1 Reduced order models of transient CO₂ and brine
2 leakage along abandoned wellbores from geologic carbon
3 sequestration reservoirs

4 Dylan R. Harp, Rajesh Pawar, J. William Carey, Carl W. Gable*
5 *Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los*
6 *Alamos, NM, 87544*

7 **Abstract**

8 We have developed Reduced Order Models (ROMs) for CO₂ and brine leakage
9 rates along wellbores including abandoned wells at geologic CO₂ storage sites
10 using a Multivariate Adaptive Regression Splines (MARS) algorithm. The
11 ROMs were developed for use within systems level performance assessment
12 models such as Los Alamos National Laboratory's CO₂-PENS model. The
13 ROMs are used to compute leakage rates as a function of wellbore properties
14 including effective permeability, depth as well as pressures and saturations
15 in the reservoir where the wellbore intercepts the reservoir. The ROMs were
16 created using results of complex, 3-D multi-phase numerical simulations of
17 large-scale CO₂ injection at a generic CO₂ storage site with an abandoned
18 wellbore. The generic site included not only the primary storage reservoir
19 but also a groundwater aquifer and an intermediate permeable zone. Two
20 sets of simulations were performed, one with and one without an abandoned
21 wellbore in order to capture the effect of coupling between the storage reser-

*Corresponding author: Dylan R. Harp (dharp@lanl.gov)

Preprint submitted to International Journal of Greenhouse Gas Control November 14, 2015

22 voir and wellbore in a system level model where it is assumed that they
23 are decoupled. Cross-validation against the complex, multi-phase numerical
24 simulation results were used to evaluate the ability of the ROMs to repro-
25 duce numerical simulation results. Further our ROM development approach
26 effectively captures transient CO₂ and brine leakage during and after CO₂
27 injection as well as the effects of an intermediate permeable zone on leakage
28 to a shallow groundwater aquifer and to the atmosphere. Ultimately, the
29 ROM is a computationally efficient model that effectively captures many of
30 the complex underlying processes taking place during CO₂ and brine leakage
31 along a wellbore at a geologic CO₂ storage site.

32 *Keywords:* reduced-order model, wellbore, leakage, CO₂, geologic
33 sequestration

34 1. Introduction

35 Geologic carbon sequestration (GCS) is being considered as a method to
36 reduce CO₂ emissions to the atmosphere (Metz et al., 2005). As of 2014, there
37 were 21 large-scale GCS projects in operation or under construction and an
38 additional 39 in various stages of planning globally (Global CCS Institute,
39 2014). One of the key considerations in the development of any new GCS site
40 is an assessment of the risk of leakage of either CO₂ or brine (Environmental
41 Protection Agency, 2013). Leakage of CO₂ not only counteracts the purpose
42 of GCS, but can also threaten groundwater resources by mobilizing trace
43 contaminants (Wilkin and DiGiulio, 2010). Leakage of saline reservoir fluids

44 (brines) can degrade groundwater resources as well (Keating et al., 2010;
45 Gleason et al., 2011). Given the high number of abandoned wellbores from
46 petroleum exploration at many potential GCS sites (Metz, 2005), leakage
47 along abandoned wellbores is a primary concern.

48 Due to a lack of knowledge concerning the quality of past wellbore com-
49 pletions and the lack of direct testing of wellbore integrity, the leakage po-
50 tential of abandoned wellbores is uncertain. One of the least expensive ways
51 to quantify this uncertainty and assess its impact on the viability of a se-
52 questration site is by modeling leakage for an ensemble of possible scenarios.
53 However, assessing the risk of wellbore leakage at GCS sites can easily re-
54 quire thousands to millions of simulations. Some previous efforts to address
55 the wellbore leakage risks have focused on analytical solutions (Nordbotten
56 et al., 2005, 2008; Tao et al., 2011) or hybrid analytical/numerical approaches
57 (Gasda et al., 2009; Pan et al., 2011). Accurate models of GCS sites can re-
58 quire the inclusion of complicated hydrogeological, geochemical, geological,
59 and geomechanical factors. While numerical models provide the ability to
60 capture physical processes and geometrical details of a GCS site, they are
61 computationally demanding. As a result, systems-level models that use sim-
62 plified numerical representations have been developed to represent wellbore
63 leakage processes including those of Zhou et al. (2005); Viswanathan et al.
64 (2008); Oldenburg et al. (2009); Guen et al. (2011); Nicot et al. (2013); Jordan
65 et al. (2015); Bai et al. (2015) In this study, we reduce the computational
66 burden by developing an approach to extract computationally efficient re-

duced order models (ROMs) from a few hundred computationally intensive, 3D multi-phase numerical simulations. The ROMs allow a comprehensive analysis of project risk induced by uncertainty in wellbore leakage rates.

Multi-phase 3D numerical simulations of CO₂ injection and subsequent relaxation were performed using the Finite Element Heat and Mass (FEHM) code (Zyvoloski, 2007). We used the Multivariate Adaptive Regression Splines (MARS) algorithm (Friedman, 1991) to extract ROMs from the inputs and outputs of these simulations. MARS produces ROMs that capture trends in the numerically simulated leakage using computationally efficient algebraic models. MARS is less affected by nonlinear dependencies and input interactions than other response surface algorithms.

