SAND2020-10035C

Uncertainty Quantification in Large Scale Computational Models

Habib N. Najm

Sandia National Laboratories
Livermore, CA, USA
hnnajm@sandia.gov

Data Assimilation and Uncertainty Quantification at Exascale workshop

University College London, UK
24-25 Sep, 2020

oratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia,
neywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract

Acknowledgement

Bert J. Debusschere, Khachik Sargsyan, Cosmin Safta, Mike Eldred
Gianluca Geraci, Tiernan Casey — Sandia National Labs

Xun Huan — Univ. Mich
Joe Oefelein — Georgia Tech
Guilhem Lacaze — SpaceX
Zachary Vane — Blue Origin
Youssef Marzouk — MIT
Roger Ghanem — U. South. California, Los Angeles, CA

This work was supported by:

- DOE Office of Basic Energy Sciences, Div. of Chem. Sci., Geosci., & Biosci.
- DOE Office of Advanced Scientific Computing Research (ASCR)
- DOE ASCR Scientific Discovery through Advanced Computing (SciDAC) program
- DOE ASCR Applied Mathematics program
- DOD Defense Advanced Research Projects Agency (DARPA)

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & 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-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Probabilistic Forward UQ — $y = f(x)$

Represent uncertain quantities using probability theory

Random sampling, Monte Carlo

- Generate random samples $\{x^i\}_{i=1}^N$ from the PDF of x, $p(x)$
- Bin the corresponding $\{y^i\}$ to construct $p(y)$
- Not feasible for computationally expensive $f(x)$
  - slow convergence of MC/QMC methods
    $\Rightarrow$ very large N required for reliable estimates

Build a cheap surrogate for $f(x)$, then use Monte Carlo/others

- Collocation – interpolants
- Regression – fitting
- Galerkin methods
  - Polynomial Chaos (PC) methods

Challenges with Surrogate Construction

- Choice of surrogate function is informed by structure of $f(x)$
  - Structure of $f(x)$ not known *a priori*
  - Discontinuities, say at some x^*, require particular care
    - Local versus global surrogates
  - Nonlinearities, shape ...
    - *e.g.* polynomials have trouble with sigmoid response
    - Surrogate complexity can grow
- High dimensionality in x
  - Large number of uncertain parameters
  - Non-smooth random fields
- Large computational cost for $f(x)$
  - *e.g.* a global climate simulation
  - Can only afford a few samples

Role of Surrogates in Probabilistic UQ

Computational forward model, parameter vector λ

$$y = f(x, \lambda)$$

Forward UQ

- Given PDF $p(\lambda)$, estimate $p(y)$ or $M_q(y) = \mathbb{E}[y^q]$
- General *non-intrusive* methods rely on sampling λ
- Require many samples $(\lambda_k, f(x, \lambda_k))$, $k = 1, \dots, N$

Inverse UQ

- Given data $D := \{(x_i, y_i),\ i = 1, \dots, M\}$, estimate $p(\lambda|D)$
- Bayesian methods often use Markov Chain Monte Carlo (MCMC)
- Require many samples $(\lambda_k, f(x_i, \lambda_k))$, $k = 1, \dots, K$, $\forall i$

Require a cheap surrogate $S_\alpha(x, \lambda) \simeq f(x, \lambda),\ \alpha \in \mathbb{R}^L$

High dimensionality is a major challenge in forward UQ

- High dimensionality is the result of
  - Large number of uncertain parameters/inputs
  - Large number of degrees of freedom in random field inputs
- Sparse-quadrature requires an unfeasible number of model evaluations for very high dimensional systems
- Monte Carlo requires similarly large number of samples when the number of important dimensions is very high
  - However, typically, physical model output quantities of interest are *smooth* $\Rightarrow$ Only a small number of inputs are important
- In this case, the way out is:
  - Use global sensitivity analysis (GSA) with Monte Carlo to identify important parameters
  - Use polynomial chaos expansions (PCE) with sparse quadrature on the reduced dimensional space for accurate forward UQ

