SAND2019-4151C

High Fidelity Surrogate Modeling of Fuel Dissolution for Probabilistic Assessment of Repository Performance

P.E. Mariner, D.T. Seidl, B.J. Debusschere, J. Vo, J.M. Frederick, L.P. Swiler
Sandia National Laboratories, USA
and J.L. Jerden
Argonne National Laboratory, USA

2019 IHLRWM Conference

Knoxville, TN
April 16, 2019

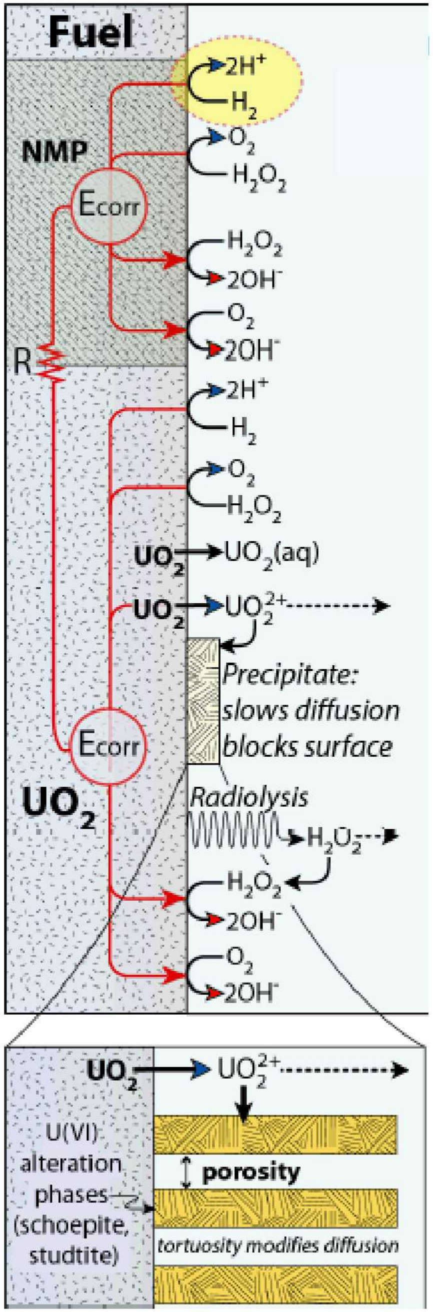
Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. This presentation is approved as SAND2019.

Is high-fidelity modeling at waste-package scale possible in a geologic repository performance assessment (PA) model?

https://www.deviantart.com/fractalmonster/art/For-Emz1-127400950
Labeled for noncommercial reuse

Fuel Dissolution Process Model

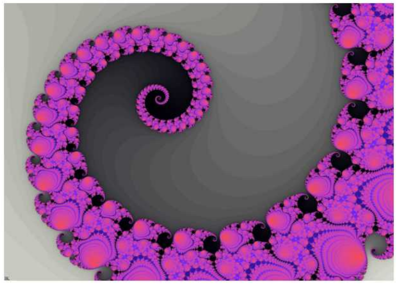
- Complex set of processes
  - Radiolysis
  - Oxidation of H_2 via noble metal particle (NMP) catalyst
  - 1-D reactive transport through alteration layer
  - Growth of the alteration layer
  - Diffusion of reactants and products through the alteration layer
- Expensive in a repository PA calculation
  - Slow, iterative solution is required for each call to the process model
  - ~1 billion calls per probabilistic PA simulation
    - (Thousands of waste packages) × (Thousands of time steps) × (Hundreds of realizations)
  - => Process model much too slow to be directly used in a repository PA calculation



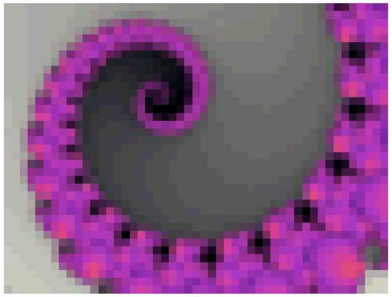
(Figure adapted from Jerden et al. 2017)