The ROMs were developed for use in the CO₂-PENS systems model (Viswanathan et al., 2008; Stauffer et al., 2008; Pawar et al., 2013), which is an Integrated Assessment Model (AIM) used to determine CO₂ storage risk profiles within the National Risk Assessment Partnership (NRAP). The ROMs described here build on those developed by Jordan et al. (2015) and include the following improvements: (1) effects of leakage on the reservoir pressures and saturations, (2) transient effects, (3) inclusion of a thief zone (permeable intermediate zone within the caprock), and (4) direct evaluation of MARS generated algebraic equations as opposed to lookup tables. Similar to ROMs developed in Jordan et al. (2015), flow for all phases is Darcian and the wellbore properties are considered homogeneous. The ROMs represent scenarios where the leakage paths in wellbores can be represented by Darcy

90 flow (e.g., flow through a permeable cement matrix) but also include scenar-
 91 ios in which leakage can be approximated by an effective Darcy permeability
 92 such as fracture flow. The use of the ROMs for cases with higher leakage
 93 rates associated with open wellbores would be inappropriate. Our definition
 94 of a thief zone is slightly more general than the common definition associ-
 95 ated with well development in that we do not require the thief zone to be
 96 under-pressurized. Our thief zones are initialized at hydrostatic pressures,
 97 which in general will be a conservative assumption.

98 The CO₂-PENS systems model is decoupled into reservoir, wellbore, and
 99 shallow aquifer models. The reservoir model provides reservoir pressures and
 100 saturations at the point where CO₂ enters the wellbore (i.e., along the top
 101 of the reservoir at the point of wellbore penetration) to the wellbore model.
 102 In turn, the wellbore model provides leakage rates to the aquifer model. The
 103 reservoir model does not include the effects of wellbore leakage on reservoir
 104 pressures and saturations. In order to include these effects, we ran paired
 105 simulations. One with a wellbore (coupled reservoir/wellbore/aquifer simu-
 106 lation) and one without (decoupled, reservoir-only simulation). Therefore,
 107 the coupled simulation contained the complete system while the decoupled
 108 model represented the reservoir only. We train the ROMs with reservoir pres-
 109 sures and saturations that they will receive as inputs in the system model
 110 and use flow rates that include leakage effects as ROM outputs. In this way,
 111 the wellbore ROMs capture coupled effects even though the reservoir satu-
 112 rations and pressures they receive do not include these effects. This assumes

113 that the wellbores are sufficiently spaced apart that there is no interaction
114 in leakage between nearby wellbores.

115 This paper is organized as follows. Methods of the research are discussed
116 in Section 2, including a description of the numerical model and mesh (Sec-
117 tion 2.1), our framework and workflow for ROM development (Section 2.2),
118 and the use of cross-validation (CV) to evaluate the predictive capacity of
119 the ROMs (Section 2.3). Results and analysis are presented in Section 3 in-
120 cluding correlation analysis between ROM inputs and outputs (Section 3.1),
121 ROM development diagnostics (Section 3.2), and MARS input importance
122 ranking (Section 3.3). Section 4 provides conclusions of the research. We
123 provide details on the process of ROM development using MARS in the ap-
124 pendix.

125 2. Methods

126 2.1. Description of the Mesh and Numerical Model

127 3D multi-phase simulations of heat and mass transfer of water and super-
128 critical, liquid, and gas CO₂ through leaking wells are performed using the
129 FEHM code (Zyvoloski, 2007). Dissolution, residual saturations, and brine
130 density were not considered, but would increase estimates of flow resulting
131 in conservative risk assessments. This conservatism will be counteracted to
132 some degree by neglecting brine density.

133 Figure 1 presents a schematic of the model with reservoir, thief zone,
134 aquifer, and caprock layers connected by a wellbore. The reservoir, thief

135 zone, aquifer, and upper caprock layer are 51.2, 22.4, 19.2, and 11.2 m thick,
 136 respectively. The total depth of the model is varied within the ensemble
 137 from 960 to 3200 m. These depths target scenarios where injected CO₂
 138 will remain supercritical, but well development will not be too costly. The
 139 location of the thief zone can vary from 30 to 70% of the model depth.
 140 Other than a column of caprock surrounding the wellbore and the layer
 141 above the aquifer, the caprock has been removed from the model. This
 142 reduces the memory requirements of the model. Therefore we do not consider
 143 flow within the caprock, however, heat flow is allowed within the column of
 144 caprock surrounding the wellbore. The caprock surrounding the wellbore is
 145 intended to allow transfer of heat due to phase change between the wellbore
 146 and caprock. In some realizations, the limited extent of the caprock may
 147 lead to slightly increased temperatures in the wellbore. The sides of the
 148 reservoir, thief zone, and aquifer are Dirichlet boundaries allowing outflow
 149 only of CO₂ and water at pressures above hydrostatic. The remainder of the
 150 boundaries are no-flow boundaries. The model is initialized to a geothermal
 151 gradient of 0.03 °C/m with a temperature of 20 °C along the top. Pressures
 152 are initialized to a vertical gradient of 9.81×10^{-3} MPa/m with a pressure of
 153 0.2 MPa along the top. CO₂ is injected for 20 years followed by 30 years of
 154 relaxation for a total simulation time of 50 years.

155 We used LaGriT (<http://lagrit.lanl.gov>) (George et al., 1999) to generate
 156 the meshes for the ensemble. Figure 2 presents a plan view of a representa-
 157 tive mesh from the ensemble. Vertices are computational nodes and lines are

Figure 1: Side view schematic of model not drawn to scale. ROM input parameter symbols are identified near their associated location in the schematic (refer to Table 1).

158 nodal connectivity. The mesh is composed of tetrahedral cells (finite volumes)
 159 associated with each node. The mesh is composed of a base mesh discretized
 160 at 204.8 m horizontally and 3.2 m vertically. The wellbore is comprised of a
 161 vertical line of nodes located at the center of the model at horizontal coordi-
 162 nates (0,0) in meters. The base mesh is octree refined horizontally along the
 163 wellbore to 0.1 m cells (refer to inset in Figure 2). This requires 11 orders
 164 of horizontal refinement ($2^{11} \times 0.1 \text{ m} = 204.8 \text{ m}$). This results in 20–40% of
 165 the mesh nodes residing in the area of refinement, depending on the reservoir
 166 depth. Octree refinement horizontally, but not vertically, reduces mesh ef-
 167 fects on the CO₂ saturation plume along the top of the reservoir (saturation
 168 with a buoyant plume are dependent on cell thickness). Therefore, we are
 169 representing the wellbore by a 10 cm square column of nodes. Injection of
 170 CO₂ is simulated along a vertical line of cells fully penetrating the reservoir
 171 at (-1024,-1024), approximately 1448 m from the wellbore.