Global sensitivity analysis: Sobol indices

Global sensitivity analysis (GSA) (Saltelli:2004,2008)

- For a given quantify of interest (QoI) ...
- QoI variance decomposed into contributions from each parameter
- Sobol indices rank parameters by their contributions (Sobol:2003)

Total effect $$S_{T_i} = \frac{\mathbb{E}_{\lambda_{\sim i}}[\mathrm{Var}_{\lambda_i}(f(\lambda)|\lambda_i)]}{\mathrm{Var}(f(\lambda))}$$

S_{T_i} small $\Rightarrow$ low impact parameter $\Rightarrow$ fix value (eliminate dimension)

How to compute?

- Monte Carlo estimators (Saltelli:2002,2010) still prohibitive if used directly for large scale computational models

Hi-dimension with large-scale computational models

When the number of feasible samples for GSA is highly limited due to computational costs:

- Reliable MC-estimation of sensitivity indices requires regularization
- Presuming smoothness, use MC samples to fit a PCE, which is subsequently used to estimate the sensitivity indices
- Employ ℓ_1-norm constrained regression to discover a sparse PCE
  - compressive sensing
- Employ Multilevel Monte Carlo (MLMC), as well as Multilevel Multifidelity (MLMF) methods
  - Optimal combination of coarse/fine mesh and low/high fidelity models to minimize computational costs for a given accuracy

Similarly for forward PC UQ:

- Employ generalized adaptive non-isotropic sparse quadrature with MLMF methods on reduced dimensional input space

Estimation of GSA Sobol' Indices with PC regression

- When # samples is small, GSA indices can be computed with improved accuracy, relying on PC regression/smoothing
- Polynomial Chaos expansion (PCE): $u(\boldsymbol{\xi}) = \sum_{\boldsymbol{\alpha}\in\mathfrak{I}} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$
  - Germ: $\boldsymbol{\xi} = \{\xi_1, \dots, \xi_d\}$, Multi-index $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_d\}$,
  - Polynomials, orthogonal w.r.t. $p(\boldsymbol{\xi})$, $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \prod_{i=1}^{d} \psi_{\alpha_i}(\xi_i)$
- Use regression with MC samples to fit a PCE to the data

$$\underset{c_{\boldsymbol{\alpha}}}{\text{argmin}} \sum_{s=1}^{N} \left(f(\lambda(\boldsymbol{\xi}^{(s)}))) - \sum_{\boldsymbol{\alpha}\in\mathfrak{I}} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}^{(s)}) \right)^2$$

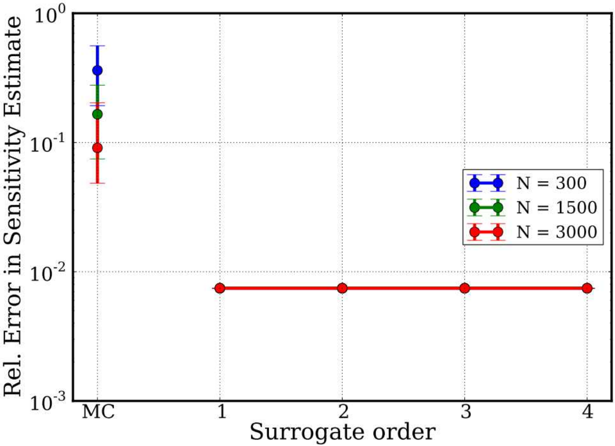
- Use PCE to evaluate Sobol indices

$$S_{T_i} = \frac{\sum_{\boldsymbol{\alpha}\in\mathfrak{I}|\alpha_i>0} c_{\boldsymbol{\alpha}}^2 \langle \Psi_{\boldsymbol{\alpha}}^2 \rangle}{\sum_{\boldsymbol{\alpha}\in\mathfrak{I}|\boldsymbol{\alpha}\neq\mathbf{0}} c_{\boldsymbol{\alpha}}^2 \langle \Psi_{\boldsymbol{\alpha}}^2 \rangle}$$

