LA-UR-19-23466 (Accepted Manuscript)

Frankenstein's ROMster: Avoiding pitfalls of reduced-order model development

Chen, Bailian
Harp, Dylan Robert
Pawar, Rajesh J.
Stauffer, Philip H.
Viswanathan, Hari S.
Middleton, Richard Stephen

To be published in: International Journal of Greenhouse Gas Control

DOI to publisher's version: 10.1016/j.ijggc.2019.102892

Permalink to record: http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-19-23466

Frankenstein's ROMster: Avoiding Pitfalls of Reduced-Order Model Development

Bailian Chen, Dylan R. Harp, Rajesh J. Pawar, Philip H. Stauffer,
Hari S. Viswanathan, and Richard S. Middleton*

Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87544, USA

**Corresponding author, E-mail:* rsm@lanl.gov

Abstract

Reduced-order models (ROMs) are a widely used and powerful approach to reducing the complexity of predictive physics-based numerical simulations for a wide range of applications, including electronics and fluid mechanics such as geologic CO_2 sequestration (GCS). ROMs are critical for optimization, sensitivity analysis, model calibration and uncertainty quantification where full-order models cannot be feasibly executed many times. Traditional approaches generate a single ROM for each simulated response (e.g., CO_2 injection rates, pH changes) based on a set of training simulations. Here, we demonstrate that a single ROM can display excellent overall predictive statistics, but have predictions that dramatically and unacceptably deviate from simulator responses especially when the response variable has a large range (i.e., vary over multiple orders of magnitude). For example, we show that a traditional statistically-high-performing GCS ROM (coefficient of determination R^2 of 0.99) can have average absolute relative errors of over 200%. To address this, we propose a new and novel approach where a set of sub-ROMs are generated to overcome the potential pitfalls in traditional single ROM development. The effectiveness of the proposed approach—the ROMster framework—is demonstrated using a case study of predicted CO_2 injection rates for GCS. We find our approach is a robust and general framework for ROM development, reducing the average "error" from 200% to only 4% in our case study.

Keywords: Reduced-order models; Potential pitfall; Traditional ROMs; Combing ROMs; Geologic CO_2 sequestration

Introduction

Geologic CO_2 sequestration in reservoirs—such as deep saline aquifers, depleted oil/gas reservoirs and the deep ocean—is being considered as an important method to reduce anthropogenic greenhouse gas emissions to the atmosphere (Saito et al., 2000; Stauffer et al., 2011; Elliot and Celia, 2012; Middleton et al., 2012; Bachu, 2015; Teng and Zhang, 2018; Shabani and Vilcáez, 2018; Cui et al., 2018). Many insights into safe and secure sequestration of CO_2 have been gained through laboratory experiments (Wollenweber et al., 2009; Kneafsey and Pruess, 2010; Newell and Carey, 2012; Börner et al., 2013; Cao et al., 2015). However, for predictions in large-scale GCS operations, such as the assessment of CO_2/brine leakage rates along abandoned wellbores and the impact of leaking CO_2 in shallow aquifers, the only option is to rely on numerical reservoir simulations (Pruess et al., 2001; Kumar et al., 2005; Nordbotten and Celia, 2011; Hou et al., 2014; Carroll et al., 2014; Keating et al., 2016b). These large-scale reservoir simulations often require computationally expensive multi-phase, multi-component numerical simulations that result in highly nonlinear system behavior. Due to a large number of uncertainties—uncertain reservoir permeability, reservoir porosity and unknown leak locations—the forecast will be inherently uncertain. Thus, it is necessary to develop computationally efficient reduced-order models (ROMs) to make reliable predictions that fully account for potential uncertainties. Often, thousands of Monte Carlo realizations may be needed to bound system uncertainty.

In GCS, the response variables such as CO_2 injection rate, CO_2/brine leakage rates, and CO_2 plume area usually have large ranges (i.e., vary over multiple orders of magnitude) and, as a result, the corresponding ROM will have to cover a large predictive range. To the best of our knowledge, all previous ROMs development in GCS (Viswanathan et al., 2008; Stauffer et al., 2008; Pawar et al., 2013; Dai et al., 2013; Carroll et al., 2014; Schuetter et al., 2014; Pawar et al., 2016; Zhang et al., 2016; Keating et al., 2016a; Guyant et al., 2016; Sun and Tong, 2017; Chen et al., 2018; Jia et al., 2018) use only a single ROM covering the entire predictive range for the objective of interest (e.g., CO_2 leakage rates through wellbores, pH in shallow aquifers). When the predictive capacity is evaluated for such ROMs, such as through cross-validation, an acceptable predictive statistic (a high R^2 value) is usually achieved. We hypothesize that a traditional single ROM which covers a relatively large predictive range (i.e., vary over multiple orders of magnitude), although statistically performing well over the entire predictive space, could perform poorly particularly for

critical parts of the prediction space. Evidence to support this hypothesis can be found in the literature for GCS examples including CO_2 and brine leakage. For instance, Jordan et al. (2015) generated ROMs for CO_2 leakage rates to the surface and an overlying aquifer with strongly significant the R^2 values (0.9986 and 0.9968 respectively), suggesting excellent ROM performance (see Figure 1). However, a re-analysis of the ROM prediction of leakage shows poor predictive performance at low rates (10^{-6}–10^{-4} kg/s) compared with higher leakage rates (10^{-3}–10^{-1} kg/s). Visually, this can be seen as many points in the low range falling far from the 1:1 prediction line. Further, the average absolute relative errors are 314.40% and 972.84% respectively. In this case, the R^2 weights high rates more than low rates, and therefore they are inaccurate at low rates. As a result, ROM performance arguably should be measured through other statistics, such as average absolute relative error for prediction, and not just an R^2. Similar behavior is demonstrated in Harp et al. (2016) who developed ROMs of transient CO_2 and brine leakage along abandoned wellbores from GCS reservoirs with R^2 values for all ROMs greater than 0.95. However, the scatter in the plots of simulator response versus ROM predictions for CO_2 and water flow rates show significant deviation for low values. A third example is Keating et al. (2016b) who constructed ROMs for assessing CO_2 impacts in shallow unconfined aquifers. Again, only a single ROM is developed for each simulator response of interest; for example only one ROM was developed for the prediction of pH plume volume in shallow aquifers. A reanalysis of the ROMs shows more validation points fell into a lower range than in a higher range (see Figure 2). In each of these cases, we postulate that the *practically* poor performance is due to building only a single ROM for each simulator response of interest across the entire predictive range. In other words, building a single ROM often requires compromising between matching one interval of the range of responses well at the expense of another interval because it is impossible to obtain a good match across the entire range.

