Sparse Polynomial Chaos Surrogate for ACME Land Model via Iterative Bayesian Compressive Sensing

K. Sargsyan¹, D. Ricciuto², P. Thornton²
C. Safta¹, B. Debusschere¹, H. Najm¹

¹Sandia National Laboratories
Livermore, CA

²Oak Ridge National Laboratory
Oak Ridge, TN

Sponsored by DOE, Biological and Environmental Research, under Accelerated Climate Modeling for Energy (ACME).

Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-AC04-94AL85000.
• Surrogates needed for complex models

• Polynomial Chaos (PC) surrogates do well with uncertain inputs

• Bayesian regression provide results with uncertainty certificate

• Compressive sensing ideas deal with high-dimensionality
A single-site, 1000-yr simulation takes \( \sim 10 \) hrs on 1 CPU

- Involves \( \sim 70 \) input parameters; some dependent
- Non-smooth input-output relationship
Surrogate construction: scope and challenges

Construct surrogate for a complex model \( f(\lambda) \) to enable

- Global sensitivity analysis
- Optimization
- Forward uncertainty propagation
- Input parameter calibration
- ...

- Computationally expensive model simulations, data sparsity
  - Need to build accurate surrogates with as few training runs as possible
- High-dimensional input space
  - Too many samples needed to cover the space
  - Too many terms in the polynomial expansion
Surrogate construction: scope and challenges

Construct surrogate for a complex model $f(\lambda)$ to enable

- Global sensitivity analysis
- Optimization
- Forward uncertainty propagation
- Input parameter calibration
- ...

- Computationally expensive model simulations, data sparsity
  - Need to build accurate surrogates with as few training runs as possible

- High-dimensional input space
  - Too many samples needed to cover the space
  - Too many terms in the polynomial expansion
Surrogate construction: scope and challenges

Construct surrogate for a complex model $f(\lambda)$ to enable

- Global sensitivity analysis
- Optimization
- Forward uncertainty propagation
- Input parameter calibration
- ... 

- Computationally expensive model simulations, data sparsity
  - Need to build accurate surrogates with as few training runs as possible

- High-dimensional input space
  - Too many samples needed to cover the space
  - Too many terms in the polynomial expansion
Polynomial Chaos surrogate for $f(\lambda)$

- Scale the input parameters $\lambda_i \in [a_i, b_i]$

$$\lambda_i = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{2} x_i$$

- Forward function $f(\cdot)$, output $u$

$$u(x) = f(\lambda(x)) \approx \sum_{k=0}^{K-1} c_k \Psi_k(x) \equiv g(x)$$

- Global sensitivity information for free
  - Sobol indices, variance-based decomposition.

- Bayesian inference useful for finding $c_k$:

$$P(c|u) \propto P(u|c)P(c)$$
Polynomial Chaos surrogate for $f(\lambda)$

- Scale the input parameters $\lambda_i \in [a_i, b_i]$

  $$\lambda_i = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{2} x_i$$

- Forward function $f(\cdot)$, output $u$

  $$u(x) = f(\lambda(x)) \approx \sum_{k=0}^{K-1} c_k \Psi_k(x) \equiv g(x)$$

- Global sensitivity information for free
  - Sobol indices, variance-based decomposition.

- Bayesian inference useful for finding $c_k$:

  $$P(c|u) \propto P(u|c)P(c)$$
Polynomial Chaos surrogate for $f(\lambda)$

- Scale the input parameters $\lambda_i \in [a_i, b_i]$

$$\lambda_i = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{2} x_i$$

- Forward function $f(\cdot)$, output $u$

$$u(x) = f(\lambda(x)) \approx \sum_{k=0}^{K-1} c_k \Psi_k(x) \equiv g(x)$$

- Global sensitivity information for free
  - Sobol indices, variance-based decomposition.

- Bayesian inference useful for finding $c_k$:

$$P(c|u) \propto P(u|c)P(c)$$
Polynomial Chaos surrogate for $f(\lambda)$

- Scale the input parameters $\lambda_i \in [a_i, b_i]$
  \[
  \lambda_i = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{2} x_i
  \]

- Forward function $f(\cdot)$, output $u$
  \[
  u(x) = f(\lambda(x)) \approx \sum_{k=0}^{K-1} c_k \Psi_k(x) \equiv g(x)
  \]

- Global sensitivity information for free
  - Sobol indices, variance-based decomposition.

