Surrogate models via Polynomial Chaos Expansions

Diane Donovan, Brodie Lawson, Thomas McCourt, Bevan Thompson, Stephen Tyson and Fengde Zhou

What is a surrogate model?

- A **surrogate model** approximates a computationally expensive model.
- Following the behaviour of the original model and honouring the underlying physics.
- Accurately and efficiently performing:
  - uncertainty propagation;
  - sensitivity analysis;
  - parameter finding.

What makes a good surrogate?

- Honours the underlying physics of the geological model.
- Uses a small set of training and validation data.
- Fast evaluations across the entire parameter space.
- Enables key parameter identification (sensitivity analysis).
- Respects the statistical distributions of uncertain input parameters.

Why Polynomial Chaos Expansion (PCE)?

- Surrogate models constructed by summing combinations of polynomials.
- Polynomial functions are fast to evaluate.
- Resulting response surfaces predict model output with low error.
- Choosing orthogonal polynomials reduces the complexity and allows for propagation of uncertainty in the input parameters.

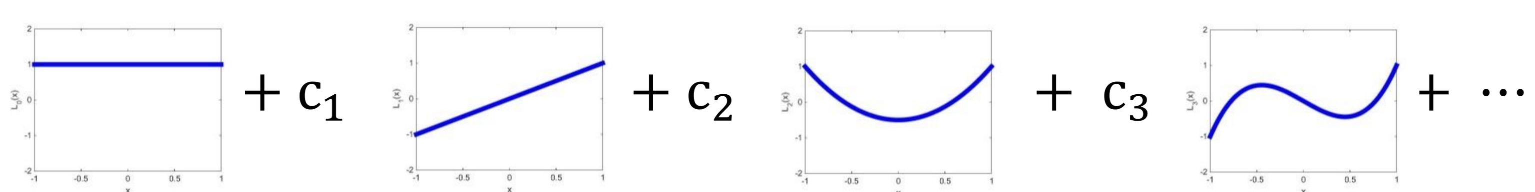
How do you construct a PCE surrogate model?

- A PCE represents the model as a sum of carefully chosen polynomials each individually **weighted** to give an accurate approximation.

The mean

Capturing how the model varies

$\mathcal{M}(x) = c_0 \quad + c_1 \quad + c_2 \quad + c_3 \quad + \cdots$



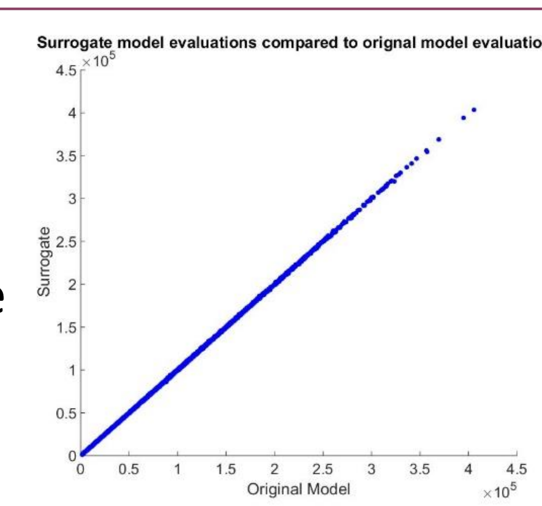
- The method naturally generalises to multiple input parameters.
- The polynomials are **orthogonal** with respect to the input parameters' statistical distributions:
  - reducing the complexity;
  - capturing the uncertainty in the input parameters;
  - allowing for efficient identification of key parameters and key parameter interactions.

How does it honour the geophysics?

- The weights $c_0, c_1, c_2, \ldots$ are derived from the underlying data (often via evaluations of the original model).

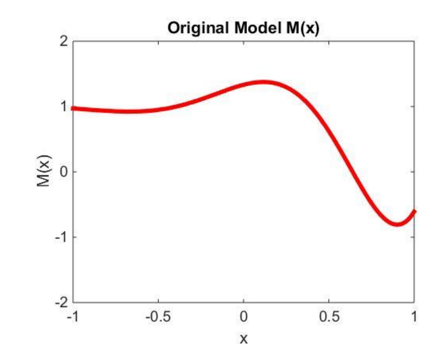
Example – PCE validation:

The first six 1D polynomials for 4 input parameters can be combined to construct a 5D response surface for a CMG model for peak gas extraction in which the mean absolute percentage error across the entire surface is 0.33 %.