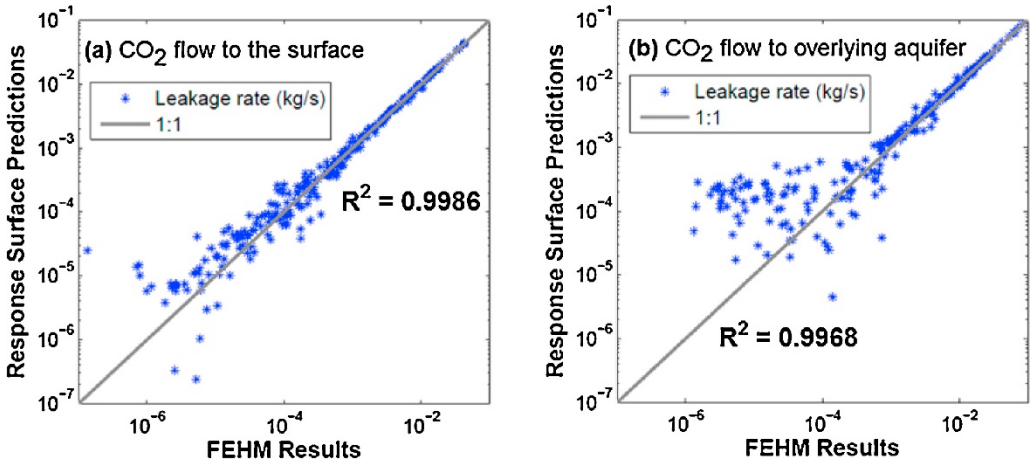


Figure 1. Performance of ROMs for predicting CO_2 leakage as compared to full FEHM simulations for (a) leakage to the surface and (b) leakage to an overlying aquifer. (Jordan et al., 2015)

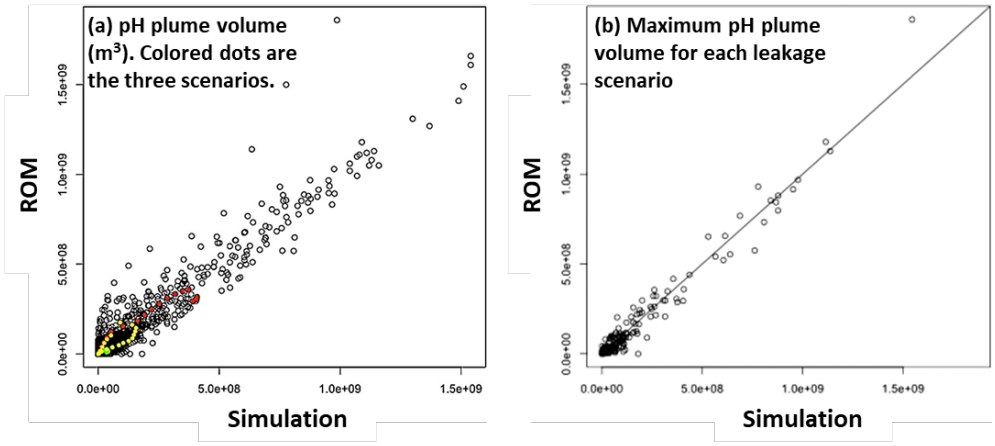


Figure 2. Performance of ROMs for predicting pH plume volume compared to full simulations. (Adapted from Keating et al. (2016b))

ROMs are also widely used in oil and gas applications. Note ROMs are also called response surface models, proxy models, meta-models, or surrogate models. White et al. (2001) applied polynomial regression to develop response surface models for sensitivity analysis and estimation of probability distributions of geologic model parameters. Peng and Gupta (2004) utilized kriging and bootstrapping methods to build surrogate models to quantify hydrocarbon in-place probability distribution curves. Yeten et al. (2005) compared different algorithms such as polynomials, kriging, splines and neural networks to construct proxy models for uncertainty quantification and sensitivity analysis. Osterloh (2008) used a kriging method to construct meta-models to assist

reservoir simulation forecasting and history matching. Zubarev (2009) compared different reduced-order modeling methods including polynomial regression, kriging, splines and neural networks as an alternative to a full reservoir simulation approach in history matching, production optimization and forecasting. Luo et al. (2018) utilized neural networks to develop ROMs based on data from the Bakken shale and conducted production optimization with the constructed ROMs. Chen and Pawar (2019) compared different machine learning algorithms including Multivariate Adaptive Regression Splines (MARS), support vector regression (SVR), and random forests to construct ROMs for the estimation of CO_2 storage and oil recovery potentials in residual oil zones. After tuning the parameters of three different algorithms, they showed that MARS gives the highest predictive accuracy for calculating CO_2 storage capacity and oil recovery.