Surrogate Models

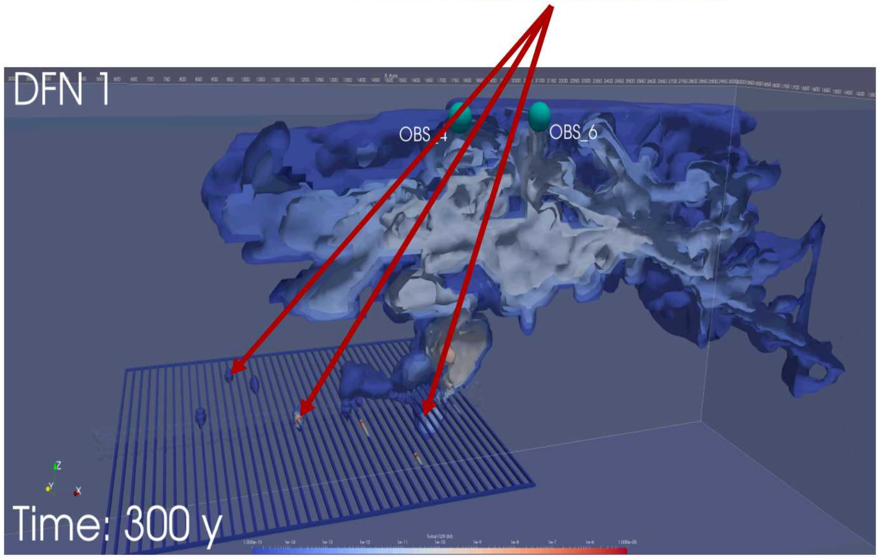
- Can capture the important effects of high-fidelity process models
- Can run orders of magnitude faster than process models
- Can be used to
  - Identify important parameters in the process model
  - Track uncertainty introduced to the PA model by the surrogate model



Process model



Surrogate model

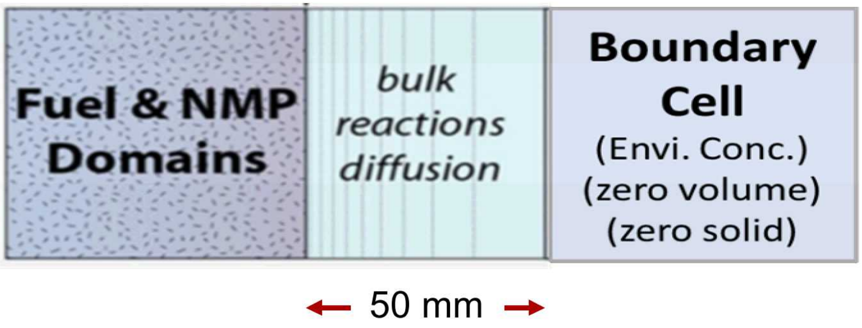


PA model

Objective of Study

- Develop two surrogate models of the Fuel Matrix Degradation (FMD) process model for use by PFLOTRAN in *GDSA Framework*
  - One continuous function surrogate model
    - Parametric surrogate model: polynomial linear regression
  - One lookup table surrogate model
    - Non-parametric surrogate model: k-Nearest Neighbors regression (kNNR)
- Assess error and simulation run time of these models relative to the coupled FMD process model

Fuel Matrix Degradation (FMD) Model

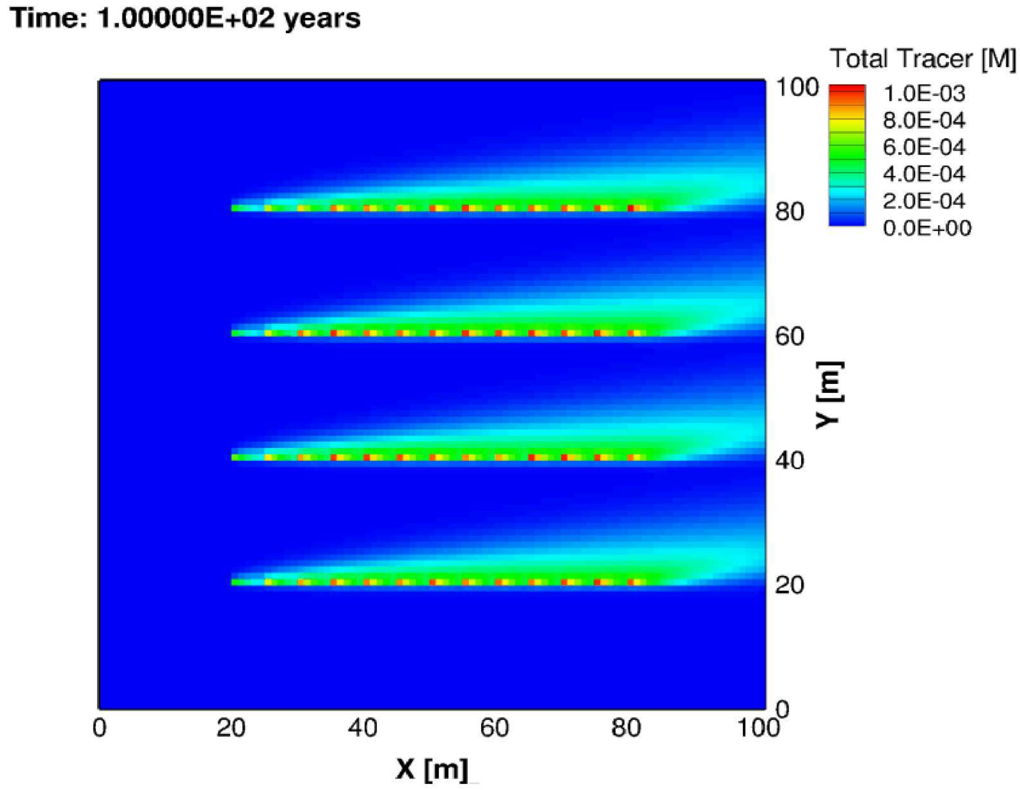


- Domain
  - 1D, fuel surface to bulk water
- Processes
  - Radiolysis, alteration layer growth, diffusion of reactants through the alteration layer, temperature, and interfacial corrosion potential
- FMD process model coded in Matlab
- Inputs/outputs each time step