Sudret, 2008; Crestaux, 2009; Sargsyan, 2017; Ricciuto, 2018

Estimation of GSA Sobol' Indices with PC regression

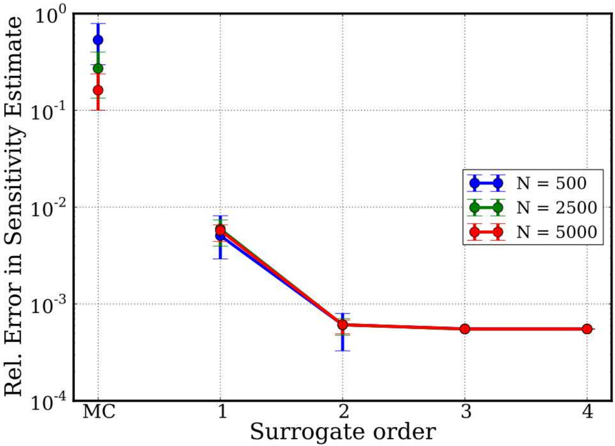
$d = 1$



Sargsyan, 2017

Estimation of GSA Sobol' Indices with PC regression

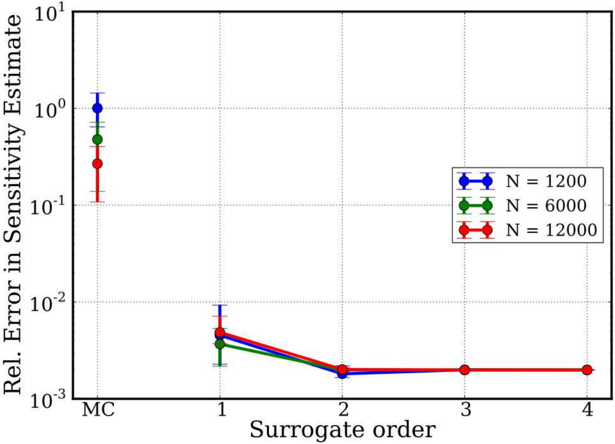
$d = 3$



Sargsyan, 2017

Estimation of GSA Sobol' Indices with PC regression

$d = 10$



Sargsyan, 2017

Sparse regression

Model: $$y = f(\boldsymbol{\xi}) \simeq \sum_{\boldsymbol{\alpha} \in \mathfrak{I}} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$$

- With N samples $(\boldsymbol{\xi}^1, y^1), \ldots, (\boldsymbol{\xi}^N, y^N)$, estimate K terms $c_{\boldsymbol{\alpha}}$

$$\min ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}||_2^2$$

With $N << K \Rightarrow$ under-determined, need regularization

- Use ℓ_1 norm regularization to discover sparsity
- Discover a sparse fitted PCE – many zero coefficients

Compressive Sensing; LASSO; basis pursuit; etc ...

$\min_{\boldsymbol{c}} \{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2\}$	subject to $\|\boldsymbol{c}\|_1 \leq \epsilon$	LASSO
$\min_{\boldsymbol{c}} \{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 + \lambda \|\boldsymbol{c}\|_1\}$		uLASSO
$\min_{\boldsymbol{c}} \{\|\boldsymbol{c}\|_1\}$	subject to $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{c}$	BP
$\min_{\boldsymbol{c}} \{\|\boldsymbol{c}\|_1\}$	subject to $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 \leq \epsilon$	BPDN

Unconstrained LASSO (uLASSO) – Practicalities

A broad range of methods exists for solving the optimization problem:

$$\boldsymbol{c}^* = \operatorname*{argmin}_{\boldsymbol{c}} \left\{ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 + \lambda \|\boldsymbol{c}\|_1 \right\}$$

l1_ls (Kim 2007), SpaRSA (Wright 2009), CGIST (Goldstein 2010), FPC_AS (Wen 2010), ADMM (Boyd 2010)