In this study, we show that building a single ROM for each simulator response is indeed a real and potentially massive pitfall especially when the response variable has a large range (i.e., varies over multiple orders of magnitude). It is important to note that when the variation of response variable is not significant, one may not observe the poor performance of a single ROM (Anbar, 2010; Wriedt et al., 2014; Schuetter and Mishra, 2015). Specifically, we develop a GCS ROM for CO_2 injection rates that has excellent overall statistical properties (R^2 >0.99) over the entire training range but with almost zero predictive value (R^2 <0.15) for specific ranges of interest (injection rates ≤1 $MtCO_2$/year) and overall average absolute relative errors of 261%. We then introduce and demonstrate a novel approach that "Frankenstein's" or pieces together separate sub-ROMs (hence "Frankenstein's ROMster") and packages them into a single ROMster framework. This new approach requires developing an approach to select the sub-ROM most appropriate for the desired predictive range. The ROMster is able to reduce the average absolute relative error to only 4%. We also compare our new ROMster approach with the more traditional approach of transforming training outputs into log space before creating a single ROM. Although this has not been successful for other GCS related applications, the log transform approach performs very well for our GCS example, reducing absolute relative error to 8% with a single ROM.

Materials and Methods

To simulate CO_2 sequestration in saline aquifers, we employ a cylindrical reservoir model which we represent using a 2D radial model indicated in orange in Figure 3. The maximum radius of this

model is 25km. The input parameters for the ROM are geothermal gradient (geo_grad), storage reservoir depth, reservoir permeability, porosity and thickness. Parameter ranges are listed in Table 1. To train the ROMs, we use a ±10% range for each parameter to enhance the performance; ROM performance is known to begin to deteriorate as they approach their lower and upper parameter ranges. All the parameters are assumed to be uniformly distributed. The initial pressure (P; MPa) and temperature (T; ºC) fields are initialized based on depth, which are given by the following equations:

$$P = 0.101325 + 0.00981 * \text{depth}, \quad (1)$$

$$T = 4.4 + \text{geo\_grad} * \text{depth}. \quad (2)$$

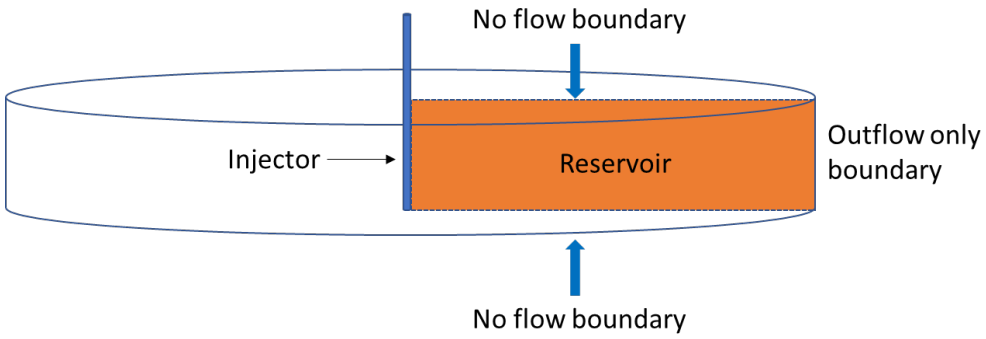


Figure 3. Simplified illustration of the 2D radial model used to represent a radially-symmetric 3D cylindrical model.

Table 1. Uncertain parameter and its range for reservoir model.

Uncertain parameters	User		ROMs development		Units
	Minimum	Maximum	Minimum	Maximum	
Geothermal gradient	0.015	0.045	0.0135	0.0495	$℃/m$
Depth	1000	5000	900	5500	m
Permeability	1	1000	0.9	1100	mD
Porosity	0.05	0.4	0.03	0.44	\
Thickness	5	100	3	110	m

We generated 10,000 parameter combinations using Latin hypercube Sampling (LHS) (Iman, 2008). The LHS method has been widely used to generate different parameter combinations for

building ROMs in many applications including CO_2 sequestration and oil/gas reservoir management communities (Kalla and White, 2007; Ekeoma and Appah, 2009; Maschio et al., 2010; Zhang and Sahinidis, 2012; Sun et al., 2013; He et al., 2016a; He et al., 2016b; Chen et al., 2017; Ampomah et al., 2017). The numerical mesh for the reservoir simulation model (2D radial model) was generated using the LaGriT grid generation toolkit (https://lagrit.lanl.gov) (George et al., 1999). All the full-order training simulations were performed using the Finite Element Heat and Mass transfer simulator (FEHM; https://fehm.lanl.gov) (Zyvoloski et al., 1997). CO_2 is injected at a constant maximum pressure (P_{inj}) equivalent to 80% of lithostatic pressure (Deng et al., 2012). Lithostatic pressure (P_{lith}) in MPa is computed by (Stauffer et al., 2008)

$$P_{lith} = 0.101325 + [\rho_w * \emptyset + \rho_r * (1 - \emptyset)] * 9.81 * \text{depth} * 10^{-6}, \quad (3)$$

where the average porosity ($\emptyset$) above the storage reservoir is assumed to be 0.15; $\rho_w = 1{,}000$ and $\rho_r = 2{,}650$, respectively, are water and rock density (kg/m^3). The LHS and concurrent simulation execution and response collection were performed using the Model Analysis ToolKit (MATK) Python package (https://matk.lanl.gov). A Multivariate Adaptive Regression Splines (MARS) algorithm (Friedman, 1991), which has been widely utilized to develop ROMs in GCS related research (Harp et al., 2016; Jordan et al., 2015; Keating et al., 2016b; Keating et al., 2016a; Chen et al., 2018), was implemented to build the ROM for average CO_2 injection rate (equal to the cumulative mass of CO_2 injection divided by injection time period). Here, we give a brief summary of the main steps for the MARS based ROM development. Step 1: Perform experimental design using the LHS method to generate 10,000 training samples. Step 2: Perform training simulations with each of the training samples using the FEHM simulator. Step 3: Collect inputs (parameter combinations) and outputs (CO_2 injection rates) from the training simulations. Step 4: Generate ROMs for the response variable (i.e., CO_2 injection rate) using MARS. Step 5: Validate the predictive accuracy of the ROMs.