Inputs	Outputs
• Initial concentration profiles across 1D corrosion/water layer ($UO_2(s)$, $UO_3(s)$, $UO_4(s)$, H_2O_2, UO_2^{2+}, UCO_3^{2-}, UO_2, CO_3^{2-}, O_2, Fe^{2+}, and H_2) • Initial corrosion layer thickness • Dose rate at fuel surface (= f (time, burnup)) • Temperature • Time, time step length • Environmental concentrations (CO_3^{2-}, O_2, Fe^{2+}, and H_2)	• Final concentration profiles across 1D corrosion/water layer • Final corrosion layer thickness • Fuel dissolution rate

Coupled FMD Model

- Coupled to PFLOTRAN in 2015
  - Recoded in Fortran
- Tested on a 2D layout
  - 52 breached spent fuel waste packages in a steady state flow field
  - 100 time steps
  - 45-minute simulation
  - 67% of computational time due to FMD process model
- Too expensive for PA



Polynomial Surrogate

- Two polynomial surrogates developed
  - Linear regression model for input parameters x_i
    - $\hat{f}(\boldsymbol{x}) \approx c_0 + \sum_{i=1}^{m} c_i\, x_i$
  - Second order polynomial regression (aka quadratic regression model)
    - $\hat{f}(\boldsymbol{x}) \approx c_0 + \sum_{i=1}^{m} c_i\, x_i + \sum_{i=1}^{m} \sum_{j \geq i}^{m} c_{ij}\, x_i x_j$
- Coefficients (c_0, c_i, c_{ij})
  - Determined by minimizing sum-of-squared error (SSE) between the surrogate model and the actual data y_i
  - $SSE = \sum_{i=1}^{n} \left(\hat{f}(\boldsymbol{x}_i) - y_i\right)^2$
  - Linear solve for linear regression model

Surrogate Training/Testing Data

- Training and testing data
  - 2,800 Matlab FMD model simulations
    - Each consisting of 101 points in time, logarithmically spaced from 0 to 10^5 yr
    - For polynomial surrogate, half used for training, the other half for testing
  - Inputs (not temperature) and outputs log-transformed prior to regressions
  - Latin hypercube sampling (LHS) of input parameters
  - Six-dimensional space

Parameter	Distribution	Min.	Max.
Init. Temp. (C)	Uniform	298	373
Burnup (Gwd/MTU)	Uniform	20	90
Env. CO_3^{2-} (mol/m^3)	Log-uniform	10^{-6}	10^{0}
Env. O2 (mol/m3)	Log-uniform	10^{-6}	10^{-1}
Env. Fe^{2+} (mol/m^3)	Log-uniform	10^{-6}	10^{-5}
Env. H_2 (mol/m^3)	Log-uniform	10^{-6}	10^{-1}

Two Polynomial Surrogate Models

- Input parameters for feature sets A and B

Feature Set	A	B
Initial (previous) concentrations of UO_2^{2+}, $UO_2(CO_3)_2^{2-}$, UO_2, and H_2O_2 at the bulk water boundary cell	X	
Initial (previous) corrosion layer thickness	X	
Dose rate at fuel surface	X	X
Temperature	X	X
Time	X	X
Environmental concentrations of CO_3^{2-}, O_2, Fe^{2+}, and H_2	X	X
Initial (previous) UO_2 surface flux (dissolution rate)		X

