Polynomial Chaos for the Geosciences

Diane Donovan, Steve Tyson, Brodie Lawson, Marvin Tas, Bevan Thompson

Abstract

Uncertainty in input parameters is a serious challenge to traditional simulation-based methods for model prediction in geoscience and engineering. For example given limited subsurface information, significant additional computational costs are incurred when applying Monte Carlo methods to predict subsurface flows in aquifers and reservoirs and to study the effect of upscaling of parameters.

This project seeks to apply Polynomial Chaos Expansion techniques to quantify uncertainty in model inputs, providing significantly faster model predictions with reduced computational costs.

Some Questions

- Can Polynomial Chaos Expansions (PCE) take flow rates or pressure drop data and determine porosity/permeability?
- Can PCE translate cumulative probability or confidence intervals for model inputs to model outputs?
- How will this translate to better workflows for multi-phase subsurface flows?
- How will this translate into enhanced predictions for reservoir capacity and recovery?

A Problem and Method

A simple model to describe contaminant concentration, *C,* in groundwater given advection and dispersion, at any time *t,* and point *x* in space.

$$R_f \frac{\partial C}{\partial t} = D_L \frac{\partial^2 C}{\partial x^2} - V_W \frac{\partial C}{\partial x}.$$

R_f, D_L and V_W involve uncertain input parameters, here hydraulic conductivity *K*, and organic carbon partition coefficient K_{oc}. PCE captures uncertainty through a sum

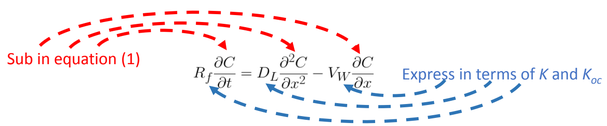
$$C(K, K_{oc}) \approx \sum_{i=0}^{m} \sum_{j=0}^{n} C_{ij} P_i(K) P_j(K_{oc}). \qquad (1)$$

where the P_i's are orthogonal polynomials. The statistical distribution of the parameters determines the form of these polynomials P_i and if required they can be constructed from experimental data. The C_{ij}'s, are determined by either *intrusive* or *non-intrusive* methods.

PCE gives a good approximation to the model output, the contamination *C,* subject to the unknown input variables *K* and K_{oc}. Increasing the values of *m* and *n* increases the accuracy of this approximation – however this also increases computational time.

Intrusive Polynomial Chaos

Intrusive PCE operates by substituting equation (1) into the model equation.



Resolving in the direction of the 'basis vectors' P_i gives a matrix system (***C*** a vector in C_{ij}'s) – a bigger system to be solved, but only once.

$$A \frac{\partial \mathbf{C}}{\partial t} = B \frac{\partial^2 \mathbf{C}}{\partial x^2} - D \frac{\partial \mathbf{C}}{\partial x}$$

Non-Intrusive Polynomial Chaos

Non-intrusive PCE takes expectations through equation (1) and uses numerical integration to find the C_{ij}'s. Using a quadrature rule, the C_{ij}'s are calculated as

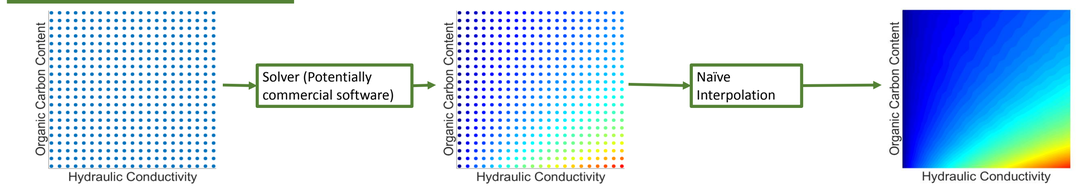
Evaluated at quadrature points using '**off the shelf solver**' — Random variable pdf's

$$C_{ij} = \frac{\sum_{\alpha=0}^{m+1} \sum_{\beta=0}^{n+1} C(K_\alpha, K_{oc\beta})\, P_i(K_\alpha)\, P_j(K_{oc\beta})\, \omega(K_\alpha)\, \omega(K_{oc\beta})\, w_\alpha w_\beta}{\langle P_i^2 \rangle \langle P_j^2 \rangle} \qquad (2)$$

A Comparison of Methods for Uncertainty Propagation

For the example, the dependence of the penetration distance of contaminant *C* on the hydraulic conductivity *K* and organic carbon partition coefficient K_{oc} can be characterised by a parameter map (on the right) built from an 'off the shelf solver'. The map can be used to identify parameter sets that maximise (minimise) the predicted value of desirable (undesirable) quantities.

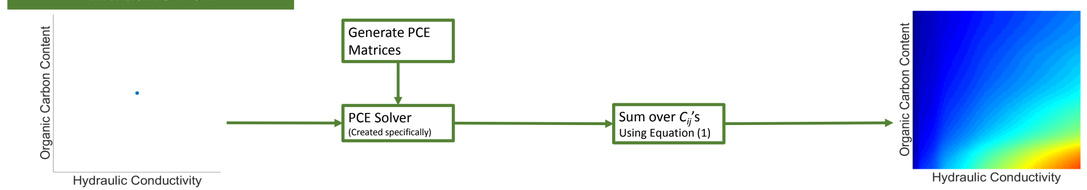
Monte Carlo



The solution is generated by running the solver for a large number of points. No approximations are introduced, except when interpolating between the points – a very fine mesh of points gives 'exact' results, when the solver can be trusted.