Results and Discussion

The single CO_2 injection ROM was developed on 9,470 successful training simulations out of 10,000. The failed simulations were mainly caused by infeasible combinations of input parameters (e.g., very high permeability combined with a low porosity). Each of the 10,000 FEHM simulations

took approximately five hours with a total solution time of around two weeks on multiple CPUs. Some individual FEHM runs took many hours or several days to run. The resultant ROMs can produce 100,000 or more simulations in a single second on one CPU. The performance of the single CO_2 injection rate ROM is shown in Figure 4(a). Almost all the points in the plot are on or very close to the 1:1 reference line with an extremely high coefficient of determination (R^2 of 0.9874). It would typically be concluded that the single ROM has high fidelity compared to the simulations. However, when the same predictions are displayed into a log-log domain (Figure 4(b)), there is a clear discrepancy between the actual numerical results and the ROM predictions, especially when the injection rate is less than 10 $MtCO_2$/year. Postprocessing these numerical results and ROM predictions by translating them into log-log space reduces the R^2 to 0.8331 (note this value is calculated based on the data in log-log space). Thus, though the developed ROM displays great overall predictive statistics, this does not guarantee that it is a good CO_2 injection rate predictor in the real world. Building ROMs based on log-transformed outputs (i.e., the CO_2 injection rates outputted by FEHM are first transformed and then the ROM is built) improves the single-ROM performance significantly but still does not acceptably remedy the situation since this shifts the problem to larger values (see Figure 5) and makes the ROM too sensitive to lower values which are inherently less accurate in physics-based models.

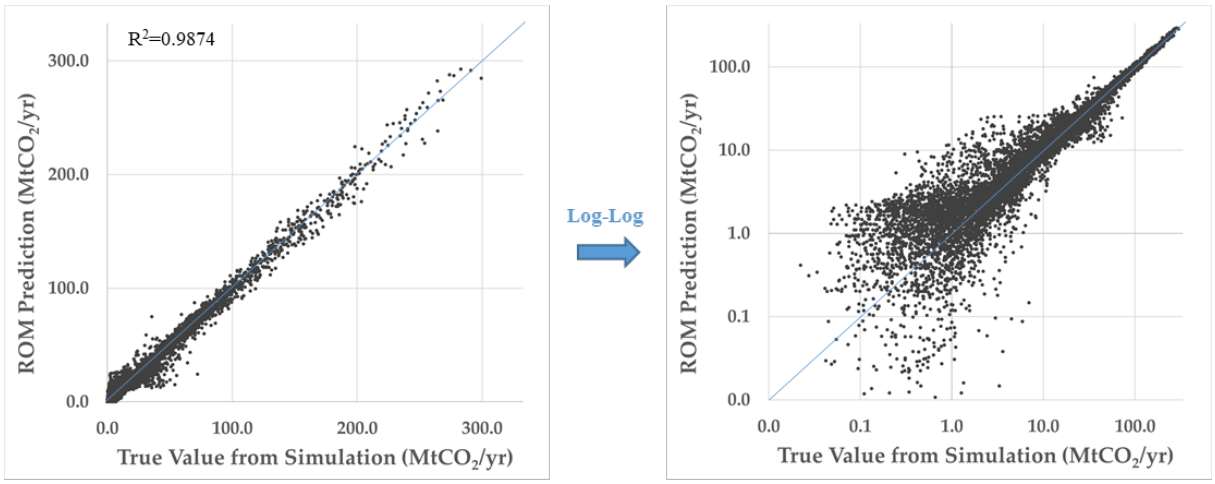


Figure 4. Performance of the ROM for predicting CO_2 injection rate as compared to full simulation results plotted in (a) the real domain and (b) the log-log domain.

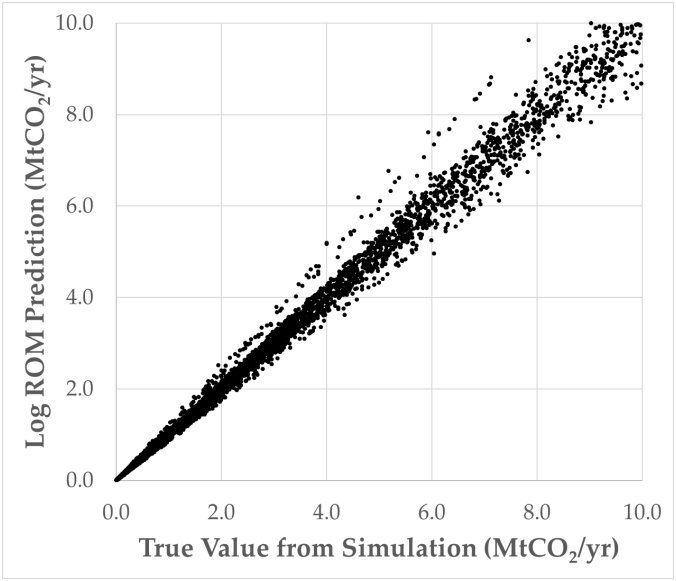


Figure 5: Performance of a single ROM developed in log-space as compared to full simulation results. Note only the practical injection range (0→10 $MtCO_2$/year) instead of 0→max is shown in this figure to demonstrate that building a ROM based on log-transformed outputs shifts the errors to larger values.