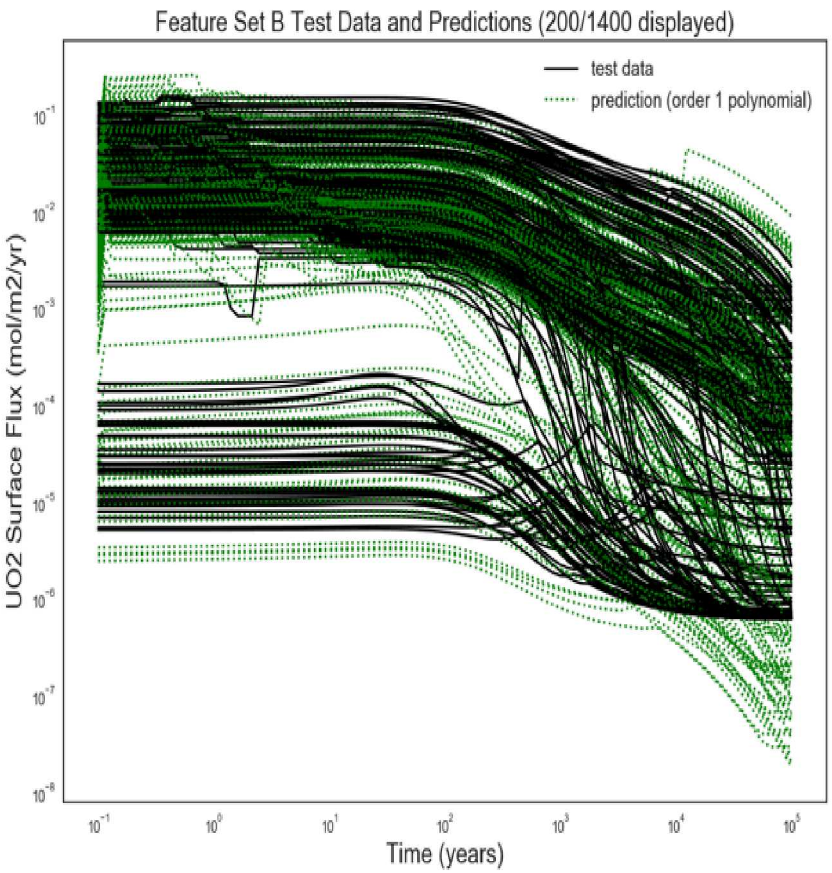
Error Analysis

- Relative pointwise absolute error (RPWAE)
  - $RPWAE = \frac{|y_{pred} - y_{true}|}{y_{true}} = |1 - \frac{y_{pred}}{y_{true}}|$ (at each data point)
- This error is averaged to obtain the *mean* RPWAE (M-RPWAE) metric for each test run

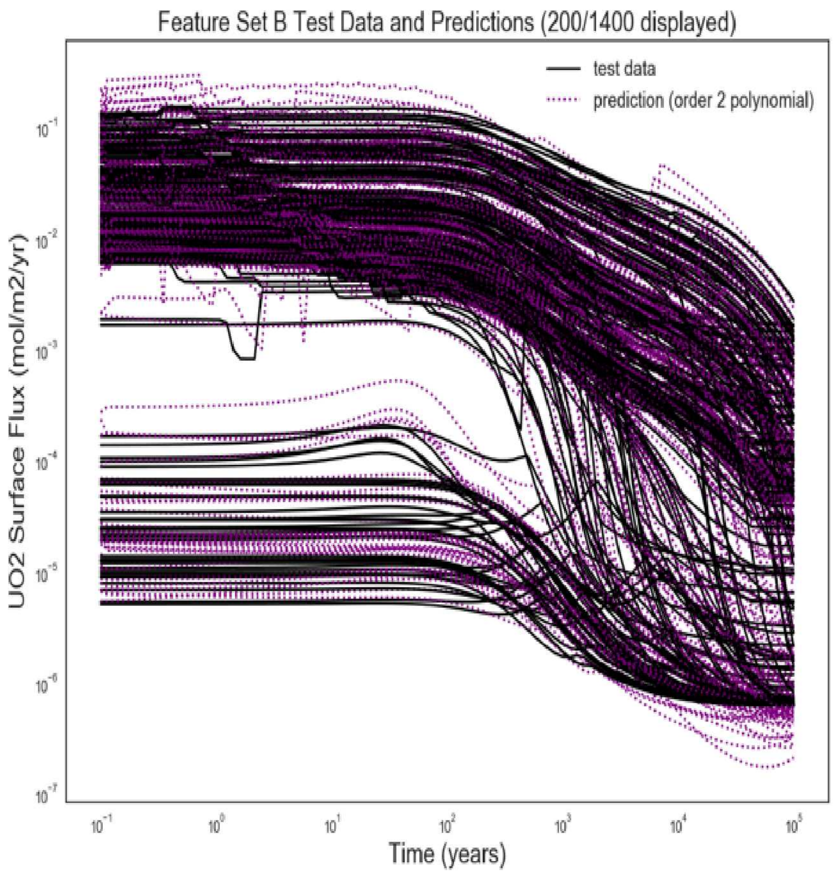
F-Set	P-Order	Terms	R^2	M-RPWAE
A	Linear	12	0.371	1.07
A	Quadratic	91	0.747	0.286
B	Linear	8	0.997	0.0515
B	Quadratic	45	0.997	0.0457