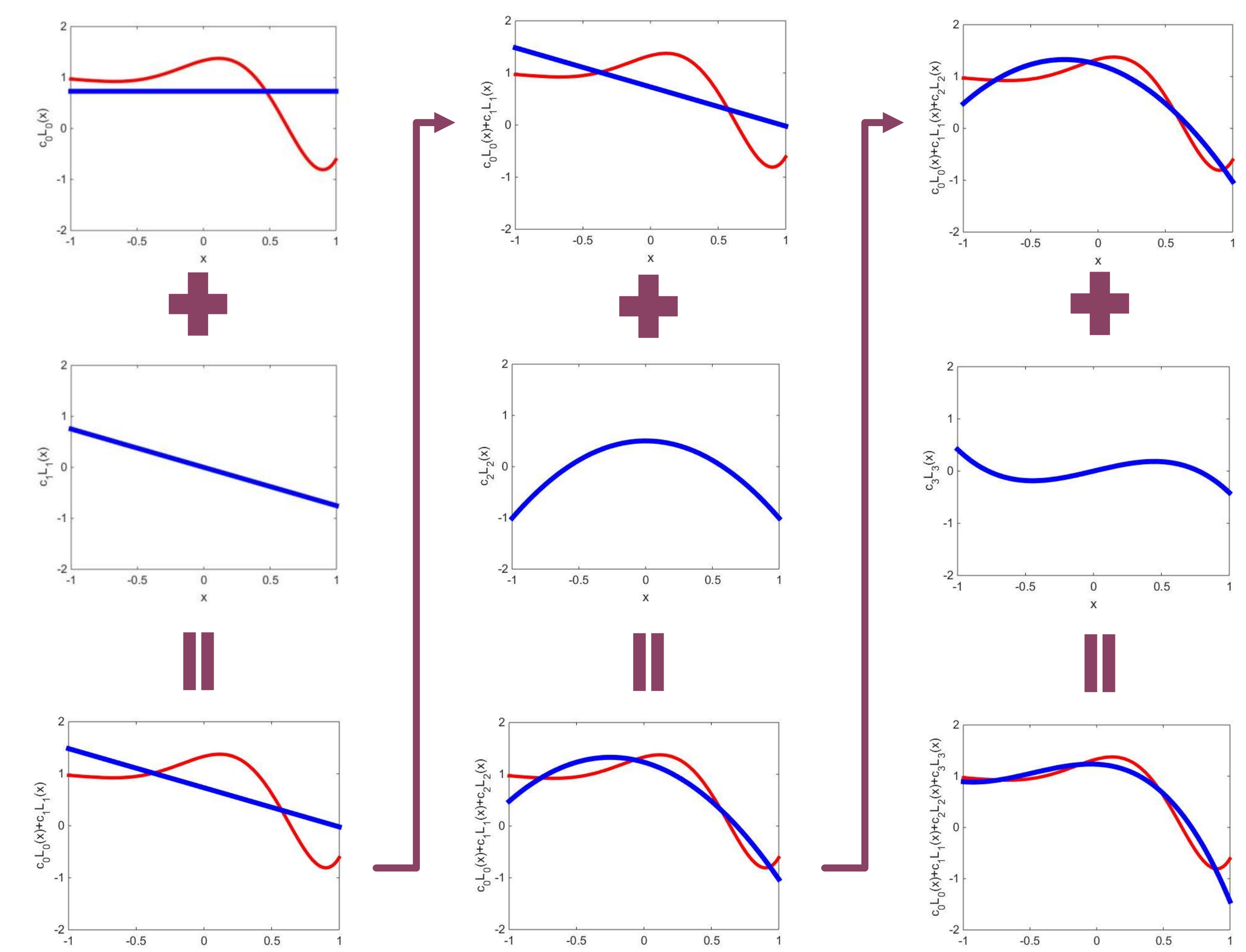


Example – A Polynomial Chaos Expansion:

A response surface for a model with a uniformly distributed uncertain input parameter on $[-1,1]$:



The incremental PCE approximation for the response surface:



What is the pay off?

Statistical information and uncertainty propagation:

- Immediately provides the mean, variance and higher moments.
- Rapidly generates cumulative distribution functions for the model outputs.

Sensitivity Analysis – identifying key parameters:

- Orthogonality allows for rapid analysis of the propagation of input parameter variance.
- Resulting **Sobol' Indices** enable identification of key parameters and key parameter interactions.

