantities at CO_2 storage site, including cumulative CO_2 leakage by utilizing different monitoring designs and has ~~been~~ demonstrated using a synthetic GCS project. The optimal monitoring design is obtained ~~by~~ by assimilating pressure data at ~~monitoring well~~ location 6 for a monitoring well.
2. The effect of different types of measurements (pressure, CO_2 saturation, and temperature) and the effect of monitoring well location on the choice of monitoring design is investigated. It is observed that pressure data has more value of information compared to CO_2 saturation, while temperature has the least value of information, though still valuable in terms of uncertainty reduction compared to no monitoring strategy.
3. Well placement optimization is important to maximize the value of information for the monitoring design. Typical operations include pairs of one monitoring well for each injection well, partly due to the cost of drilling and data acquisition. Determination of the best location provides significant benefits in reducing the uncertainty of cumulative CO_2 leakage and ensure an efficient risk management in the life-cycle of a GCS project.
4. The incremental reduction in uncertainty in the cumulative CO_2 leakage may not increase proportional to the distance from the injection well, and is a strong function of the reservoir permeability heterogeneity. Thus, an optimal monitoring well placement and measurement type is important to minimize present and future potential risks.
5. The subgrid for the possible monitoring well locations will have a major impact in the optimal monitoring design. The uncertainty reduction obtained at the monitoring locations will depend on the geologic heterogeneity, which affects the CO_2 plume movement in the subsurface. However, with proper knowledge of the geologic setting and a finer monitoring location subgrid, one can further optimize the monitoring design to further reduce the uncertainty in cumulative CO_2 leakage.

Future research in this topic includes investigating the effect of different monitoring measurement types, such as seismic sensing, or a combination of the available measurements. Similarly, multi-scale or local grid refinement to optimize the monitoring well placement can help improve the reduction in uncertainty for CO_2 leakage risks. Moreover, a global optimization strategy, such as genetic algorithm or simulated annealing, can provide more computationally

ES-MDA

efficient results for finer subgrids. Other alternatives could include incorporating other data assimilation techniques such as EnKF or ~~ESMDA,~~ or performing spatial data assimilation rather than assimilating point-wise measurements. Including other risks such as geomechanical failure can help characterize a GCS site and provide a more in-depth risk management program.

CRediT authorship contribution statement

Misael M. Morales: Conceptualization of this study, Methodology, Software, Writing - Original draft preparation. **Mohamed Mehana:** Conceptualization, Writing, ~~Funding acquisition~~. **Bailian Chen:** Conceptualization, Software, Writing, Funding acquisition.

References

[1] Bert Metz. Carbon dioxide capture and storage: special report of the intergovernmental panel on climate change, 2005.

[2] K. Michael, A. Golab, V. Shulakova, J. Ennis-King, G. Allinson, S. Sharma, and T. Aiken. Geological storage of co2 in saline aquifers—a review of the experience from existing storage operations. *International Journal of Greenhouse Gas Control*, 4(4):659–667, 2010. ISSN 1750-5836. doi: https://doi.org/10.1016/j.ijggc.2009.12.011.

[3] A. Kopp, P.J. Binning, K. Johannsen, R. Helmig, and H. Class. A contribution to risk analysis for leakage through abandoned wells in geological co2 storage. *Advances in Water Resources*, 33(8):867–879, 2010. doi: 10.1016/j.advwatres.2010.05.001. cited By 47.

[4] A. Goodman, G. Bromhal, B. Strazisar, T. Rodosta, W.F. Guthrie, D. Allen, and G. Guthrie. Comparison of methods for geologic storage of carbon dioxide in saline formations. *International Journal of Greenhouse Gas Control*, 18:329–342, 2013. doi: 10.1016/j.ijggc.2013.07.016. cited By 48.

[5] N. Castelletto, P. Teatini, G. Gambolati, D. Bossie-Codreanu, O. Vincké, J.-M. Daniel, A. Battistelli, M. Marcolini, F. Donda, and V. Volpi. Multiphysics modeling of co2 sequestration in a faulted saline formation in italy. *Advances in Water Resources*, 62:570–587, 2013. doi: 10.1016/j.advwatres.2013.04.006. cited By 25.

[6] B. Li and S.M. Benson. Influence of small-scale heterogeneity on upward co2plume migration in storage aquifers. *Advances in Water Resources*, 83:389–404, 2015. doi: 10.1016/j.advwatres.2015.07.010. cited By 84.

[7] J.S. Levine, I. Fukai, D.J. Soeder, G. Bromhal, R.M. Dilmore, G.D. Guthrie, T. Rodosta, S. Sanguinito, S. Frailey, C. Gorecki, W. Peck, and A.L. Goodman. U.s. doe netl methodology for estimating the prospective co2 storage resource of shales at the national and regional scale. *International Journal of Greenhouse Gas Control*, 51:81–94, 2016. doi: 10.1016/j.ijggc.2016.04.028. cited By 81.