To address this ROM pitfall, we introduce a new approach that combines sub-ROMs to make accurate predictions across the whole range. We consider four key predictive ranges (Table 2) to produce four sub-ROMs. The practical CO_2 injection rate is usually less than 10 $MtCO_2$/year, while the most practical injection rate in field is about 0.1→1 $MtCO_2$/year (Birkholzer and Zhou, 2009; Zhou et al., 2010; Celia et al., 2015). The four sub-ROMs are split into four predictive ranges: 0→0.1, 0.1→1, 1→10 and 10 to the highest injection rate ($MtCO_2$/year). The number of samples used to develop each sub-ROM are listed in Table 2. The sub-ROMs for different predictive ranges, respectively, are denoted by ROM A, ROM B, ROM C and ROM D. The ROM for the entire predictive range is denoted by ROM 0. The summation of the number of samples (in Table 2) used to generate ROMs A, B, C and D is equal to the number of successful training simulations (9,470). The coefficient of determination (R^2) for the sub-ROMs are also listed in Table 2. The R^2-values for each sub-ROM is higher than for ROM 0. The performance of the four sub-ROMs is shown in Figure 6.

Table 2. Statistics for the single normal-space ROM (0) and four sub-ROMS (A–D).

ROM name	Predictive range ($MtCO_2$/year)	Samples	R^2

0	0→max	9470	0.9874
A	0→0.1	475	0.9949
B	0.1→1	2505	0.9920
C	1→10	3266	0.9927
D	10→max	3224	0.9948

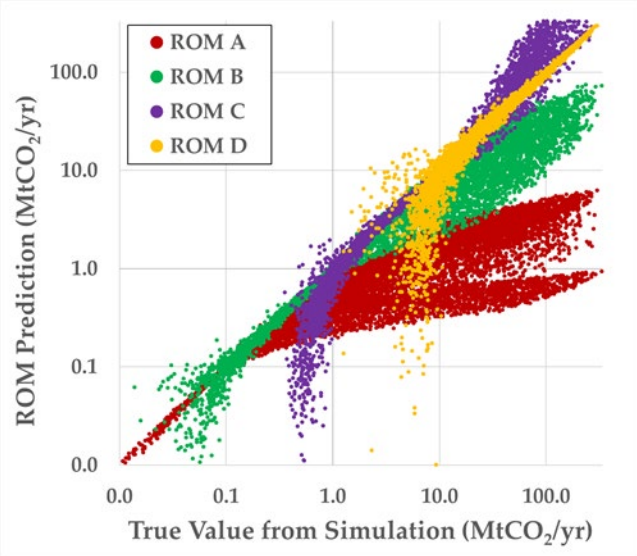


Figure 6. Performance of the four sub-ROMs.

Although the four sub-ROMs perform exceptionally well for their predictive ranges, they perform very poorly outside of their trained ranges. Consequently, the challenge is to decide which sub-ROM to use *a priori*. That is, given values for the five input parameters (i.e., geothermal gradient, depth, permeability, porosity, and thickness) and without knowing what the resultant injection rate is, which sub-ROM should be chosen? Figure 6 shows that picking the incorrect ROM is likely to result in errors similar or worse than using only a single ROM for the entire predictive space. In this case, each sub-ROM performs monotonically in the sense that picking a "too low" sub-ROM always results in a predicted injection rate above that sub-ROMs range. For example, imagine a set or parameters that lead to a 15 $MtCO_2$/year injection rate in the numerical model and thus should use ROM D (trained on 10→max $MtCO_2$/year). Those same parameters plugged into ROM A (0→0.1 $MtCO_2$/year), ROM B (0.1→1 $MtCO_2$/year), and ROM C (1→10

$MtCO_2$/year) always result in a predicted injection rate higher than the upper trained bound for each sub-ROM.

The following procedure is proposed for combining the sub-ROMs, resulting in a final framework (or ROMster) that accurately calculates injection rates across the entire predictive range. For a reservoir with given reservoir parameters, insert the parameter values into ROM A (trained under 0→0.1 $MtCO_2$/year). If the predicted value is within ROM A's range, then this value is accepted as the predicted injection rate. If the predicted value is greater than 0.1 $MtCO_2$/year, then the framework automatically applies ROMs B, C and D sequentially repeating the procedure conducted for ROM A until the calculated value falls into the predictive range of the applied ROM. In practice, the framework can pick an upper range for each ROM larger than the trained range. For example, ROM A performs very well up to 0.15 $MtCO_2$/year, 50% higher than its training range.

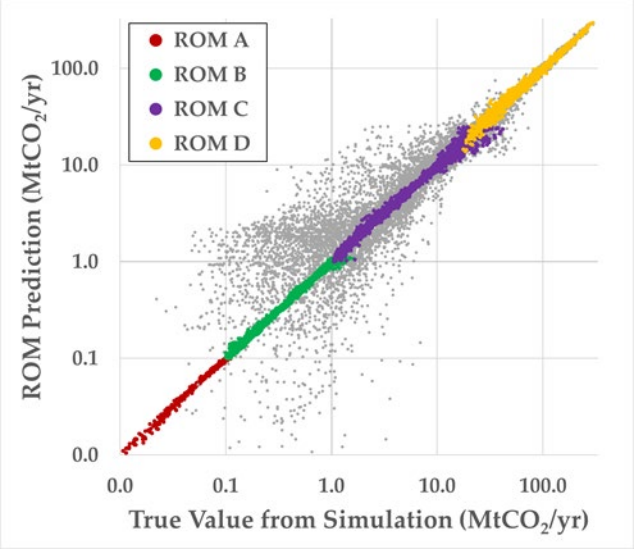
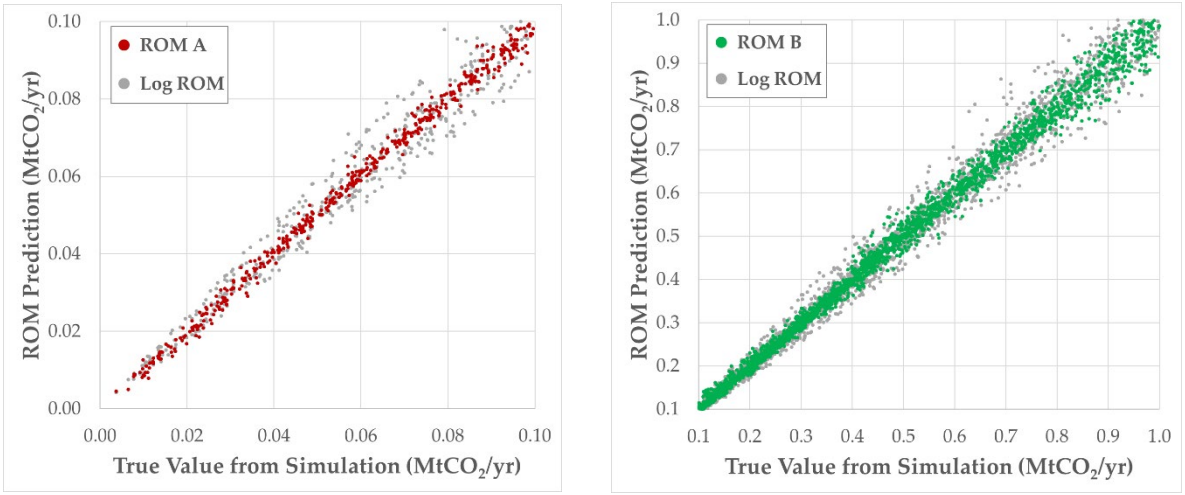


