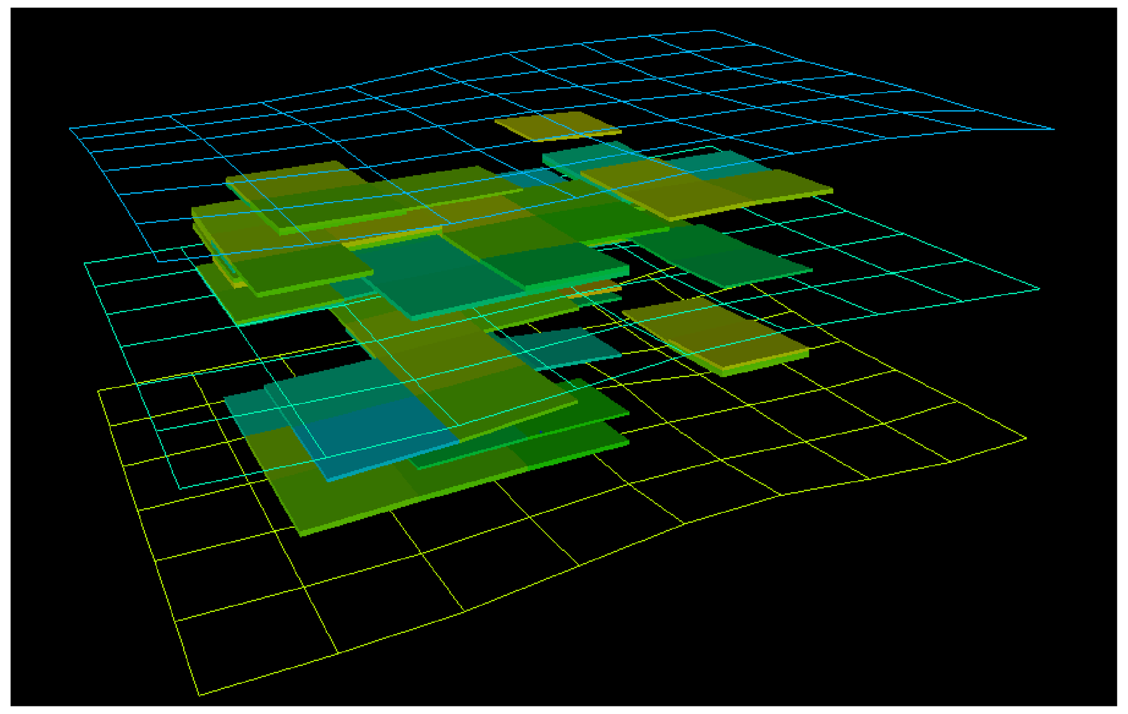
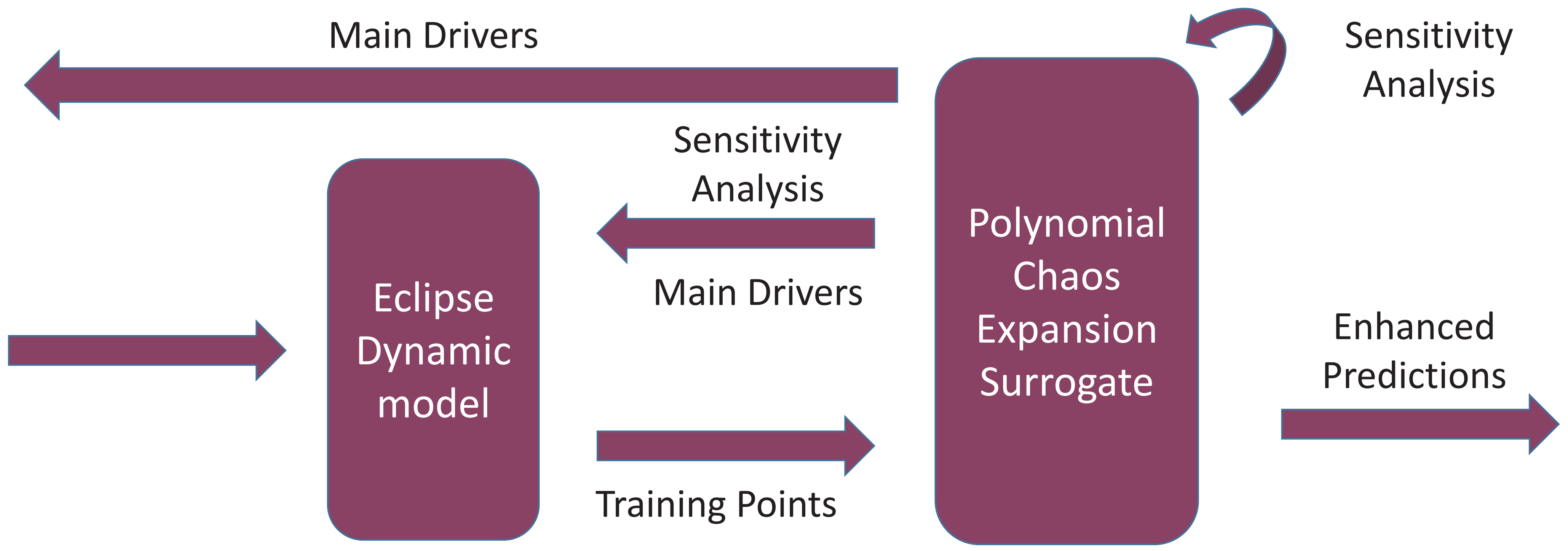
THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Machine Learning for Coal Seam Gas Production