172 As mentioned above, caprock nodes that do not significantly contribute to
 173 the flow rates along the wellbore are removed. This is shown schematically in
 174 Figure 1 and is apparent in Figure 3. All caprock between the reservoir and
 175 thief zone and thief zone and aquifer are removed except for a 6.4 m column
 176 along the wellbore. The caprock surrounding the wellbore is impermeable,
 177 but does transfer heat. A caprock layer is included above the aquifer and
 178 used in the initialization of the temperatures and pressures of the model, but
 179 is not included in subsequent wellbore leakage calculations.

180 In an example simulation, Figure 4 presents CO₂ saturations along the

Figure 2: Plan view of mesh with close-up of octree refinement at the wellbore

Figure 3: Angled view of mesh

181 borehole and along the top of the reservoir for a leaky well scenario. Close-
 182 ups of the reservoir, thief zone, and aquifer/upper caprock along the borehole
 183 are shown and the transient evolution of the CO_2 plume and wellbore leakage
 184 are given at 6, 15, 30 and 50 years. CO_2 is injected in supercritical phase,
 185 but changes to gas phase below the thief zone in this scenario.

Figure 4: CO₂ saturations along borehole and along the top of the reservoir at various times of an example simulation. CO₂ injection occurs for 20 years followed by 30 years of relaxation. The wellbore and thief zone permeabilities are $1.e-12$ and $1.e-14$ m², respectively. The model depth is 1,280 m and the depth of the thief zone is centered at 640 m.

186 2.2. ROM Development Framework and Workflow

187 The workflow for the ROM development is presented in Figure 5. The
188 workflow is comprised of (1) Latin Hypercube Sampling (LHS) of parameters,
189 (2) collecting inputs and outputs from simulations, (3) generating ROMs
190 using MARS, and (4) cross-validation (CV) of ROMs.

191 Table 1 lists the parameters sampled by LHS. The parameter ranges are
192 intended to cover a wide range of scenarios that may be encountered at GCS
193 sites. LHS from uniform distributions ensures that samples span and fill the
194 parameter space as effectively as possible. The sampled distributions are
195 not intended to represent uncertainties at potential GCS sites, but to en-
196 sure that ROMs are built with ample input/output combinations that cover
197 the potential storage scenarios for which ROMs will be used. As discussed
198 in Jordan et al. (2015), the range of wellbore permeabilities (k_w) represent
199 effective permeabilities that are intended to include effects of annular separa-
200 tion of materials. Therefore, these values are larger than available estimates
201 of degraded cement permeabilities (Carey et al., 2007; Crow et al., 2010).
202 Smaller values of k_w are not considered since they do not lead to significant
203 leakage. The thief zone permeability range was selected to cover a represen-
204 tative range of values leading to the capture of leaking CO₂. Smaller values
205 will lead to low levels of capture. Larger values are possible, but less likely
206 in most geologic settings. CO₂ injection rates ($\dot{m}_{inj}$) were selected to be
207 large enough to induce significant leakage and to approach, but not exceed
208 lithostatic pressure. The largest lithostatic pressure in the ensemble is ap-

Figure 5: ROM development flow diagram

proximately 12 MPa, while the overpressure (ΔP_{res}) does not exceed 10 MPa (refer to Figure 6).

All of these parameters are included as inputs to MARS except for the CO₂ injection rate, $\dot{m}_{inj}$. The purpose of varying $\dot{m}_{inj}$ is to allow variations in the transient pressures and CO₂ saturations at the reservoir/wellbore intersection to emerge through both the coupled and decoupled simulations. As illustrated in Figure 5, transient pressure changes (ΔP_{res}) and CO₂ saturations (S_{res}) at the top of the reservoir at the wellbore are collected from the reservoir simulation and are inputs to MARS. These values of pressures and saturations do not include leakage effects and are representative of the inputs that the ROMs will be provided in the systems model framework. As mentioned above, this allows coupled effects to be included in decoupled ROMs. Pressure changes are calculated relative to the initial pressure as $\Delta P_{res} = P_{res} - P_{res,0}$. ΔP_{res} is used in place of P_{res} as P_{res} and model depth parameter D are directly correlated. MARS outputs (CO₂ and brine flow rates) are collected from the coupled wellbore-reservoir simulations.

The mesh is regenerated for various values of D and f_{thf} by “sculpting” a pre-generated, octree-refined, cube-shaped mesh. The pre-generated mesh has a depth of 3200 m. Top layers of the mesh are removed to the thickness D and caprock between the reservoir and thief zone and thief zone and aquifer are removed using f_{thf} to determine the vertical center of the thief zone. As the pre-generated, cube-shaped mesh is discretized vertically at 3.2 m, the total depth D is sampled discretely at this interval. The depth of the thief

Table 1: Parameters sampled by LHS

Input	Lower Bound	Upper Bound	Units	Description
$\dot{m}_{inj}$	0.2	2.0	kg/s	CO ₂ injection rate at each injection node
D	1000	4000	m	Total depth
f_{thf}	0.3	0.7	–	Center of thief zone as fraction of total depth
k_w	-14 (10)	-10 (10000)	log m ² mD	Wellbore permeability
k_{thf}	-14 (10)	-12 (1000)	log m ² mD	Thief zone permeability

zone must also adhere to this discretization, and is calculated by rounding the value of $D * f_{thf}$ to the nearest multiple of 3.2 m.

The FEHM model simulations prescribed by the LHS samples were executed concurrently on Los Alamos National Laboratory high performance computational resources (clusters). Generation of LHS samples and subsequent concurrent execution of FEHM model simulations were managed by the Model Analysis ToolKit (MATK; <http://matk.lanl.gov>).