Feature Set A – Polynomial Surrogate

Linear

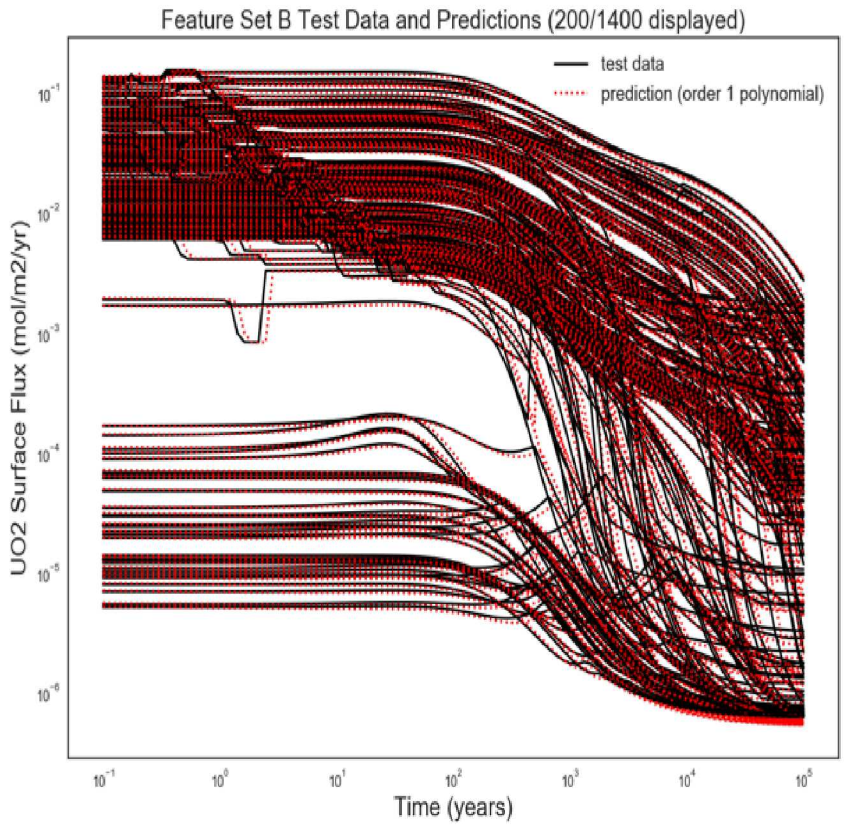


Quadratic

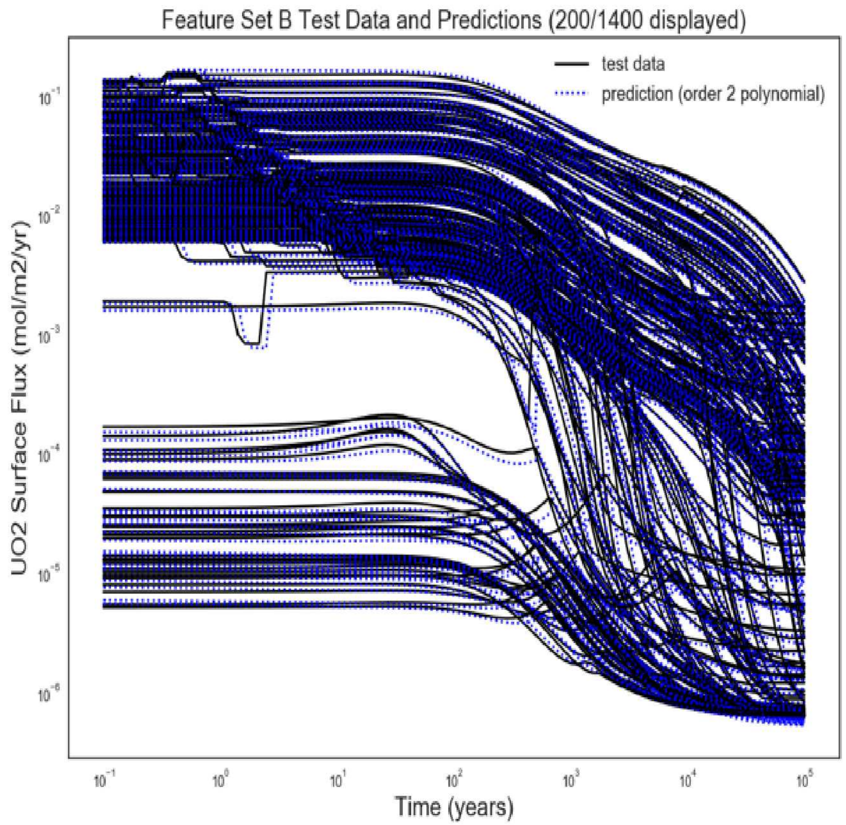


Feature Set B – Polynomial Surrogate

Linear

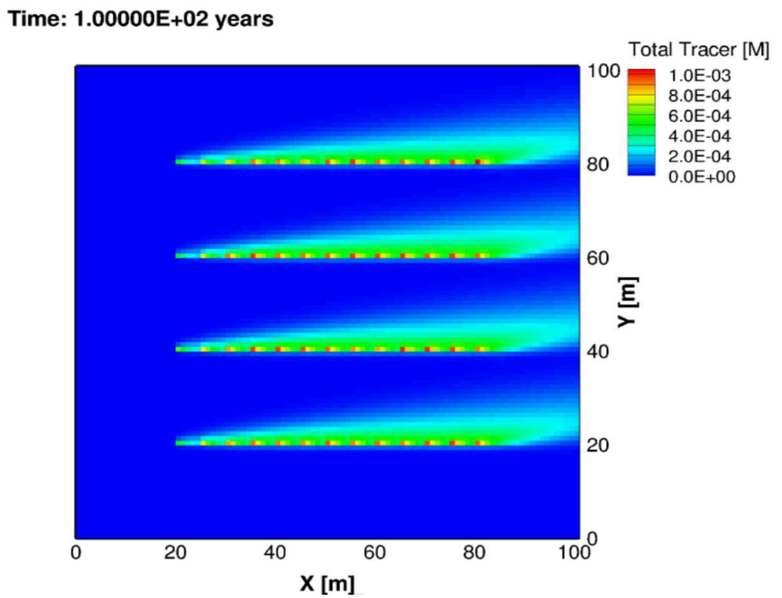


Quadratic



Polynomial Surrogate Assessment

- Polynomial surrogate model coupled to PFLOTRAN
- Tested on 2D example
  - Fast (see table)
  - Relative accuracy will be evaluated after coupled process model is updated to latest process model



Run Time (s)

Module	Coupled FMD Process Model	Coupled Polynomial Surrogate Model
Flow	168	194
Transport	244	278
Waste Form	1522	8

kNNR Lookup Surrogate

- k Nearest Neighbors regression (kNNR) surrogate model
  - Supervised, non-parametric, machine-learning method
  - Tabulates data points for making predictions on the fly
  - k is the number of nearest data points used in a prediction
  - Distance from the interrogation point depends on the metric, e.g.
    - Minkowski metric: $\left(\sum_{i=1}^{d}|x_i - y_i|^p\right)^{\frac{1}{p}}$, with $p \geq 1$
      - For the popular Euclidean metric, $p = 2$
  - An inverse of the distance to each neighbor may be used to determine how influential the neighbor is in calculating the weighted average
  - Tabulations may be of various forms
    - E.g., a table, K-D Tree, or Ball Tree
  - No need for global smoothness – kNNR acts locally
  - Requires sufficiently dense tabulation of data in sampled areas

kNNR Surrogate Setup

- Same 2,800 simulations used for training and testing
  - 10% used for testing
  - Remainder used for training in different training set sizes to examine the effects of training set size
- Manhattan distance metric
  - Same as Minkowski metric for $p = 1$
  - Better suited for higher-dimensional domain space
- Ball Tree tabulation (for same reason)
- Distance-weighted method used