Diane Donovan, Suzanne Hurter, Thomas McCourt, Iain Rodger, Bevan Thompson and Ryan Blackmore

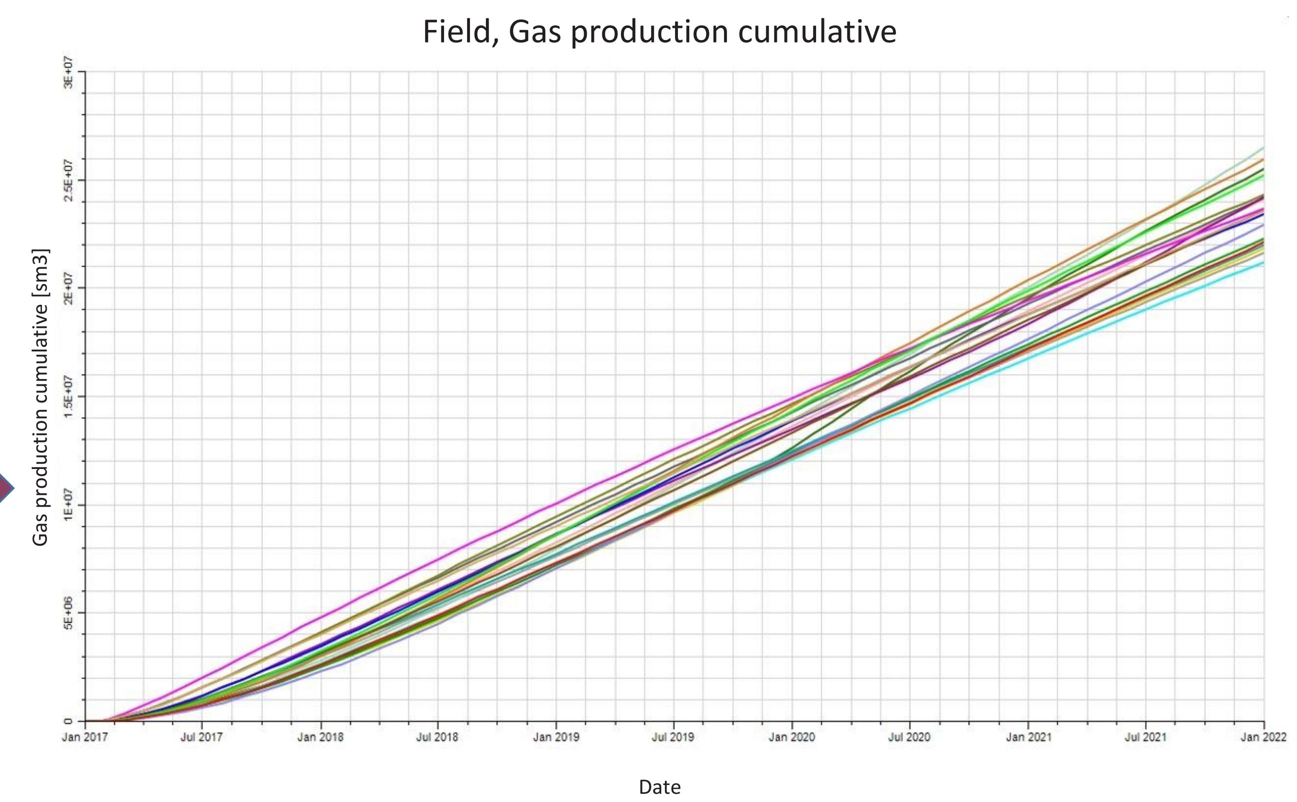


Static Reservoir Simulation Model: QDEX data
4.2km x 3km x 300m, with 5 wells



87 – 174 Active Parameters

Fast and Efficient Machine Learning



How do we apply machine learning?