2.3. Evaluating Predictive Capacity of ROMs by Cross-Validation

Cross-validation (CV) is the process of repetitively partitioning a dataset into build and test sets. For each repetition, the model is generated using the build set and then tested using the test set. A goodness-of-fit metric is calculated on the test data to evaluate the model in each CV iteration. In the present case, this metric is the coefficient of determination (RSq; equation A.2). The collection of RSq values provides an indication of the

246 predictive capability of the model. A collection of large RSq values with
 247 small variance indicates high predictive capacity. The predictive capacity of
 248 the full model, generated from the entire dataset, is expected to be at least
 249 as good as that indicated by the CV results.

250 In the present case, the dataset of input/output samples is split into
 251 10 subsets. Ten CV repetitions are performed where one of the subsets is
 252 used as the test set and the remainder as the build set. In this way, all
 253 the data is used in testing the predictive capacity. In each repetition, 10%
 254 of the data is used to test a model built with 90% of the data. CV RSq
 255 values (i.e. RSq of the CV test data) are collected during the MARS forward
 256 pass to be compared with the full model RSq (RSq of the build data, the
 257 full model has no test data) and the generalized coefficient of determination
 258 (GRSq; equation A.5). GRSq is analogous to RSq, but includes a penalty
 259 term that increases as terms are added to the ROM. This promotes a more
 260 parsimonious and generalized ROM that avoids over fitting. Therefore, we
 261 expect the GRSq to increase as additional terms are added to the ROM to a
 262 point where the penalty due to additional terms will cause it to decline. The
 263 ROM is selected at the maximum value of the GRSq.

264 **3. Results and Analysis**

265 *3.1. Statistical Evaluation of ROM Inputs and Outputs*

266 Figure 6 presents histograms of the MARS inputs. Although D , f_{thf} , k_w ,
 267 and k_{thf} are sampled by LHS, it is apparent that the presented samples are

268 biased to lower values of k_w . This is the result of incomplete simulations
 269 due to numerical instabilities that are more prevalent at higher values of k_w .
 270 Although this does introduce bias in the sampling, for the purposes of ROM
 271 development, the histograms of the LHS sampled inputs do appear to cover
 272 the input ranges with ample number of examples across the k_w range.

273 Larger values of k_w may lead to non-Darcian flow, and may explain the
 274 numerical difficulties of some of the simulations. Based on densities and vis-
 275 cosities encountered in the ensemble, the approximate maximum CO₂ and
 276 brine flow rate encountered (8×10^{-3} kg/s; refer to Figure 8), and a Reynolds
 277 number ($Re = \rho v d_{30} / \mu$) limit of 10 for Darcian groundwater flow, the repre-
 278 sentative grain size diameter (d_{30} ; 30% passing size from a grain size analysis)
 279 would have to be smaller than approximately 1.3 mm for supercritical CO₂,
 280 0.13 mm for water, and 35 mm for CO₂ gas to remain in a Darcy flow regime.
 281 The value of d_{30} for water would increase slightly for brine based on salinity.
 282 The limiting value for ensuring Darcian groundwater flow is $d_{30} = 0.13$ mm
 283 for water, which corresponds to a fine sand. Since the simulations assume a
 284 degraded, cemented wellbore, the assumption of Darcian groundwater flow
 285 for the ensemble is reasonable.

286 ΔP_{res} and S_{res} are inputs that emerge from the FEHM simulations and
 287 are therefore not expected to have unbiased histograms. It is also important
 288 to note that there are a collection of values of ΔP_{res} and S_{res} for each LHS
 289 sample. LHS sampled inputs prescribe the simulation characteristics, while
 290 ΔP_{res} and S_{res} are transient values that emerge from reservoir runs.

Figure 6: Histograms of MARS inputs.

Figure 7: Time series of median (blue lines) and 95% confidence band (shaded region) of (a) pressures (P_{res}) and (b) CO₂ saturations (S_{res}) at top of the reservoir at the location of the wellbore collected from the reservoir simulations

Figure 7 presents the pressures (P_{res}) and CO₂ saturations (S_{res}) at the top of the reservoir at the location of the wellbore collected from the reservoir simulations. ΔP_{res} is calculated from P_{res} by subtracting the initial (hydrostatic) value of P_{res} from the transient values. There are a great number of reservoir simulations with S_{res} values that either reach full saturation of 1 at early times (i.e., $t < 10$ y) or remain at relatively low values for the entire simulation. While S_{res} values in between these two end members are less common, there are examples that span the range.

Figure 8 presents time-series of flow rates of CO₂ and brine into the thief zone, aquifer, and atmosphere. It is apparent that there is a decline in CO₂ flow rates between the thief zone and aquifer, followed by a larger and more significant decline between the aquifer and atmosphere. Water leakage into the thief zone is similar in scale to CO₂, but declines to relatively

Figure 8: Time series of CO₂ flow into the (a) thief zone, (b) aquifer, (c) atmosphere and water flow into the (d) thief zone and (e) aquifer.

insignificant values for flow into the aquifer. Although there are significant differences between the FEHM model developed in Jordan et al. (2015) and the one developed here (i.e., 2D-radial vs 3D, aquifer thickness, steady-state vs transient, thief zone, etc.), the flow rates in both studies are of the same order of magnitude (less than 1×10^{-2} kg/s). This indicates that although additional complexities were considered here (e.g., 3D transient leakage vs 2D radial steady state leakage), the simulations are consistent.