Figure 7. The performance of combining sub-ROMs. The grey points correspond to the performance of ROM 0 (predictive range from 0 $MtCO_2$/year to the largest injection rate).

Figure 7 illustrates the performance of the new approach by combining the four sub-ROMs. All the points (red, green, purple and yellow for ROMs A, B, C, and D respectively) for the combined sub-ROM predictions versus simulator response values are on or very close to the 1:1 reference line, indicating accurate performance across the entire predictive range. Note that the grey points

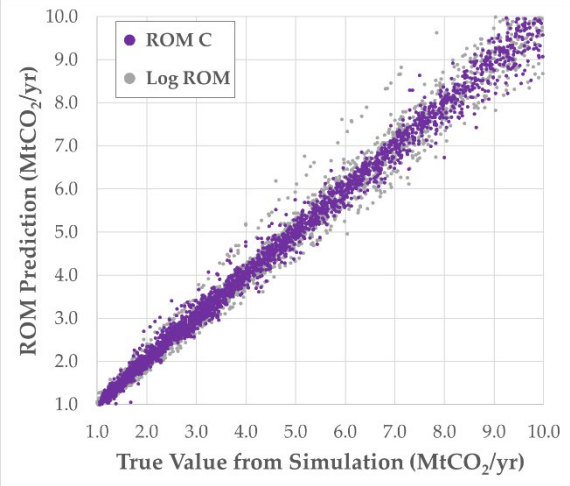
254 in Figure 7 correspond to the performance of ROM 0 (predictive range from 0 $MtCO_2$/year to the
255 highest injection rate). A common pitfall is to build only a single ROM (ROM 0) to cover the
256 entire predictive range, but results show that the average absolute relative error for the predictions
257 using ROM 0 is 260.99%, while the average absolute relative error for the predictions using the
258 ROMster approach is only 4.34%.

259 The performance of combining sub-ROMs in the practical injection range (i.e., 0 to 10
260 $MtCO_2$/year) is also compared with the performance of the ROM based on log-transformed
261 outputs from FEHM (Log ROM). Figure 8 shows that the predictions from combined ROMs
262 versus simulator response values in all different regions are much closer to the 1:1 reference line
263 than the predictions for the Log ROM versus simulator responses. The average absolute relative
264 errors for the combined ROM predictions in the three regions (i.e., 0→0.1, 0.1→1, 1→10
265 $MtCO_2$/year), respectively, are 3.16%, 3.11% and 3.96%. The corresponding errors obtained
266 from Log ROM are 7.33%, 5.07% and 4.69% respectively. This indicates that the combining
267 sub-ROMs (ROMster) has a better performance than the Log ROM.



268

269 (a) ROM A vs. Log ROM (b) ROM B vs. Log ROM



(c) ROM C vs. Log ROM

Figure 8. Comparison between the combined ROMs and the Log ROM.

Implications and Conclusions

The development of ROMs is critical for optimization, sensitivity analysis, model calibration, uncertainty quantification and risk assessment in a wide range of applications including CO_2 sequestration. However, a common pitfall in ROM generation is the tendency to build only a single ROM to cover the entire predictive range; the actual predictive error for single ROM can be unacceptably large especially when the response variable has a large range (i.e., vary over multiple orders of magnitude). Although we have only demonstrated this for a specific GCS example, we believe that many of the lessons learned will be applicable to other applications. We demonstrate a new approach to building a smart ROM that accurately estimates CO_2 injection across the entire predictive range and thus avoids this pitfall. Although our CO_2 injection framework relied on the monotonicity of the four sub-ROMS, other approaches are possible including developing a separate partial-space ROMs (e.g., a ROM based on 0 to 1 $MtCO_2$/year parameters) to predict which sub-ROM should be used or simultaneously analyzing the output from all sub-ROMs for a set of input parameters to estimate the correct sub-ROM to use. We believe that many researchers in GCS, hydrology, petroleum and other communities have not realized this potential pitfall and thus the ROMster framework will provide a powerful approach for future ROM development.

Acknowledgements

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 is funded by the US DOE's Fossil Energy Office through the "Southwest Regional Partnership on Carbon Sequestration (SWP)" (Award No. DE-FC26-05NT42591) and "*SimCCS*: Development and Applications" (Award No. FE-1017-18-FY18), and through the "US-China Clean Energy Research Center (CERC) Advanced Coal Technology Consortium (ACTC)" (Award No. DE-PI0000017).

References

Ampomah, W. et al., 2017. Optimum design of CO_2 storage and oil recovery under geological uncertainty. Applied Energy, 195: 80-92.