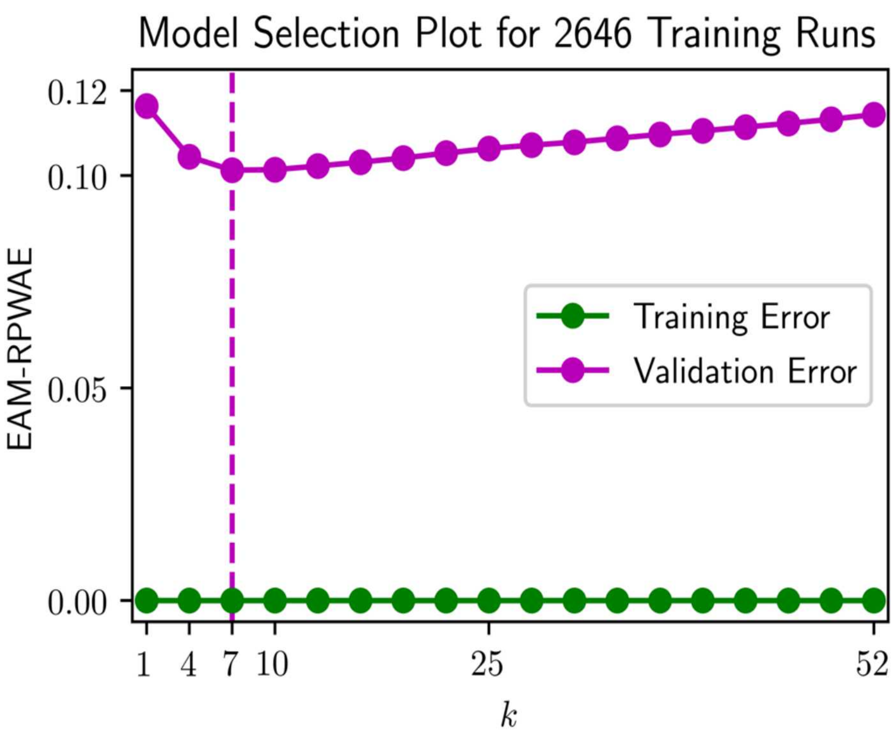
kNNR Surrogate Model

- Input parameters

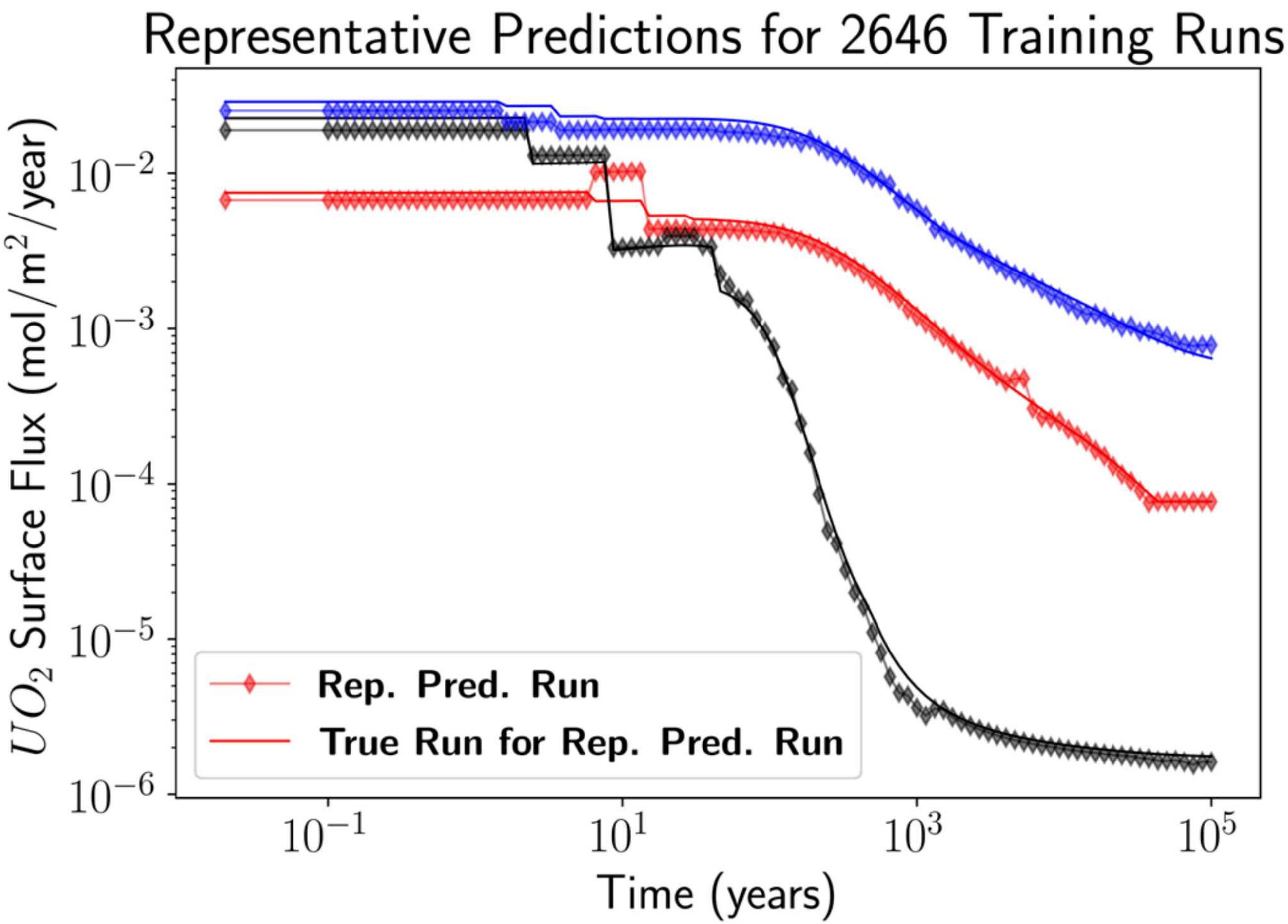
Feature Set
H_2 concentrations at the leftmost and rightmost endpoints of the spatial mesh inside the FMD model
H_2O_2 concentrations at the leftmost and rightmost endpoints of the spatial mesh
Dose rate at the leftmost endpoint of the spatial mesh

kNNR Surrogate Model

- 7 nearest neighbors optimal
- Decrease in error with increasing training set size