- A **surrogate model** is built to approximate a computationally expensive model.
- It emulates the behaviour of the original model, honouring the underlying physics.
- It accurately and efficiently performs:
  - uncertainty propagation; and
  - sensitivity analysis.
- It facilitates processes such as EUR calculations and history matching.

What are the desirable properties?

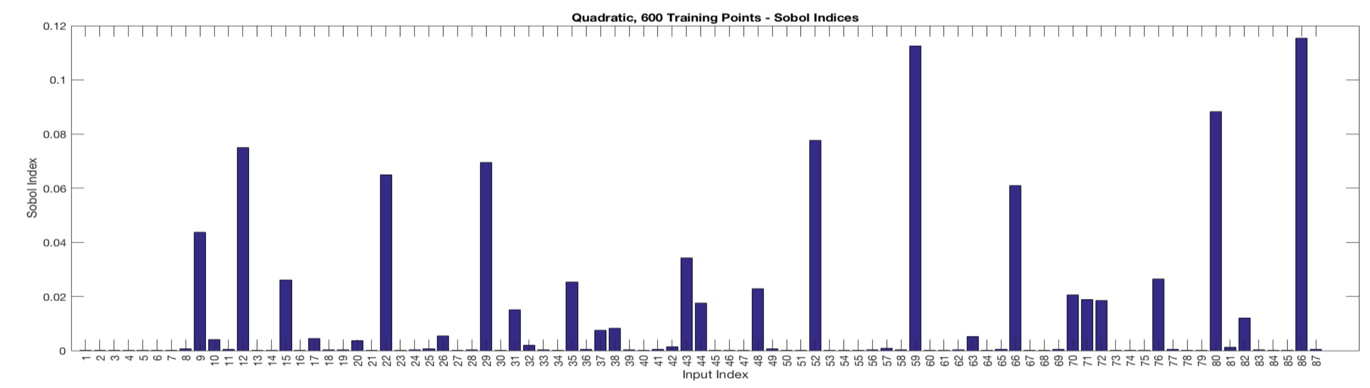
- Accurate predictions using small sets of training and validation data.
- Fast evaluations across the entire parameter space.
- Respects the statistical distributions of uncertain input parameters.
- Direct access to sensitivity analysis.

What is the pay off?

- Statistical information and uncertainty propagation: mean, variance and higher moments, and cumulative distribution functions.
- Sensitivity analysis – identifying key inputs and parameter variance.
- History matching through fast and comprehensive exploration of the response surface.

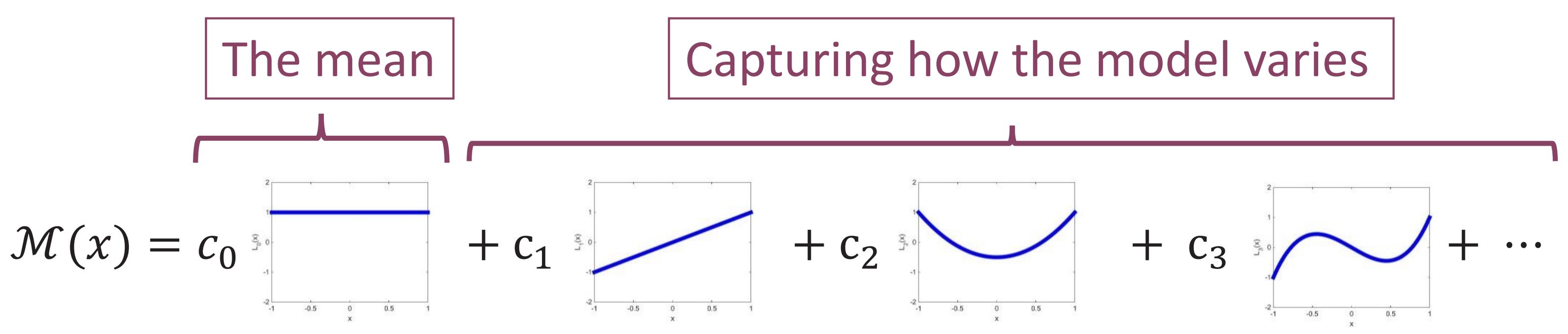
Example – Identifying the sources of uncertainty.