Anbar, S., 2010. Development of a predictive model for carbon dioxide sequestration in deep saline carbonate aquifers, SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Bachu, S., 2015. Review of CO_2 storage efficiency in deep saline aquifers. International Journal of Greenhouse Gas Control, 40: 188-202.

Birkholzer, J.T. and Zhou, Q., 2009. Basin-scale hydrogeologic impacts of CO_2 storage: Capacity and regulatory implications. International Journal of Greenhouse Gas Control, 3(6): 745-756.

Börner, J.H., Herdegen, V., Repke, J.-U. and Spitzer, K., 2013. The impact of CO_2 on the electrical properties of water bearing porous media–laboratory experiments with respect to carbon capture and storage. Geophysical Prospecting, 61(sup1): 446-460.

Cao, P., Karpyn, Z.T. and Li, L., 2015. Self-healing of cement fractures under dynamic flow of CO2-rich brine. Water Resources Research, 51(6): 4684-4701.

Carroll, S.A. et al., 2014. Key factors for determining groundwater impacts due to leakage from geologic carbon sequestration reservoirs. International Journal of Greenhouse Gas Control, 29: 153-168.

Celia, M.A., Bachu, S., Nordbotten, J. and Bandilla, K., 2015. Status of CO_2 storage in deep saline aquifers with emphasis on modeling approaches and practical simulations. Water Resources Research, 51(9): 6846-6892.

Chen, B., Harp, D.R., Lin, Y., Keating, E.H. and Pawar, R.J., 2018. Geologic CO_2 sequestration monitoring design: A machine learning and uncertainty quantification based approach. Applied Energy, 225: 332-345.

Chen, B., He, J., Wen, X.-H., Chen, W. and Reynolds, A.C., 2017. Uncertainty quantification and value of information assessment using proxies and Markov chain Monte Carlo method for a pilot project. Journal of Petroleum Science and Engineering, 157: 328-339.

Chen, B. and Pawar, R.J., 2019. Characterization of CO_2 storage and enhanced oil recovery in residual oil zones. Energy, 183: 291-304.
Cui, G. et al., 2018. Assessing the combined influence of fluid-rock interactions on reservoir properties and injectivity during CO_2 storage in saline aquifers. Energy, 155: 281-296.
Dai, Z. et al., 2013. An integrated framework for optimizing CO2 sequestration and enhanced oil recovery. Environmental Science & Technology Letters, 1(1): 49-54.
Deng, H., Stauffer, P.H., Dai, Z., Jiao, Z. and Surdam, R.C., 2012. Simulation of industrial-scale CO_2 storage: Multi-scale heterogeneity and its impacts on storage capacity, injectivity and leakage. International Journal of Greenhouse Gas Control, 10: 397-418.
Ekeoma, E. and Appah, D., 2009. Latin hypercube sampling (lhs) for gas reserves, Nigeria Annual International Conference and Exhibition. Society of Petroleum Engineers.
Elliot, T. and Celia, M., 2012. Potential restrictions for CO_2 sequestration sites due to shale and tight gas production. Environmental science & technology, 46(7): 4223-4227.
Friedman, J.H., 1991. Multivariate adaptive regression splines. The annals of statistics: 1-67.
George, D., Kuprat, A., Carlson, N. and Gable, C., 1999. LaGriT–Los Alamos Grid Toolbox.
Guyant, E., Han, W.S., Kim, K.-Y., Park, E. and Yun, S.-T., 2016. Leakage and pressurization risk assessment of CO_2 reservoirs: A metamodeling modeling approach. International Journal of Greenhouse Gas Control, 54: 345-361.
Harp, D.R., Pawar, R., Carey, J.W. and Gable, C.W., 2016. 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.
He, J. et al., 2016a. Proxy-based work flow for a priori evaluation of data-acquisition programs. SPE Journal, 21(04): 1,400-1,412.
He, J., Xie, J., Wen, X.-H. and Chen, W., 2016b. An alternative proxy for history matching using proxy-for-data approach and reduced order modeling. Journal of Petroleum Science and Engineering, 146: 392-399.
Hou, Z. et al., 2014. Uncertainty analyses of CO_2 plume expansion subsequent to wellbore CO_2 leakage into aquifers. International Journal of Greenhouse Gas Control, 27: 69-80.
Iman, R.L., 2008. Latin hypercube sampling. Wiley StatsRef: Statistics Reference Online.
Jia, W., McPherson, B., Pan, F., Dai, Z. and Xiao, T., 2018. Uncertainty quantification of CO_2 storage using Bayesian model averaging and polynomial chaos expansion. International Journal of Greenhouse Gas Control, 71: 104-115.
Jordan, A.B., Stauffer, P.H., Harp, D., Carey, J.W. and Pawar, R.J., 2015. A response surface model to predict CO_2 and brine leakage along cemented wellbores. International Journal of Greenhouse Gas Control, 33: 27-39.
Kalla, S. and White, C.D., 2007. Efficient Design of Reservoir Simulation Studies for Development and Optimization. SPE Reservoir Evaluation & Engineering, 10(06): 629-637.
Keating, E. et al., 2016a. Applicability of aquifer impact models to support decisions at CO_2 sequestration sites. International Journal of Greenhouse Gas Control, 52: 319-330.
Keating, E.H., Harp, D.H., Dai, Z. and Pawar, R.J., 2016b. Reduced order models for assessing CO_2 impacts in shallow unconfined aquifers. International Journal of Greenhouse Gas Control, 46: 187-196.
Kneafsey, T.J. and Pruess, K., 2010. Laboratory flow experiments for visualizing carbon dioxide-induced, density-driven brine convection. Transport in porous media, 82(1): 123-139.
Kumar, A. et al., 2005. Reservoir simulation of CO_2 storage in aquifers. Spe Journal, 10(03): 336-