Figure 9 presents correlation matrices for (a) Pearson (Pearson, 1895) and (b) Spearman (Spearman, 1904) correlation coefficients. Pearson correlation is a measure of linear dependence while Spearman correlation is a measure of monotonic dependence based on ranked variables. Spearman correlation is able to identify monotonic nonlinear dependence. Increases in correlation coefficients between Pearson and Spearman indicate the existence of mono-

317 tonic nonlinear dependence. For example, it is apparent that MARS outputs
 318 have monotonic nonlinear dependence on S_{res} in particular. However, it is
 319 also possible to decrease correlations from Pearson to Spearman, for example
 320 for correlations between k_w and $\dot{m}_{h2o}^{aqu}$.

321 The outputs have the strongest linear (Pearson) correlation with the well-
 322 bore permeability (k_w), with a maximum coefficient of 0.58 for $\dot{m}_{h2o}^{aqu}$. For non-
 323 linear monotonic (Spearman) correlations, larger values are also observed for
 324 k_w . Larger magnitude values are also observed for S_{res} in this case, with
 325 positive correlations for CO₂ flow rates and negative correlations for brine
 326 flow. The largest Spearman correlation is 0.67 for $\dot{m}_{co2}^{thf}$. The lack of large
 327 correlations (> 0.7 for instance) indicates that the outputs are not strongly
 328 linearly or monotonically dependent on any single parameter and are likely
 329 dependent on interactions among parameters.

330 To further investigate the dependence of leakage on LHS sampled param-
 331 eters, Figure 10 presents the paired plots between LHS sampled parameters
 332 and net mass of CO₂ and brine leaked into the thief zone, aquifer, and at-
 333 mosphere over the 50 year simulations. Leakage of brine to the atmosphere
 334 is not included as an output. Significant correlation is apparent between
 335 $\dot{m}_{inj}$ and net CO₂ leakage to the thief zone ($\dot{m}_{co2}^{thf}$) and aquifer ($\dot{m}_{co2}^{aqu}$). An
 336 initial strong upward trend followed by an asymptotic trend is apparent in
 337 the loess smooths (red lines in lower triangle of matrix plots). This implies
 338 that a maximum net leakage is attained at around $\dot{m}_{inj} = 17$ kg/s. The
 339 strong positive linear correlation is also apparent in the Pearson correlation

(a) Pearson correlation coefficient matrix

(b) Spearman correlation coefficient matrix

Figure 9: Correlation of input values to MARS with output results in terms of Pearson and Spearman coefficients.

coefficients and the scatterplots between k_w and net brine leakage to the thief zone (m_{h2o}^{thf}) and aquifer (m_{h2o}^{aqu}). Most other correlations are small and trends in the scatterplots difficult to discern.

Significant correlations exist between net CO₂ leakages to the thief zone and aquifer and the aquifer and atmosphere but not between the thief zone and atmosphere. Net brine leakage to the thief zone and aquifer are highly correlated. The correlation analysis suggests that MARS input/output dependence is complex and likely involves significant nonlinearities and input interaction. These complex dependencies emerge from the 3D multi-phase numerical simulations and would likely be difficult to simulate within an analytical solution. Some of the complexities may be explained by the fact that we are correlating inputs and outputs in a decoupled fashion, using pressures and saturations without leakage effects as ROM inputs and flow rates that include leakage effects as ROM outputs. These complexities are required if we intend to capture coupled effects of leakage in a system model with decoupled ROMs.

Figure 11 presents a histogram of the total injected CO₂ for the sampled simulations, with values ranging from a little over 1 Mt to around 21.3 Mt. Figure 12 presents the fraction of the total injected CO₂ in the thief zone, aquifer, and leaked to the atmosphere at 50 years. The extremely small fraction of the total injected CO₂ that has leaked up the wellbore is apparent with values between 1×10^{-20} to 1×10^{-3} .

Figure 10: Paired plots between LHS sampled parameters and net leakage of CO₂ and brine to the thief zone, aquifer and atmosphere. Diagonal plots are histograms with density estimate curves. Upper triangular plots contain Pearson correlation coefficients between row and column parameters. Lower triangular plots are histograms between row and column parameters with loess smooths (red lines).

Figure 11: Histogram of total injected CO₂ from reservoir simulations calculated using FEHM.

Figure 12: Histograms of fraction of total injected CO₂ in the (a) thief zone, (b) aquifer, and (c) atmosphere at 50 years from reservoir simulations calculated with FEHM.

362 3.2. ROM diagnostics

363 Figure 13 presents diagnostic plots for the ROM builds for CO₂ leakage
364 to the thief zone, aquifer, and atmosphere and for brine leakage to the thief
365 zone and aquifer. The subplots present the GRSq (Equation A.5) of the full
366 model, the RSq (equation A.2) for the full model and for the test (out-of-fold)
367 data for each CV repetition, and the mean RSq of the CV repetitions as a
368 function of the number of model terms. These values, along with Generalized
369 Cross Validation (GCV; equation A.6) and Sum-of-Squared Residual (SSR;
370 equation A.3) values, for the selected models are shown in tabular form in
371 Table 2. GCV approximates the predictive error of the ROM that would be
372 obtained using leave-one-out CV, while SSR is the cumulative measure of
373 mismatch between build data outputs and ROM simulated values.

374 The model is selected at the point where the GRSq is maximized, indicat-
375 ing that the penalty for additional terms outweighs the improved fit (refer
376 to equations A.5 to A.7). In all cases, the model is selected close to the
377 maximum number of terms considered, except for CO₂ flow to the thief zone
378 (Figure 13a). This is indicated in the plots by a black vertical dotted line.
379 The thin, light red lines are the RSq values of the test data in each of the
380 10 CV repetitions. The tight clustering of these curves and their similarity
381 to the GRSq/RSq of the full model indicates that the full model will have
382 good predictive capability (the RSq of the full model is obscured behind the
383 GRSq in all cases). This is also apparent in the similarity in the GRSq, RSq,
384 and CV RSq values in Table 2.