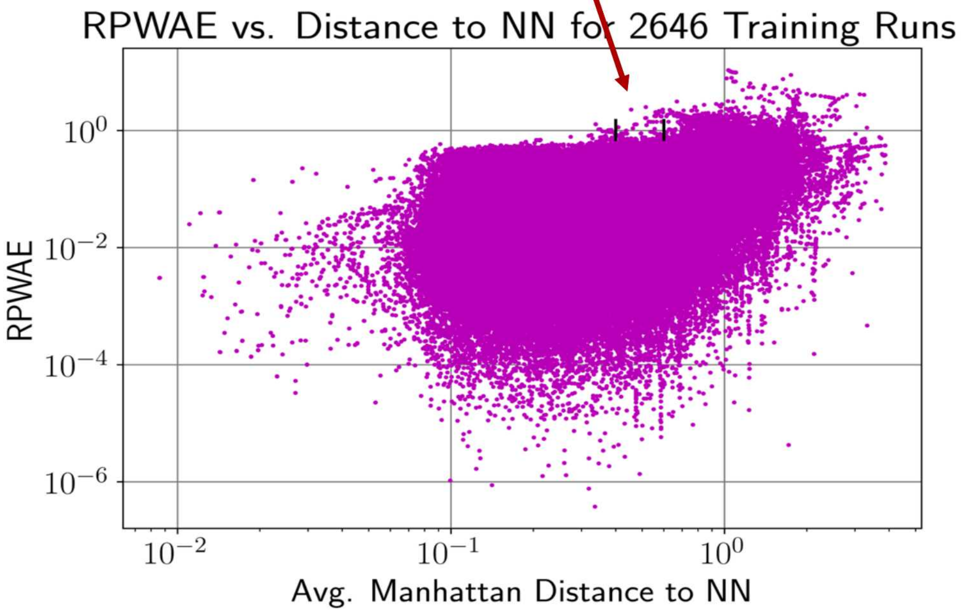


kNNR Surrogate Model



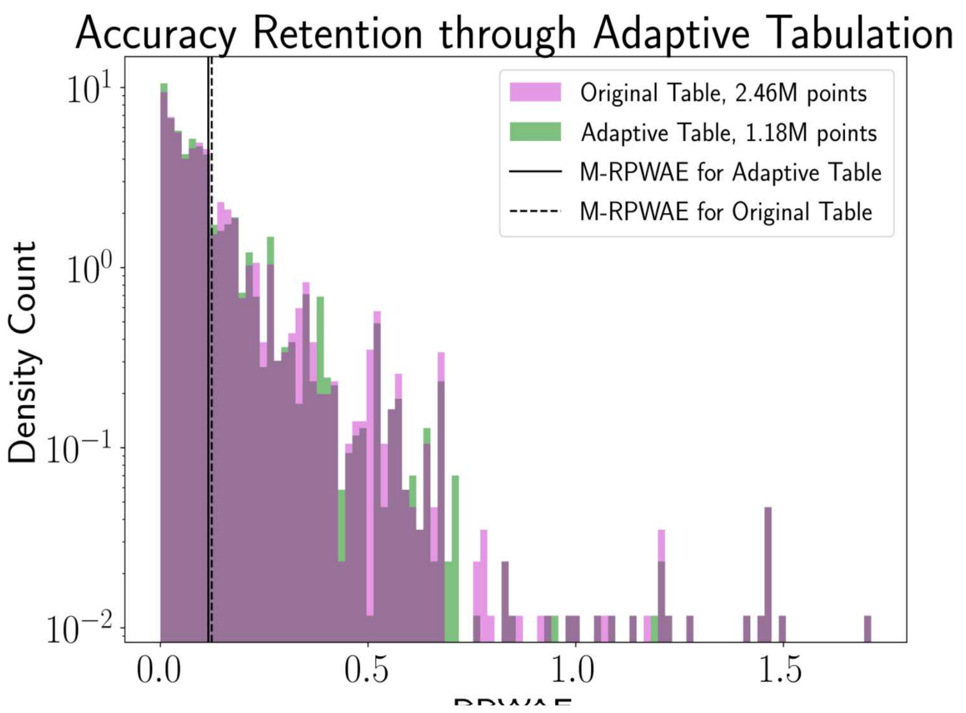
kNNR Surrogate Model

- Errors can increase above 100% when the average Manhattan distance exceeds 0.4 to 0.6 (denoted by black hashes)
- Results imply that
  - A higher density of training data is needed (limited effect here),
  - A distance cutoff is needed for nearest neighbors, and/or
  - Additional predictors may need to be added to the table (likely)



Run Time – kNNR Surrogate

- kNNR surrogate model not yet coupled to PFLOTRAN
- The standalone kNNR model appears to be faster than the coupled polynomial surrogate model
  - However, can't compare speeds very well until coupled



- Test
  - 5,000 lookups (done 30 times)
- Original table (2.46M points)
  - 4.14 seconds
- Adaptive table (1.18M points)
  - 1.79 seconds
  - Table thinned by 52% by requiring minimum distance of 0.05 (in the Manhattan norm)
  - Equivalent M-RPWAE

Conclusions

- Polynomial surrogate
  - Linear and quadratic fits produce similar accuracy
  - Coupled to PFLOTRAN
  - Very fast – increases speed of 2D example by a factor nearly 200
  - Work ongoing to reduce error for a feature input set that excludes fuel dissolution flux from previous time step
- kNNR surrogate
  - 7 nearest neighbors optimum
  - Fast and accurate (M-RPWAE < 0.1)
  - Not yet coupled to PFLOTRAN
  - Work ongoing to reduce error, reduce run time, and couple to PFLOTRAN

Is high-fidelity modeling at waste-package scale possible in a geologic repository performance assessment (PA) model?

https://www.deviantart.com/fractalmonster/art/For-Emz1-127400950
Labeled for noncommercial reuse

Questions?

https://www.deviantart.com/fractalmonster/art/For-Emz1-127400950
Labeled for noncommercial reuse