Runtime: 19.4 seconds (Resolution as pictured) **461 seconds (High Resolution)**

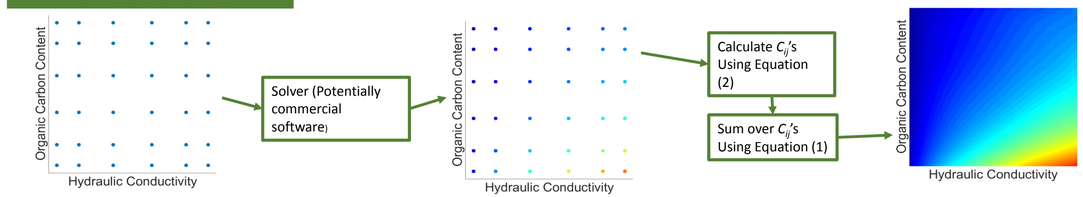
Intrusive PCE



A one off solver is used to solve a matrix system. Accuracy is obtained by increasing the order of the polynomials, but at the computational cost of solving a complex matrix system.

Runtime: 168.8 seconds (High Resolution)

Non-Intrusive PCE



The model problem is solved at a few 'intelligently' chosen points, using an '**off the shelf solver**' and C_{ij}'s are calculated. Overall this is computationally very cheap, with the visualised surface matching the high-resolution Monte Carlo map almost exactly.

Runtime: 1.7 seconds (High Resolution)

Results

Non-intrusive PCE captures the dependence of contaminant penetration on hydraulic conductivity and organic carbon partition coefficient with far less computational load than a Monte Carlo approach, whilst still using the same black box solver. Because the parameter dependence is explicitly stated in terms of polynomials, the expected output for *any* set of input parameters can be calculated without further interpolation or grid refinement.
A trade-off is that the PCE approximation may require 'sufficiently' large values of *m* and *n* for accuracy. Here, *m* = *n* = 5 proved sufficient.

Non-intrusive PCE provides a 'smart' interpolation method, needing far fewer simulations to generate an accurate map of the parameter space.
An additional benefit is that the mean and variance of model outputs (in response to uncertain inputs) can be calculated explicitly from the Polynomial Chaos Expansion.

Application – Parameters from Data

As new field data becomes available models need to be recalibrated. Here the problem is quantify the *error* between model predictions and new data. In this case, the parameter space of uncertain inputs must still be mapped, but the focus is on using PCE to predict parameter values that minimise error.

Accurate estimates of these physical properties are very important for correctly predicting total site output.

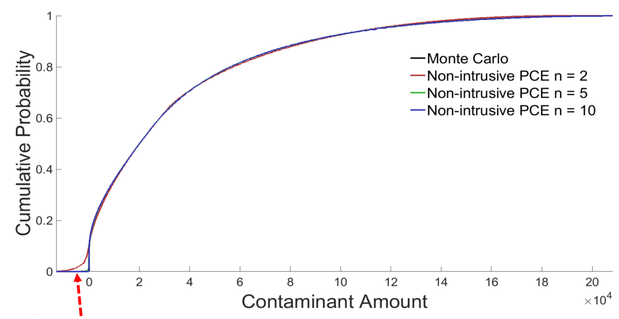
True Values: K = 7.70 x 10^{-4}, K_{oc} = 254.6			
Method	**Predicted K**	**Predicted K_{oc}**	**Runtime (s)**
Interpolated Monte Carlo, 11x11 pts.	8.48 x 10^{-4}	285.0	19.8
Interpolated Monte Carlo, 51x51 pts.	7.60 x 10^{-4}	250.4	123.5
Interpolated Monte Carlo, 101x101 pts.	7.60 x 10^{-4}	250.4	451.3
Intrusive PCE, n = 2	8.46 x 10^{-4}	284	88.2
Intrusive PCE, n = 5	7.88 x 10^{-4}	261.9	192.2
Non-intrusive PCE, n = 2	8.18 x 10^{-4}	270.6	2.1
Non-intrusive PCE, n = 5	7.88 x 10^{-4}	261.9	6.5
Non-intrusive PCE, n = 10	7.74 x 10^{-4}	256.2	20.8

Non-intrusive PCE predicts, in a significantly shorter time, the unknown hydraulic conductivity and organic carbon partition coefficient with good accuracy.

Successive iterations of PCE on smaller and smaller parameter space 'windows' provides fast refinement of the approximation.

Application – CDF Generation

The distributions of the uncertain input parameters can be used to generate statistical information about the solution. For example, it might be desirable to calculate confidence intervals for contaminant spread – specifically how much contaminant will leech past a certain point. Since the PCE is built from the distributions of the uncertain input variables, the CDFs of model outputs are easily generated.



The Monte Carlo CDF cannot be seen because the higher order PCE predictions match almost exactly. Generation of the highest order PCE required only 17.6 seconds, as opposed to 388.5 seconds for Monte Carlo.