- Choice of $\lambda \geq 0$ controls the degree of overfitting vs underfitting
- This choice can be viewed as a model selection problem
    - Can base the choice on Bayesian model evidence maximization
- A cross-validation (CV) λ-choice strategy: minimize K-fold CV error

$$\lambda^* = \operatorname*{argmin}_{\lambda \geq 0} E_{\text{CV}}(\lambda)$$

- For expensive models, also target optimal data sample size
    - Increase sample size m adaptively
    - Stop sampling when the rate of decrease of λ-optimal CV error with increasing m drops below a given threshold Huan, SIAM JUQ 2018

Bayesian Regression

- Bayes formula

$$p(\boldsymbol{c}|D) \propto p(D|\boldsymbol{c})\pi(\boldsymbol{c})$$

- Bayesian regression: prior as a regularizer, *e.g.*
    - Log Likelihood $\Leftrightarrow$ $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2$
    - Log Prior $\Leftrightarrow$ $\|\boldsymbol{c}\|_p^p$
- Laplace sparsity priors $\pi(c_k|\alpha) = \frac{1}{2\alpha} e^{-|c_k|/\alpha}$
- uLASSO (Tibshirani 1996, Van den Berg 2008) ... formally:

$$\min \left\{ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{c}\|_2^2 + \lambda \|\boldsymbol{c}\|_1 \right\}$$

Solution $\sim$ the posterior mode of $\boldsymbol{c}$ in the Bayesian model

$$y \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{c}, I_N), \qquad c_k \sim \frac{1}{2\alpha} e^{-|c_k|/\alpha}$$

- Bayesian LASSO (Park & Casella 2008)

Bayesian Compressive Sensing (BCS)

- BCS (Ji 2008; Babacan 2010)— hierarchical priors:
  - Gaussian priors $\mathcal{N}(0, \sigma_k^2)$ on the c_k
  - Gamma priors on the σ_k^2

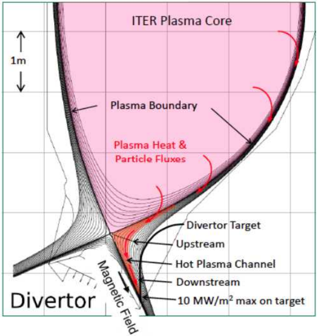
  $\Rightarrow$ Laplace sparsity priors on the c_k
- Evidence maximization establishes maximum likelihood estimates of the σ_k
  - many of which are found $\approx 0 \quad \Rightarrow \quad c_k \approx 0$
  - iteratively include terms that lead to the largest increase in the evidence
- Iterative BCS (iBCS) (Sargsyan 2012):
  - adaptive iterative order growth
  - BCS on order-p Legendre-Uniform PC
  - repeat with order-$p+1$ terms added to surviving p-th order terms

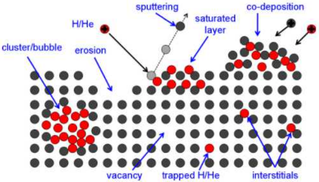
Demonstration in Cluster Dynamics Computations

- Material damage processes associated with plasma surface interactions in the ITER fusion reactor – He in W/Be
- Xe gas bubble transport in nuclear fuel rods in fission reactors
- "Xolotl" C++ cluster dynamics code for prediction of gas bubble evolution in solids
- Solves PDE (x, t) for concentration of clusters of different sizes
- 2D/3D - relies on PETSc solvers
- https://github.com/ORNL-Fusion/xolotl