- Bayesian inference useful for finding $c_k$:
  \[
  P(c|u) \propto P(u|c)P(c)
  \]
Bayesian inference of PC surrogate: high-d, low-data regime

\[ y = u(x) \approx \sum_{k=0}^{K-1} c_k \Psi_k(x) \]

\[ \Psi_k(x_1, x_2, \ldots, x_d) = \psi_k_1(x_1) \psi_k_2(x_2) \cdots \psi_k_d(x_d) \]

• **Issues:**
  - how to properly choose the basis set?
  - need to work in underdetermined regime \( N < K \): fewer data than bases (d.o.f.)

• Discover the underlying low-d structure in the model
  - get help from the machine learning community
Bayesian inference of PC surrogate: high-d, low-data regime

\[ y = u(x) \approx \sum_{k=0}^{K-1} c_k \Psi_k(x) \]

\[ \Psi_k(x_1, x_2, \ldots, x_d) = \psi_{k_1}(x_1) \psi_{k_2}(x_2) \cdots \psi_{k_d}(x_d) \]

- Issues:
  - how to properly choose the basis set?
  - need to work in underdetermined regime \( N < K \): fewer data than bases (d.o.f.)

- Discover the underlying low-d structure in the model
  - get help from the machine learning community
Bayesian inference of PC surrogate: high-d, low-data regime

\[ y = u(x) \approx \sum_{k=0}^{K-1} c_k \Psi_k(x) \]

\[ \Psi_k(x_1, x_2, \ldots, x_d) = \psi_{k_1}(x_1) \psi_{k_2}(x_2) \cdots \psi_{k_d}(x_d) \]

- Issues:
  - how to properly choose the basis set?
  - need to work in underdetermined regime \( N < K \): fewer data than bases (d.o.f.)
  - Discover the underlying low-d structure in the model
  - get help from the machine learning community
Bayesian inference of PC surrogate: high-d, low-data regime

\[ y = u(x) \approx \sum_{k=0}^{K-1} c_k \Psi_k(x) \]

\[ \Psi_k(x_1, x_2, \ldots, x_d) = \psi_{k_1}(x_1) \psi_{k_2}(x_2) \cdots \psi_{k_d}(x_d) \]

- Issues:
  - how to properly choose the basis set?
  - need to work in underdetermined regime
    \[ N < K: \text{fewer data than bases (d.o.f.)} \]

- Discover the underlying low-d structure in the model
  - get help from the machine learning community

K. Sargsyan (ksargsy@sandia.gov)  AGU Fall Meeting 2015  Dec 18, 2015
Bayesian inference of PC surrogate: high-d, low-data regime

\[ y = u(x) \approx \sum_{k=0}^{K-1} c_k \Psi_k(x) \]

\[ \Psi_k(x_1, x_2, \ldots, x_d) = \psi_{k_1}(x_1) \psi_{k_2}(x_2) \cdots \psi_{k_d}(x_d) \]

- Issues:
  - how to properly choose the basis set?
  - need to work in underdetermined regime \( N < K \): fewer data than bases (d.o.f.)
  - Discover the underlying low-d structure in the model
  - get help from the machine learning community
In a different language....

- $N$ training data points $(x_n, u_n)$ and $K$ basis terms $\Psi_k(\cdot)$
- Projection matrix $P^{N\times K}$ with $P_{nk} = \Psi_k(x_n)$
- Find regression weights $c = (c_0, \ldots, c_{K-1})$ so that

$$u \approx Pc \quad \text{or} \quad u_n \approx \sum_k c_k \Psi_k(x_n)$$

- The number of polynomial basis terms grows fast; a $p$-th order, $d$-dimensional basis has a total of $K = (p + d)!/(p!d!)$ terms.
- For limited data and large basis set ($N < K$) this is a sparse signal recovery problem ⇒ need some regularization/constraints.