[8] Mai Bui, Claire S. Adjiman, André Bardow, Edward J. Anthony, Andy Boston, Solomon Brown, Paul S. Fennell, Sabine Fuss, Amparo Galindo, Leigh A. Hackett, Jason P. Hallett, Howard J. Herzog, George Jackson, Jasmin Kemper, Samuel Krevor, Geoffrey C. Maitland, Michael Matuszewski, Ian S. Metcalfe, Camille Petit, Graeme Puxty, Jeffrey Reimer, David M. Reiner, Edward S. Rubin, Stuart A. Scott, Nilay Shah, Berend Smit, J. P. Martin Trusler, Paul Webley, Jennifer Wilcox, and Niall Mac Dowell. Carbon capture and storage (ccs): the way forward. *Energy Environ. Sci.*, 11:1062–1176, 2018. doi: 10.1039/C7EE02342A. URL `http://dx.doi.org/10.1039/C7EE02342A`.

[9] Energy 2020. European commission. In *A strategy for competitive, sustainable and secure energy*, 2010.

[10] United nations. Agreement, p. *United Nations Treaty Collect*, pages 1–27, 2015.

[11] Zhenxue Dai, Hari Viswanathan, Richard Middleton, Feng Pan, William Ampomah, Changbing Yang, Wei Jia, Ting Xiao, Si Yong Lee, Brian McPherson, Robert Balch, Reid Grigg, and Mark White. Co2 accounting and risk analysis for co2 sequestration at enhanced oil recovery sites. *Environmental Science and Technology*, 50(14):7546–7554, 7 2016. ISSN 15205851. doi: 10.1021/acs.est.6b01744.

[12] D.R. Harp, R. Pawar, J.W. Carey, and C.W. Gable. Reduced order models of transient co2 and brine leakage along abandoned wellbores from geologic carbon sequestration reservoirs. *International Journal of Greenhouse Gas Control*, 45:150–162, 2 2016. ISSN 1750-5836. doi: 10.1016/j.ijggc.2015.12.001. cited By 38.

[13] J. Song and D. Zhang. Comprehensive review of caprock-sealing mechanisms for geologic carbon sequestration. *Environ Sci Technol*, 47 (1):9–22, 2012.

[14] W. Sifuentes, M.J. Blunt, and M.A. Giddins. Modeling co2 storage in aquifers: Assessing the key contributors to uncertainty. volume 1, pages 148–160, 2009. cited By 38.

[15] J.M. Nordbotten, B. Flemisch, S.E. Gasda, H.M. Nilsen, Y. Fan, G.E. Pickup, B. Wiese, M.A. Celia, H.K. Dahle, G.T. Eigestad, and K. Pruess. Uncertainties in practical simulation of co2 storage. *International Journal of Greenhouse Gas Control*, 9:234–242, 2012. doi: 10.1016/j.ijggc.2012.03.007. cited By 78.

[16] S.M. Benson and L. Myer. 2003.

[17] E. Keating, D. Bacon, S. Carroll, K. Mansoor, Y. Sun, L. Zheng, D. Harp, and Z. Dai. Applicability of aquifer impact models to support decisions at co2 sequestration sites. *International Journal of Greenhouse Gas Control*, 52:319–330, 2016. doi: 10.1016/j.ijggc.2016.07.001.

[18] J. Condor, D. Unatrakarn, M. Wilson, and K. Asghari. A comparative analysis of risk assessment methodologies for the geologic storage of carbon dioxide. volume 4, pages 4036–4043, 2011. doi: 10.1016/j.egypro.2011.02.345.

[19] L. De Lary, J.-C. Manceau, A. Loschetter, J. Rohmer, O. Bouc, I. Gravaud, C. Chiaberge, P. Willaume, and T. Yalamas. Quantitative risk assessment in the early stages of a co2 geological storage project: Implementation of a practical approach in an uncertain context. *Greenhouse Gases: Science and Technology*, 5(1):50–63, 2015. doi: 10.1002/ghg.1447.

[20] Q. Li and G. Liu. *Risk assessment of the geological storage of CO2: A review*. 2016. doi: 10.1007/978-3-319-27019-7_13. cited By 39.

[21] J.-P. Nicot, C.M. Oldenburg, J.E. Houseworth, and J.-W. Choi. Analysis of potential leakage pathways at the cranfield, ms, u.s.a., co2 sequestration site. *International Journal of Greenhouse Gas Control*, 18:388–400, 2013. doi: 10.1016/j.ijggc.2012.10.011. cited By 38.

671–687, 2013. doi: 10.1007/s10596-013-9350-6. cited By 36.

[107] Mingliang Liu and Dario Grana. Petrophysical characterization of deep saline aquifers for co2 storage using ensemble smoother and deep convolutional autoencoder. *Advances in Water Resources*, 142, 8 2020. ISSN 03091708. doi: 10.1016/j.advwatres.2020.103634.