Table 2: ROM build statistics

ROM	GCV	SSR	GRSq	RSq	CV RSq
$\dot{m}_{co_2}^{thf}$	1.60E-08	4.44E-03	0.95	0.95	0.95
$\dot{m}_{co_2}^{aqu}$	5.63E-09	1.56E-03	0.97	0.97	0.97
$\dot{m}_{co_2}^{atm}$	8.96E-10	2.48E-04	0.98	0.98	0.97
$\dot{m}_{h_2o}^{thf}$	1.26E-09	3.49E-04	0.99	0.99	0.99
$\dot{m}_{h_2o}^{aqu}$	1.58E-12	4.38E-07	0.97	0.97	0.97

385 While residual analysis of the ROM fitting (not presented here) provided
386 indications of heteroscedasity and non-normality in the residuals, these issues
387 are not as severe a problem for MARS ROMs as for linear models (Milbor-
388 row, 2013). Based on the repeatably large CV RSq values, we have decided
389 that the level of heteroscedasity and non-normality are acceptable for our
390 purposes.

391 To further illustrate ROM fitting by MARS, Figure 14 presents scatter-
392 plots of ROM fitted outputs versus build data outputs (FEHM generated).
393 Figure 15 presents histograms comparing the FEHM and MARS ROM out-
394 puts (flow rates) for the ROM generated with all the data. This provides
395 an indication of the ability of the ROMs to reproduce the ensemble behav-
396 ior of the full FEHM simulations. While for many individual leakage sce-
397 narios, ROM predictions provide good approximations of FEHM, there are
398 cases with significantly differences. Capturing uncertainties in wellbore leak-
399 age simulations in an ensemble sense is the intended purpose of the ROMs.

Figure 13: Analysis of MARS ROM building for CO₂ flow to the (a) thief zone ($\dot{m}_{co2}^{thf}$), (b) aquifer ($\dot{m}_{co2}^{aqu}$), and (c) atmosphere ($\dot{m}_{co2}^{atm}$) and water flow to the (d) thief zone ($\dot{m}_{h2o}^{thf}$) and (e) aquifer ($\dot{m}_{h2o}^{aqu}$). The diagnostic metrics GRSq and RSq are plotted as functions of the number of terms added to the ROMs. ROMs are selected at the maximum value of the GRSq (solid black line) as indicated by the vertical dashed black line. The RSq value is obscured in all cases as it is behind the GRSq. The light red lines are RSq values for test (out-of-fold) data for individual CV runs. The thick red line is the average of the CV RSq values, where the vertical dashed red line indicates the location of the maximum value.

Figure 14: Scatterplots of full model build data flow rates (FEHM) versus fitted ROM flow rates for CO₂ flow to the (a) thief zone ($\dot{m}_{co2}^{thf}$), (b) aquifer ($\dot{m}_{co2}^{aqu}$), and (c) atmosphere ($\dot{m}_{co2}^{atm}$) and water flow to the (d) thief zone ($\dot{m}_{h2o}^{thf}$) and (e) aquifer ($\dot{m}_{h2o}^{aqu}$). A 1:1 line is plotted for reference.

400 Therefore, these results indicate that the ROMs have the potential to achieve
401 this goal.

402 3.3. Input importance

403 MARS' rankings of input importance are presented in Figure 16. Missing
404 inputs for ROMs indicate that MARS deemed the input unimportant and
405 has not included it in the ROM. Comparing Figure 16 with Figure 9 (b),
406 it is apparent that there are some similarities between MARS' importance
407 ranking and Spearman correlation coefficient magnitude. For example, k_w is
408 ranked the most important and has high Spearman correlation in all cases.
409 However, there are many differences as well. For example, f_{thf} has low
410 correlation in all cases, however, it is ranked third in importance by MARS
411 for $\dot{m}_{h2o}^{aqu}$. The differences in these figures may be partly explained by the
412 fact that MARS is able to identify some non-monotonic dependencies, while
413 Spearman correlation cannot.

414 4. Conclusions

415 We present an approach to build CO₂ and brine wellbore leakage ROMs
416 from numerical simulations that include transient leakage during and after
417 injection, thief zone, and leakage effects on reservoir pressures and saturations
418 in decoupled ROMs (i.e., where reservoir pressures and saturations provided
419 to the wellbore leakage ROMs do not include the effects of leakage).

420 The following conclusions have been drawn from this research:

Figure 15: Comparison of histograms of FEHM (blue) and MARS RS (red) flow rates to the (a) thief zone ($\dot{m}_{co2}^{thf}$), (b) aquifer ($\dot{m}_{co2}^{aqu}$), and (c) atmosphere ($\dot{m}_{co2}^{atm}$) and water flow to the (d) thief zone ($\dot{m}_{h2o}^{thf}$) and (e) aquifer ($\dot{m}_{h2o}^{aqu}$).

Figure 16: MARS importance ranking of inputs. Missing inputs indicate that MARS deemed the inputs insensitive and did not use them in the ROM.

1. Computational requirements can be significantly reduced through the use of mesh sculpting
2. Octree refinement considerably reduces mesh effects on CO_2 saturation plumes
3. Many input/output correlations from the decoupled models exhibit nonlinear dependence, non-monotonic dependence, and input interactions that would be difficult to capture in analytical solutions
4. It is possible to capture the ensemble characteristics of transient coupled effects in decoupled wellbore leakage ROMs
5. Of the sampled parameters, flow rates are most dependent on wellbore permeability

432 6. Significant nonlinear, monotonic dependence of flow rates on CO₂ sat-
433 urations is also identified, positive for CO₂ and negative for brine

434 Acknowledgments

435 This work was funded by the US DOE’s Fossil Energy Office through the
436 National Risk Assessment Partnership (NRAP) managed by the National
437 Energy Technology Laboratory (NETL). Numerical simulations were per-
438 formed on Los Alamos National Laboratory clusters supported by the High
439 Performance Computing Division.