- Least-squares
  
  $$\text{argmin}_c \left\{ \| u - Pc \|_2 \right\}$$

- The ‘sparsest’
  
  $$\text{argmin}_c \left\{ \| u - Pc \|_2 + \alpha \| c \|_0 \right\}$$

- Compressive sensing
  
  $$\text{argmin}_c \left\{ \| u - Pc \|_2 + \alpha \| c \|_1 \right\}$$
• \( N \) training data points \((x_n, u_n)\) and \( K \) basis terms \( \Psi_k(\cdot) \)
• Projection matrix \( P^{N \times K} \) with \( P_{nk} = \Psi_k(x_n) \)
• Find regression weights \( c = (c_0, \ldots, c_{K-1}) \) so that

\[
\begin{align*}
\mathbf{u} & \approx \mathbf{Pc} & \text{or} & & \mathbf{u}_n \approx \sum_k c_k \Psi_k(x_n) \\
\end{align*}
\]

• The number of polynomial basis terms grows fast; a \( p \)-th order, \( d \)-dimensional basis has a total of \( K = (p + d)!/(p!d!) \) terms.
• For limited data and large basis set \((N < K)\) this is a sparse signal recovery problem \( \Rightarrow \) need some regularization/constraints.

• Least-squares \( \text{argmin}_c \{ \| \mathbf{u} - \mathbf{Pc} \|_2 \} \)
• The ‘sparsest’ \( \text{argmin}_c \{ \| \mathbf{u} - \mathbf{Pc} \|_2 + \alpha \| c \|_0 \} \)
• Compressive sensing \( \text{Bayesian} \quad \text{argmin}_c \{ \| \mathbf{u} - \mathbf{Pc} \|_2 + \alpha \| c \|_1 \} \quad \text{Likelihood} \quad \text{Prior} \)
BCS removes unnecessary basis terms

\[ f(x, y) = \cos(x + 4y) \]

\[ f(x, y) = \cos(x^2 + 4y) \]

The square \((i, j)\) represents the (log) spectral coefficient for the basis term \(\psi_i(x)\psi_j(y)\).
Iterative Bayesian Compressive Sensing (iBCS)

- *Iterative BCS*: We implement an iterative procedure that allows increasing the order for the relevant basis terms while maintaining the dimensionality reduction [Sargsyan et al. 2014], [Jakeman et al. 2015].
- Combine basis growth and reweighting!
Basis set growth: simple anisotropic function
Basis set growth: ... added outlier term
Application of Interest: ACME Land Model

http://www.cesm.ucar.edu/models/clm/

- A single-site, 1000-yr simulation takes \( \sim 10 \) hrs on 1 CPU
- Involves \( \sim 70 \) input parameters; some dependent
- Non-smooth input-output relationship
FLUXNET experiment

- 96 FLUXNET sites covering major biomes and plant functional types
- Varying 68 input parameters over given ranges; 5 steady state outputs
- Ensemble of 3000 runs on Titan, DoE Leadership Computing Facility at Oak Ridge National Lab
Sparse PC surrogate and uncertainty decomposition for the ACME Land Model

- Main effect sensitivities: rank input parameters
- Joint sensitivities: most influential input couplings
- About 200 polynomial basis terms in the 68-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data

Site # 31
GPP

Site # 31
TLAI
Sparse PC surrogate and uncertainty decomposition for the ACME Land Model

- Main effect sensitivities: rank input parameters
- Joint sensitivities: most influential input couplings
- About 200 polynomial basis terms in the 68-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data

Site # 31

TOTVEGC

Site # 31

TOTSOMC
Sparse PC surrogate and uncertainty decomposition for the ACME Land Model

- Main effect sensitivities: rank input parameters
- Joint sensitivities: most influential input couplings
- About 200 polynomial basis terms in the 68-dimensional space
- Sparse PC will further be used for:
  - sampling in a reduced space
  - parameter calibration against experimental data
Sparse PC surrogate and uncertainty decomposition for the ACME Land Model

- Main effect sensitivities: rank input parameters
- Joint sensitivities: most influential input couplings
- About 200 polynomial basis terms in the 68-dimensional space
- Sparse PC will further be used for:
  - sampling in a reduced space
  - parameter calibration against experimental data

- GPP
  gross primary productivity

Sensitivity

K. Sargsyan (ksargsy@sandia.gov)
Sparse PC surrogate and uncertainty decomposition for the ACME Land Model

- Main effect sensitivities: rank input parameters
- Joint sensitivities: most influential input couplings
- About 200 polynomial basis terms in the 68-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data

TLAI
leaf area index

K. Sargsyan (ksargsy@sandia.gov) AGU Fall Meeting 2015 Dec 18, 2015
Sparse PC surrogate and uncertainty decomposition for the ACME Land Model

- Main effect sensitivities: rank input parameters
- Joint sensitivities: most influential input couplings
- About 200 polynomial basis terms in the 68-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data