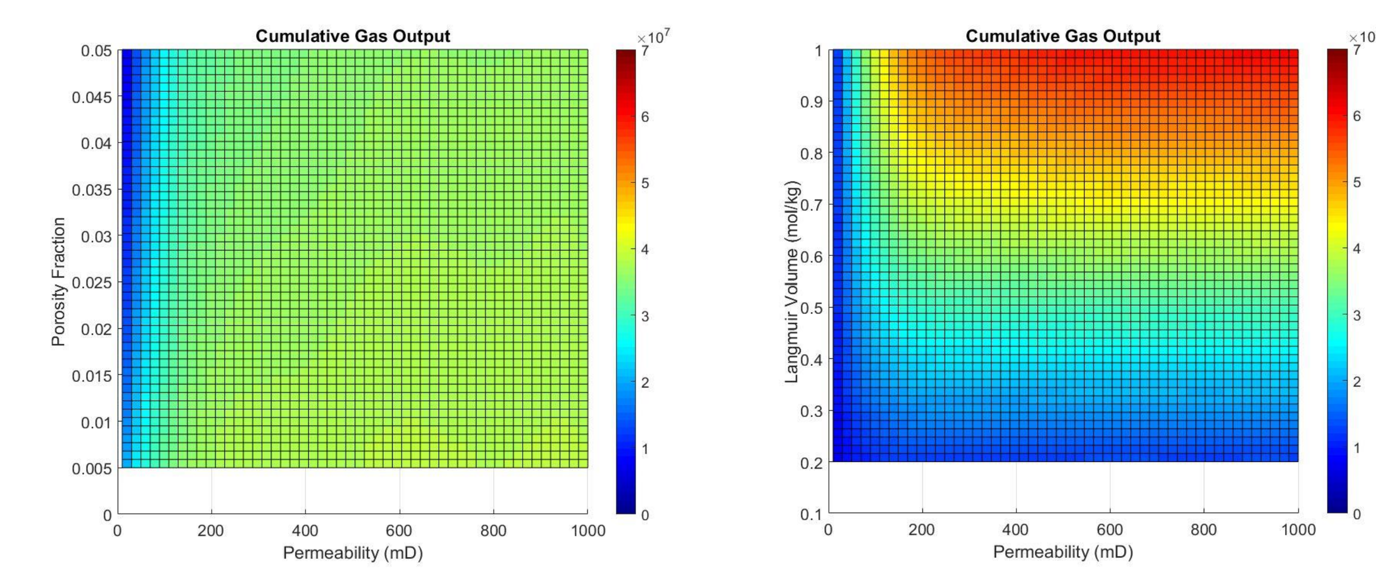
Parameter finding:

- As a PCE is fast to evaluate it enables comprehensive exploration of the response surface to conduct inverse parameter finding.

Example – Identifying Key Parameters:

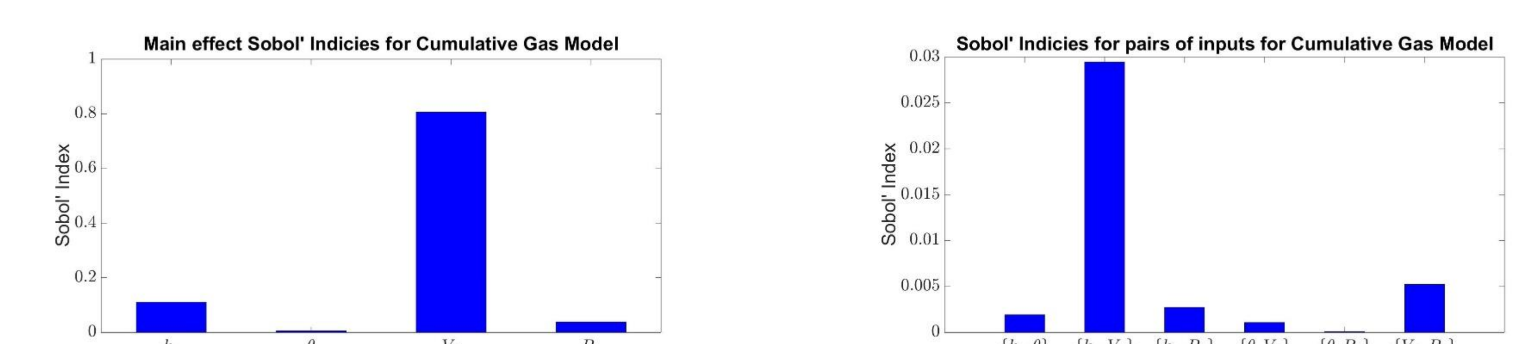
A CMG model to predict gas extraction with uncertain input parameters: fracture permeability k_x, fracture porosity ϕ, Langmuir Volume V_L and Langmuir Pressure P_L.

Plots of slices of the response surface for cumulative gas extraction:



These slices suggest certain sensitivity relationships.

A PCE easily provides a formal sensitivity analysis through the construction of Sobol' Indices, without further sampling the parameter space.



Future directions.

1. The size of the training set increases with the number of input parameters. The use of adaptive strategies and other advances in quadrature techniques will be explored to minimise this.
2. Constructing PCEs from field data, *cutting out the middleman*, i.e. no requirement for an established model.
3. Hybrid approaches combining 1 and 2.

Example – Cumulative Distribution Functions:

A CMG model for peak gas extraction, empirical CDFs plots (3000 evaluations) from the original model (taking days) and from a PCE surrogate (taking seconds).