440 Appendix A. MARS Algorithm

441 ROMs are produced using a MARS algorithm implementation in R’s
442 Earth package (Milborrow, 2013). A brief review of MARS necessary for
443 interpreting results will be provided here. For additional details, refer to
444 Friedman (1991). MARS is a non-parametric regression technique intended
445 to handle non-linear dependence and input parameter interactions. MARS
446 generates a ROM ($\hat{f}$) as an expansion of basis functions in the form

$$\hat{f}(\mathbf{x}) = \sum_{m=1}^M a_m B_m(\mathbf{x}), \quad (\text{A.1})$$

447 where $\mathbf{x}$ is a vector of input parameters, B_m are the basis functions, M is the
448 number of basis functions determined by the MARS regression, and a_m is the
449 m th constant coefficient. A single basis function takes a constant value of 1 in

order to create the model intercept. Other basis functions are either a single hinge function or the product of 2 or more hinge functions. The number of combinations of hinge functions allowed within a basis function is specified by the user. Hinge functions take the form $\max(0, x - c)$ or $\max(0, c - x)$ where c is a constant referred to as the knot. The use of hinge functions allows the MARS algorithm to partition the response surface into regions; however, the surface remains continuous with continuous derivatives.

Starting from a single basis function of 1, MARS performs a forward pass by adding mirrored pairs of new hinge functions to existing basis functions that result in the greatest increase in the Coefficient of Determination (RSq) defined as

$$RSq = 1 - \frac{SSR}{TSS} \quad (A.2)$$

where SSR is the sum-of-squared-residuals (SSR) defined as

$$SSR = \sum_i^N (y_i - \hat{f}(\mathbf{x}_i))^2 \quad (A.3)$$

and TSS is the total sum-of-squares defined as

$$TSS = \sum_i^N (y_i - \bar{y})^2 \quad (A.4)$$

where N is the number of input/output samples, y_i is the i th output, and $\bar{y}$ is the mean of y . The forward pass is terminated either when (1) an increase in RSq is less than a specified value (e.g., 0.001), (2) the RSq is greater than

466 or equal to a specified value (e.g., 0.999), (3) the maximum number of terms
 467 is reached, or (4) the Generalized RSq (GRSq) < -10 , where the GRSq is
 468 defined as

$$\text{GRSq} = 1 - \frac{\text{GCV}}{\text{GCV}_0} \quad (\text{A.5})$$

469 The GCV in Equation A.5 is the Generalized Cross Validation defined as

$$\text{GCV} = \frac{\text{SSR}}{N * (1 - N_e/N)^2}, \quad (\text{A.6})$$

470 where N is the number of input/output samples in the dataset and N_e is the
 471 effective number of terms defined as

$$N_e = N_m + \frac{N_m - 1}{2}p, \quad (\text{A.7})$$

472 where N_m is the number of MARS terms and p is the penalty for addi-
 473 tional terms. The GCV provides an approximation of the error that leave-
 474 one-out CV would produce (i.e., building the response surface leaving one
 475 input/output combination out and calculating the error in predicting that
 476 output). The analogy between RSq and GRSq is apparent by comparing
 477 Equation A.2 and A.5. GCV_0 is the GCV of the intercept only model.

478 In order to generalize the ROM and avoid overfitting, MARS performs a
 479 backward pass to prune terms from the model. As opposed to the forward
 480 pass, the backward pass can consider all terms and is able to remove one
 481 term in a paired basis function. The determination of which term to remove

482 is based on minimizing the GCV. Decreases in the GCV are possible as terms
483 are dropped during the backward pass, whereas the SSR will only increase
484 as terms are dropped. The use of the GCV during the backward pass allows
485 MARS to generate a model that compromises between goodness-of-fit and
486 model parsimony criteria.

487 After the backward pass, the coefficients $\mathbf{a} = [a_1, a_2, \dots, a_M]$ are deter-
488 mined by regressing the output samples $\mathbf{y} = [y_1, y_2, \dots, y_N]$ on the basis
489 matrix $\mathbf{B} = [B_1, B_2, \dots, B_M]$ using ordinary least squares.

490 Bai, M., Sun, J., Song, K., Li, L., Qiao, Z., 2015. Well completion and
491 integrity evaluation for CO₂ injection wells. Renewable and Sustainable
492 Energy Reviews 45, 556 – 564.

493 Carey, J. W., Wigand, M., Chipera, S. J., WoldeGabriel, G., Pawar, R.,
494 Lichtner, P. C., Wehner, S. C., Raines, M. A., Guthrie, G. D., 2007. Anal-
495 ysis and performance of oil well cement with 30 years of co 2 exposure
496 from the sacroc unit, west texas, usa. International Journal of Greenhouse
497 Gas Control 1 (1), 75–85.

498 Crow, W., Carey, J. W., Gasda, S., Williams, D. B., Celia, M., 2010. Well-
499 bore integrity analysis of a natural co 2 producer. International Journal of
500 Greenhouse Gas Control 4 (2), 186–197.

501 Environmental Protection Agency, May 2013. Underground Injection Control
502 (UIC) Program Class VI Well Area of Review Evaluation and Corrective

503 Action Guidance for Owners and Operators. Tech. Rep. EPA 816-R-13-
504 005, EPA, Office of Water (4606M).

505 Friedman, J. H., 1991. Multivariate adaptive regression splines. The annals
506 of statistics, 1–67.

507 Gasda, S. E., Nordbotten, J. M., Celia, M. A., 2009. Vertical equilibrium
508 with sub-scale analytical methods for geological co2 sequestration. Com-
509 putational Geosciences 13 (4), 469–481.

510 George, D., Kuprat, A., Carlson, N., Gable, C., 1999. LaGriT-Los Alamos
511 Grid Toolbox. <http://lagrit.lanl.gov/>.

512 Gleason, R. A., Thamke, J. N., Smith, B. D., Tangen, B. A., Chesley-Preston,
513 T., Preston, T. M., 2011. Examination of brine contamination risk to
514 aquatic resources from petroleum development in the williston basin. US
515 Geological Survey Fact Sheet 3047 (2011), 4.