- TOTVEGC
  vegetation carbon
Sparse PC surrogate and uncertainty decomposition for the ACME Land Model

- Main effect sensitivities: rank input parameters
- Joint sensitivities: most influential input couplings
- About 200 polynomial basis terms in the 68-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data

- EFLX\_LH\_TOT
  latent heat flux
Sparse PC surrogate and uncertainty decomposition for the ACME Land Model

- Main effect sensitivities: rank input parameters
- Joint sensitivities: most influential input couplings
- About 200 polynomial basis terms in the 68-dimensional space
- Sparse PC will further be used for
  - sampling in a reduced space
  - parameter calibration against experimental data

- TOTSOMC
  soil organic matter carbon
Summary

- **Surrogate** models are necessary for complex models
  - Replace the full model for both forward and inverse UQ

- **Uncertain inputs**
  - **Polynomial Chaos** surrogates well-suited

- **Limited training dataset**
  - **Bayesian** methods handle limited information well

- **Curse of dimensionality**
  - The hope is that not too many dimensions matter
  - Compressive sensing (CS) ideas ported from machine learning
  - We implemented *iteratively reweighting Bayesian CS* algorithm that reduces dimensionality and increases order on-the-fly.

---

- **Open issues**
  - Computational design. What is the best sampling strategy?
  - Overfitting still present. Cross-validation techniques help.

---

- **Software**: employed SNL-CA UQ library UQTk (www.sandia.gov/uqtoolkit)
Literature


Random variables represented by Polynomial Chaos

$$X \simeq \sum_{k=0}^{K-1} c_k \Psi_k(\eta)$$

- $\eta = (\eta_1, \cdots, \eta_d)$ standard i.i.d. r.v.
- $\Psi_k$ standard polynomials, orthogonal w.r.t. $\pi(\eta)$.

$$\Psi_k(\eta_1, \eta_2, \cdots, \eta_d) = \psi_{k_1}(\eta_1) \psi_{k_2}(\eta_2) \cdots \psi_{k_d}(\eta_d)$$

- Typical truncation rule: total-order $p$, $k_1 + k_2 + \cdots k_d \leq p$.
  Number of terms is $K = \frac{(d+p)!}{d!p!}$.

- Essentially, a parameterization of a r.v. by deterministic spectral modes $c_k$.

- Most common standard Polynomial-Variable pairs:
  (continuous) Gauss-Hermite, Legendre-Uniform,
  (discrete) Poisson-Charlier.

[Wiener, 1938; Ghanem & Spanos, 1991; Xiu & Karniadakis, 2002; Le Maître & Knio, 2010]
Bayesian inference of PC surrogate

\[ u \simeq \sum_{k=0}^{K-1} c_k \Psi_k(x) \equiv g_c(x) \]

- **Data** consists of *training runs*

\[ \mathcal{D} \equiv \{ (x_i, u_i) \}_{i=1}^N \]

- **Likelihood** with a gaussian noise model with \( \sigma^2 \) fixed or inferred,

\[ L(c) = P(\mathcal{D}|c) = \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^N \prod_{i=1}^{N} \exp \left( -\frac{(u_i - g_c(x))^2}{2\sigma^2} \right) \]

- **Prior** on \( c \) is chosen to be conjugate, uniform or gaussian.

- **Posterior** is a *multivariate normal*

\[ c \in \mathcal{MN}(\mu, \Sigma) \]

- The (uncertain) surrogate is an *gaussian process*

\[ \sum_{k=0}^{K-1} c_k \Psi_k(x) = \Psi(x)^T c \in \mathcal{GP}(\Psi(x)^T \mu, \Psi(x) \Sigma \Psi(x')^T) \]
Sensitivity information comes free with PC surrogate,

\[ g(x_1, \ldots, x_d) = \sum_{k=0}^{K-1} c_k \Psi_k(x) \]

- Main effect sensitivity indices

\[
S_i = \frac{\text{Var}[\mathbb{E}(g(x|x_i))]}{\text{Var}[g(x)]} = \frac{\sum_{k \in \mathbb{I}_i} c_k^2 ||\Psi_k||^2}{\sum_{k > 0} c_k^2 ||\Psi_k||^2}
\]