[108] Mohamed Mehana, Bailian Chen, and Rajesh Pawar. Reduced-order models for wellbore leakage from depleted reservoirs. Unconventional Resources Technology Conference (URTEC), 2022. doi: 10.15530/urtec-2022-3725868.

[109] Rajesh Pawar, Shaoping Chu, Bill Carey, David Tu, Nathan Moodie, Bailian Chen, and William Ampomah. Quantitative risk assessment of leakage through legacy wells in support of permit application for a large-scale co 2 injection project in southwestern us, 2022.

[110] Bailian Chen, Dylan R. Harp, Youzuo Lin, Elizabeth H. Keating, and Rajesh J. Pawar. Geologic co2 sequestration monitoring design: A machine learning and uncertainty quantification based approach. *Applied Energy*, 225:332–345, 9 2018. ISSN 03062619. doi: 10.1016/j.apenergy.2018.05.044.

[111] J.H. Friedman. Multivariate adaptive regression splines. *Annals of Statistics*, 19(1):1–141, 1991.

[112] D.H. Le and A.C. Reynolds. Optimal choice of a surveillance operation using information theory. *Computational Geosciences*, 18(3-4): 505–518, 2014. doi: 10.1007/s10596-014-9401-7.

[113] R.L. Iman. Latin hypercube sampling. *Latin Hypercube Sampling*, 2008.

[114] J. C. Helton and F. J. Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23–69, 7 2003. ISSN 0951-8320. doi: 10.1016/S0951-8320(03)00058-9.

[115] G.A. Zyvoloski, B.A. Robinson, Z.V. Dash, and L.L. Trease. Summary of the models and methods for the fehm application - a finite-element heat- and mass-transfer code. *Rep. LA-13307-MS*, 1997. cited By 165.

[116] Y. Xu and R. Goodacre. On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing*, 2(3):249–262, 2018. cited By 311.

[117] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[118] François Chollet et al. Keras. `https://keras.io`, 2015.

[119] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

[120] J. Caers. *Petroleum Geostatistics*, 2005. cited By 186.

[121] B. Chen and A.C. Reynolds. Optimal control of icv's and well operating conditions for the water-alternating-gas injection process. *Journal of Petroleum Science and Engineering*, 149:623–640, 2017. doi: 10.1016/j.petrol.2016.11.004.

[122] D. George, A. Kuprat, N. Carlson, and C. Gable. *LaGriT - Los Alamos Grid Toolbox*, 1999. cited By 6.

[123] S. Geisser. *Predictive Inference: An Introduction*, 1993.

[124] Diana Koschel, Jean-Yves Coxam, Laurence Rodier, and Vladimir Majer. Enthalpy and solubility data of co2 in water and nacl (aq) at conditions of interest for geological sequestration. *Fluid phase equilibria*, 247(1-2):107–120, 2006.

[125] G. Evensen. *Data assimilation: The ensemble kalman filter*. 2009. doi: 10.1007/978-3-642-03711-5. cited By 1238.

[126] A.A. Emerick and A.C. Reynolds. Ensemble smoother with multiple data assimilation. *Computers and Geosciences*, 55:3–15, 2013. doi: 10.1016/j.cageo.2012.03.011. cited By 618.

Misael M. Morales is a PhD student in the Hildebrand Department of Petroleum and Geosystems Engineering and a Graduate Research Assistant in the Digital Reservoir Characterization Technology and Formation Evaluation Consortia at the University of Texas at Austin. He obtained a MSc in Mathematics and BSc in Petroleum Engineering and Applied Mathematics from the University of Tulsa. His research focuses on machine learning applications for subsurface energy resource modeling, characterization, and simulation, specifically in dimensionality reduction techniques for flow forecasting, data assimilation, uncertainty quantification and optimization.

Mohamed Mehana, PhD, is a research scientist at the energy and natural resources security group, environmental and earth sciences division, Los Alamos National Lab. He received his **PhD** in petroleum engineering from the University of Oklahoma working on multiscale modeling of subsurface phenomena. His research interests include modeling and simulation of transport and interfacial phenomena at the meso/nanoscale, and leveraging machine learning capabilities to enable a predictive concurrent scale-bridging and real-time reservoir management.

Bailian Chen, PhD, is a staff scientist at Los Alamos National Laboratory. His research interests include reservoir optimization (well placement/control optimization, pilot/surveillance optimization), assisted history matching, uncertainty quantification, machine learning, reduced-order modeling, and geologic carbon storage. He holds a PhD degree from The University of Tulsa, and bachelor's and master's degrees from the China University of Petroleum (East China), all in petroleum engineering. He has published over ~~30~~ 60 technical papers, and delivered more than ~~20~~ 40 presentations/invited talks at various conferences and universities. He serves as an associate editor for ~~Journal of Petroleum Science & Engineering~~ Geoenergy Science & Engineering and ~~SPE Reservoir Evaluation & Engineering~~ SPE Journal