348.
Luo, G., Tian, Y., Bychina, M. and Ehlig-Economides, C., 2018. Production optimization using machine learning in Bakken shale, Unconventional Resources Technology Conference, Houston, Texas, 23-25 July 2018. Society of Exploration Geophysicists, American Association of Petroleum …, pp. 2174-2197.
Maschio, C., de Carvalho, C.P.V. and Schiozer, D.J., 2010. A new methodology to reduce uncertainties in reservoir simulation models using observed data and sampling techniques. Journal of Petroleum Science and Engineering, 72(1-2): 110-119.
Middleton, R.S. et al., 2012. The cross-scale science of CO_2 capture and storage: from pore scale to regional scale. Energy & Environmental Science, 5(6): 7328-7345.
Newell, D.L. and Carey, J.W., 2012. Experimental evaluation of wellbore integrity along the cement-rock boundary. Environmental science & technology, 47(1): 276-282.
Nordbotten, J.M. and Celia, M.A., 2011. Geological storage of CO_2: modeling approaches for large-scale simulation. John Wiley & Sons.
Osterloh, W.T., 2008. Use of multiple-response optimization to assist reservoir simulation probabilistic forecasting and history matching, SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.
Pawar, R. et al., 2013. Quantification of risk profiles and impacts of uncertainties as part of US DOE's National Risk Assessment Partnership (NRAP). Energy Procedia, 37: 4765-4773.
Pawar, R. et al., 2016. The National Risk Assessment Partnership's integrated assessment model for carbon storage: A tool to support decision making amidst uncertainty. International Journal of Greenhouse Gas Control, 52: 175-189.
Peng, C.Y. and Gupta, R., 2004. Experimental design and analysis methods in multiple deterministic modelling for quantifying hydrocarbon in-place probability distribution curve, SPE Asia Pacific Conference on Integrated Modelling for Asset Management. Society of Petroleum Engineers.
Pruess, K., Xu, T., Apps, J. and Garcia, J., 2001. Numerical modeling of aquifer disposal of CO_2, SPE/EPA/DOE Exploration and Production Environmental Conference. Society of Petroleum Engineers.
Saito, T., Kajishima, T. and Nagaosa, R., 2000. CO_2 sequestration at sea by gas-lift system of shallow injection and deep releasing. Environmental science & technology, 34(19): 4140-4145.
Schuetter, J., Ganesh, P.R. and Mooney, D., 2014. Building statistical proxy models for CO_2 geologic sequestration. Energy Procedia, 63: 3702-3714.
Schuetter, J. and Mishra, S., 2015. Experimental Design or Monte Carlo Simulation? Strategies for Building Robust Surrogate Models, SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.
Shabani, B. and Vilcáez, J., 2018. A fast and robust TOUGH2 module to simulate geological CO_2 storage in saline aquifers. Computers & Geosciences, 111: 58-66.
Stauffer, P.H. et al., 2011. Greening coal: breakthroughs and challenges in carbon capture and storage. Environmental Science & Technology, 45(20): 8597-8604.
Stauffer, P.H., Viswanathan, H.S., Pawar, R.J. and Guthrie, G.D., 2008. A system model for geologic sequestration of carbon dioxide. Environmental Science & Technology, 43(3): 565-570.
Sun, Y. and Tong, C., 2017. Dynamic reduced-order models of integrated physics-specific systems for carbon sequestration. Geomechanics and Geophysics for Geo-Energy and Geo-

Resources, 3(3): 315-325.
Sun, Y. et al., 2013. Global sampling for integrating physics-specific subsystems and quantifying uncertainties of CO_2 geological sequestration. International Journal of Greenhouse Gas Control, 12: 108-123.
Teng, Y. and Zhang, D., 2018. Long-term viability of carbon sequestration in deep-sea sediments. Science advances, 4(7): eaao6588.
Viswanathan, H.S. et al., 2008. Development of a hybrid process and system model for the assessment of wellbore leakage at a geologic CO_2 sequestration site. Environmental Science & Technology, 42(19): 7280-7286.
White, C.D., Willis, B.J., Narayanan, K. and Dutton, S.P., 2001. Identifying and estimating significant geologic parameters with experimental design. SPE Journal, 6(03): 311-324.
Wollenweber, J. et al., 2009. Caprock and overburden processes in geological CO_2 storage: An experimental study on sealing efficiency and mineral alterations. Energy Procedia, 1(1): 3469-3476.
Wriedt, J., Deo, M., Han, W.S. and Lepinski, J., 2014. A methodology for quantifying risk and likelihood of failure for carbon dioxide injection into deep saline reservoirs. International Journal of Greenhouse Gas Control, 20: 196-211.
Yeten, B., Castellini, A., Guyaguler, B. and Chen, W., 2005. A comparison study on experimental design and response surface methodologies, SPE Reservoir Simulation Symposium. Society of Petroleum Engineers.
Zhang, Y., Liu, Y., Pau, G., Oladyshkin, S. and Finsterle, S., 2016. Evaluation of multiple reduced-order models to enhance confidence in global sensitivity analyses. International Journal of Greenhouse Gas Control, 49: 217-226.
Zhang, Y. and Sahinidis, N.V., 2012. Uncertainty quantification in CO_2 sequestration using surrogate models from polynomial chaos expansion. Industrial & Engineering Chemistry Research, 52(9): 3121-3132.
Zhou, Q., Birkholzer, J.T., Mehnert, E., Lin, Y.F. and Zhang, K., 2010. Modeling basin-and plume-scale processes of CO_2 storage for full-scale deployment. Groundwater, 48(4): 494-514.
Zubarev, D.I., 2009. Pros and cons of applying proxy-models as a substitute for full reservoir simulations, SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.
Zyvoloski, G.A., Robinson, B.A., Dash, Z.V. and Trease, L.L., 1997. User's manual for the FEHM application-A finite-element heat-and mass-transfer code.