516 Global CCS Institute, 2014. The global status of CCS: 2014.
517 [http://www.globalccsinstitute.com/publications/global-status-ccs-](http://www.globalccsinstitute.com/publications/global-status-ccs-february-2014)
518 [february-2014](http://www.globalccsinstitute.com/publications/global-status-ccs-february-2014).

519 Guen, Y. L., Huot, M., Loizzo, M., Poupard, O., 2011. Well integrity risk
520 assessment of Ketzin injection well (Ktzi-201) over a prolonged sequestra-
521 tion period. In: 10th International Conference on Greenhouse Gas Control
522 Technologies. Vol. 4. pp. 4076–4083.

- 523 Jordan, A. B., Stauffer, P. H., Harp, D. R., Carey, W. J., Pawar, R. J., 2015.
524 A response surface model to predict CO_2 and brine leakage along cemented
525 wellbores. *International Journal of Greenhouse Gas Control* 33, 27–39.
- 526 Keating, E. H., Fessenden, J., Kanjorski, N., Koning, D. J., Pawar, R., 2010.
527 The impact of CO_2 on shallow groundwater chemistry: observations at a
528 natural analog site and implications for carbon sequestration. *Environmental*
529 *Earth Sciences* 60 (3), 521–536.
- 530 Metz, B., 2005. Carbon dioxide capture and storage: Special report of the
531 intergovernmental panel on climate change. Cambridge University Press.
- 532 Metz, B., Davidson, O., De Coninck, H., Loos, M., Meyer, L., 2005. IPCC,
533 2005: IPCC special report on carbon dioxide capture and storage. prepared
534 by working group iii of the intergovernmental panel on climate change.
535 Cambridge, United Kingdom and New York, NY, USA, 442 pp.
- 536 Milborrow, S., 2013. earth: Multivariate Adaptive Regression Spline Models.
537 Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan
538 Miller’s Fortran utilities with Thomas Lumley’s leaps wrapper. R package
539 version 3.2-6.
540 URL <http://CRAN.R-project.org/package=earth>
- 541 Nicot, J.-P., Oldenburg, C. M., Houseworth, J. E., Choi, J.-W., 2013. Analy-
542 sis of potential leakage pathways at the Cranfield, MS, U.S.A., CO_2 seques-
543 tration site. *International Journal of Greenhouse Gas Control* 18, 388–400.

544 Nordbotten, J. M., Celia, M. A., Bachu, S., Dahle, H. K., 2005. Semiana-
545 lytical solution for co2 leakage through an abandoned well. *Environmental*
546 *science & technology* 39 (2), 602–611.

547 Nordbotten, J. M., Kavetski, D., Celia, M. A., Bachu, S., 2008. Model for
548 co2 leakage including multiple geological layers and multiple leaky wells.
549 *Environmental science & technology* 43 (3), 743–749.

550 Oldenburg, C. M., Bryant, S. L., Nicot, J.-P., 2009. Certification framework
551 based on effective trapping for geologic carbon sequestration. *International*
552 *Journal of Greenhouse Gas Control* 3, 444–457.

553 Pan, L., Oldenburg, C. M., Pruess, K., Wu, Y.-S., 2011. Transient co2 leakage
554 and injection in wellbore-reservoir systems for geologic carbon sequestra-
555 tion. *Greenhouse Gases: Science and Technology* 1 (4), 335–350.

556 Pawar, R., Bromhal, G., Dilmore, R., Foxall, B., Jones, E., Oldenburg, C.,
557 Stauffer, P., Unwin, S., Guthrie, G., 2013. Quantification of risk profiles
558 and impacts of uncertainties as part of us doe’s national risk assessment
559 partnership (nrap). *Energy Procedia* 37, 4765–4773.

560 Pearson, K., 1895. Note on regression and inheritance in the case of two
561 parents. *Proceedings of the Royal Society of London* 58 (347-352), 240–
562 242.

563 Spearman, C., 1904. The proof and measurement of association between two

564 things. The American Journal of Psychology 15 (1), pp. 72–101.

565 URL <http://www.jstor.org/stable/1412159>

566 Stauffer, P. H., Viswanathan, H. S., Pawar, R. J., Guthrie, G. D., 2008. A

567 system model for geologic sequestration of carbon dioxide. Environmental

568 science & technology 43 (3), 565–570.

569 Tao, Q., Checkai, D., Huerta, N., Bryant, S. L., 2011. An improved model to

570 forecast co 2 leakage rates along a wellbore. Energy Procedia 4, 5385–5391.

571 Viswanathan, H. S., Pawar, R. J., Stauffer, P. H., Kaszuba, J. P., Carey,

572 J. W., Olsen, S. C., Keating, G. N., Kavetski, D., Guthrie, G. D., 2008.

573 Development of a hybrid process and system model for the assessment of

574 wellbore leakage at a geologic co2 sequestration site. Environmental science

575 & technology 42 (19), 7280–7286.

576 Wilkin, R. T., DiGiulio, D. C., 2010. Geochemical impacts to groundwater

577 from geologic carbon sequestration: controls on pH and inorganic carbon

578 concentrations from reaction path and kinetic modeling. Environmental

579 science & technology 44 (12), 4821–4827.

580 Zhou, W., Stenhouse, M., Arthur, R., 2005. Assessment of potential well

581 leakage in the Weyburn site using a stochastic approach. In: Fourth Annual

582 Conference on Carbon Capture and Sequestration, Alexandra, Virginia.

583 Zyvoloski, G., 2007. FEHM: A control volume finite element code for simulat-

584 ing subsurface multi-phase multi-fluid heat and mass transfer. Los Alamos
585 Unclassified Report LA-UR-07-3359.