\(\mathbb{I}_i\) is the set of bases with only \(x_i\) involved
Sensitivity information comes free with PC surrogate, \[ g(x_1, \ldots, x_d) = \sum_{k=0}^{K-1} c_k \Psi_k(x) \]

- Main effect sensitivity indices
  \[ S_i = \frac{\text{Var}[\mathbb{E}(g(x|x_i))] \text{Var}[g(x)]}{\text{Var}[g(x)]} = \frac{\sum_{k \in \Pi_i} c_k^2 \|\Psi_k\|^2}{\sum_{k > 0} c_k^2 \|\Psi_k\|^2} \]

- Joint sensitivity indices
  \[ S_{ij} = \frac{\text{Var}[\mathbb{E}(g(x|x_i, x_j))] \text{Var}[g(x)]}{\text{Var}[g(x)]} - S_i - S_j = \frac{\sum_{k \in \Pi_{ij}} c_k^2 \|\Psi_k\|^2}{\sum_{k > 0} c_k^2 \|\Psi_k\|^2} \]

\( \Pi_{ij} \) is the set of bases with only \( x_i \) and \( x_j \) involved
Sensitivity information comes free with PC surrogate, but not with piecewise PC

\[ g(x_1, \ldots, x_d) = \sum_{k=0}^{K-1} c_k \Psi_k(x) \]

- **Main effect sensitivity indices**

\[ S_i = \frac{\text{Var}[\mathbb{E}(g(x|x_i))]}{\text{Var}[g(x)]} = \frac{\sum_{k \in I_i} c_k^2 ||\Psi_k||^2}{\sum_{k>0} c_k^2 ||\Psi_k||^2} \]

- **Joint sensitivity indices**

\[ S_{ij} = \frac{\text{Var}[\mathbb{E}(g(x|x_i, x_j))] - S_i - S_j}{\text{Var}[g(x)]} = \frac{\sum_{k \in I_{ij}} c_k^2 ||\Psi_k||^2}{\sum_{k>0} c_k^2 ||\Psi_k||^2} \]

- For piecewise PC, need to resort to Monte-Carlo estimation [Saltelli, 2002].
Basis normalization helps the success rate
Input correlations: Rosenblatt transformation

- Rosenblatt transformation maps any (not necessarily independent) set of random variables \( \lambda = (\lambda_1, \ldots, \lambda_d) \) to uniform i.i.d.’s \( \{x_i\}_{i=1}^d \) [Rosenblatt, 1952].

\[
x_1 = F_1(\lambda_1) \\
x_2 = F_{2|1}(\lambda_2 | \lambda_1) \\
x_3 = F_{3|2,1}(\lambda_3 | \lambda_2, \lambda_1) \\
\vdots \\
x_d = F_{d|d-1,\ldots,1}(\lambda_d | \lambda_{d-1}, \ldots, \lambda_1)
\]

- Inverse Rosenblatt transformation \( \lambda = R^{-1}(x) \) ensures a well-defined input PC construction

\[
\lambda_i = \sum_{k=0}^{K-1} \lambda_{ik} \Psi_k(x)
\]

- Caveat: the conditional distributions are often hard to evaluate accurately.
Strong discontinuities/nonlinearities challenge global polynomial expansions

- Basis enrichment [Ghosh & Ghanem, 2005]

- Stochastic domain decomposition
  - Wiener-Haar expansions,
    Multiblock expansions,
    also known as Multielement PC [Wan & Karniadakis, 2009]

- Smart splitting, discontinuity detection
  [Archibald et al, 2009; Chantrasmi, 2011; Sargsyan et al, 2011; Jakeman et al, 2012]

- Data domain decomposition,
  - Mixture PC expansions [Sargsyan et al, 2010]

- Data clustering, classification,
  - Piecewise PC expansions
Cluster the training dataset into non-overlapping subsets $\mathcal{D}_1$ and $\mathcal{D}_2$, where the behavior of function is smoother.

Construct global PC expansions $g_i(x) = \sum_k c_{ik} \Psi_k(x)$ using each dataset individually ($i = 1, 2$).

Declare a surrogate

$$g_s(x) = \begin{cases} 
g_1(x) & \text{if } x \in^* \mathcal{D}_1 \\
g_2(x) & \text{if } x \in^* \mathcal{D}_2 
\end{cases}$$

* Requires a classification step to find out which cluster $x$ belongs to. We applied Random Decision Forests (RDF).

Caveat: the sensitivity information is harder to obtain.