Sobol' Indices are used to rank the impact of the porosity and permeability of the various coal bodies (see figure top left).



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.



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

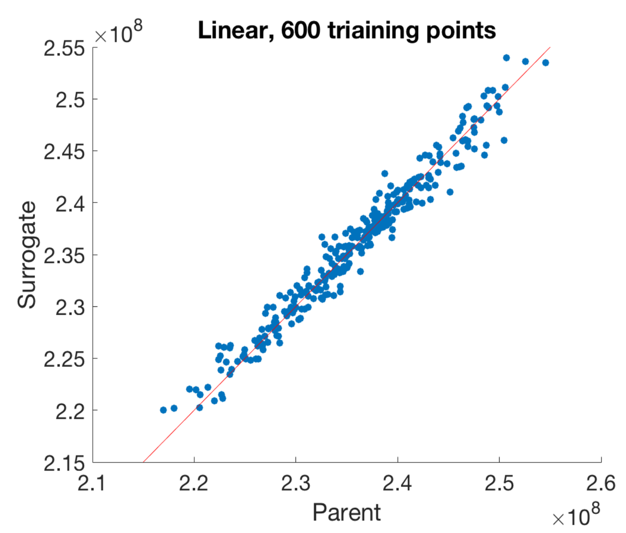
- The method naturally generalises to multiple input parameters.
- The polynomials are **orthogonal** with respect to the input parameters' statistical distributions:
  - reducing the complexity;
  - capturing the uncertainty in the input parameters;
  - allowing for efficient identification of key parameters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

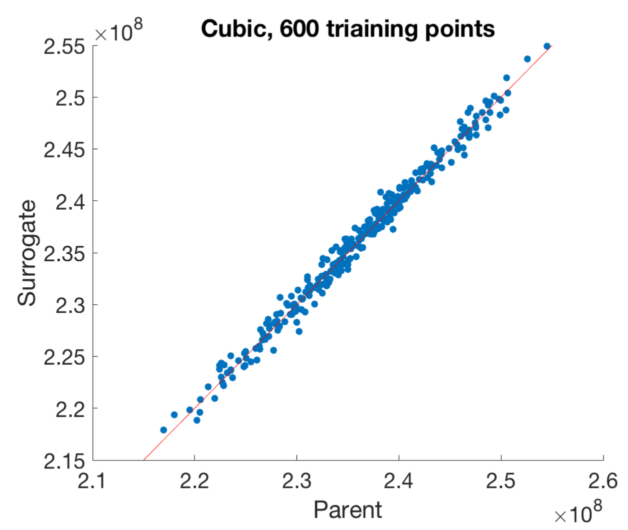
How does it honour the geophysics?

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

Example – Using a surrogate to predict cumulative gas production.

Order	Number of coefficients	RMSE x 10^6	Relative RMSE x 10^{-3}	Mean APE x 10^{-1} %	Median APE x 10^{-1} %
1	88	1.3967	5.9249	4.5627	3.6436
2	175	1.1634	4.9325	3.8976	3.2392
3	300	0.8599	3.6569	2.8667	2.3450





How do we optimise the construction process?

- Construct surrogate models using moments or approximations of the moments for the inputs, thus allowing for unknown distributions.
- Use regression techniques for approximating key coefficients, thus reducing the required number of training points.
- Two types of regression techniques to solve the same minimum argument equation; Ordinary Least Squares (OLS) and Least Angle Regression (LARS).
- LARS is preferred for higher dimensionality cases as it preferences the 'most important' coefficients and hence can generate a higher order surrogate model.

Future directions.

1. Exploring the relationship between the size of the training set, the number of input parameters and the accuracy of the surrogate model.
2. Machine learning from field data, *cutting out the middleman*, i.e. no requirement for an established model.

Schlumberger is acknowledged for providing Petrel and Eclipse software. This research receives funding and support from the UQ Centre for Coal Seam Gas and its industry partners (Arrow Energy, APLNG, Santos and QGC/Shell). Energi Simulation is thanked for partial funding support.

UQ-CCSG
Centre for Coal Seam Gas