Brian Wirth, Sophie Blondel – Oak Ridge National Lab

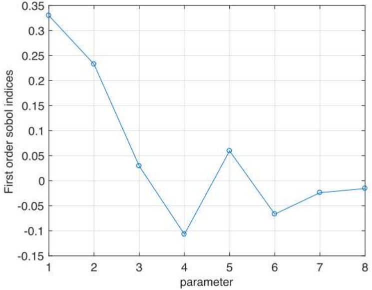
ITER Plasma Material Interface





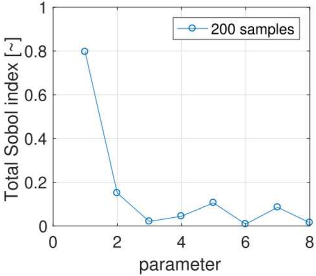
S. Blondel *et al.*, COSIRES, 2018

GSA in Xolotl

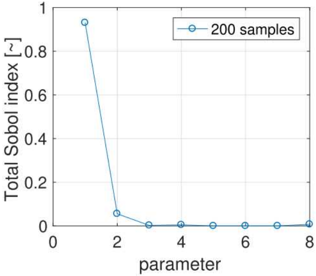


- 400 samples GSA MC study
- 8 parameters
- Noise/negativity in indices eliminated with sparse PC regression @ 200 samples

LSQ-PC: Total:



BCS-PC: Total:

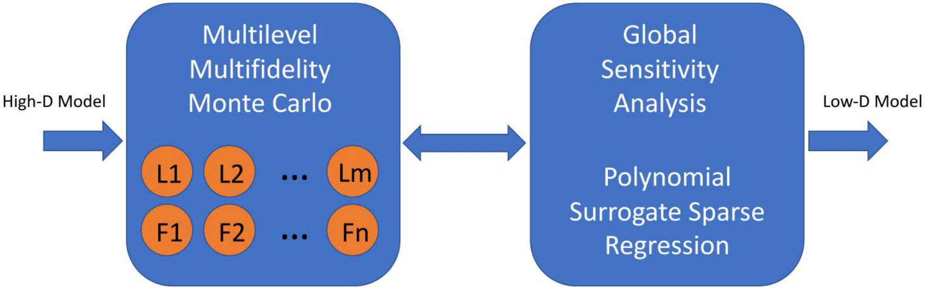


Multilevel Multifidelity (MLMF) Methods for UQ

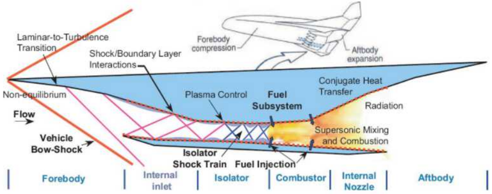
When the computational model is quite expensive, we still seek more reduction in the required number of expensive samples

- Multilevel Multifidelity (MLMF) methods allow further savings by combining information judiciously from low/high-resolution and how/high-fidelity models
- Use many low resolution/fidelity model computations and a minimal necessary number of high resolution/fidelity model computations to achieve target accuracy with MC
- Choice of how many simulations to run at low and high fidelity/resolution is done adaptively

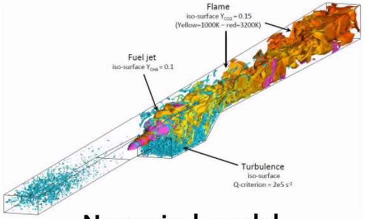
Forward UQ GSA Workflow



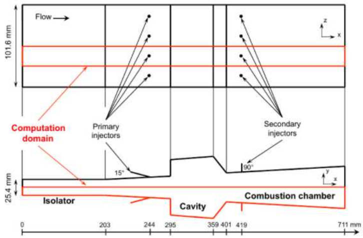
Supersonic Combusting **Ramjet** (scramjet)

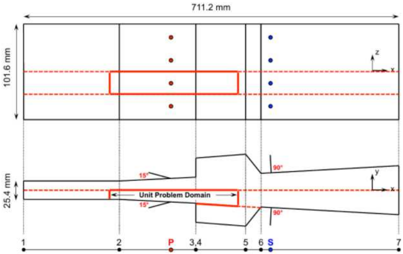


In flight



Numerical model

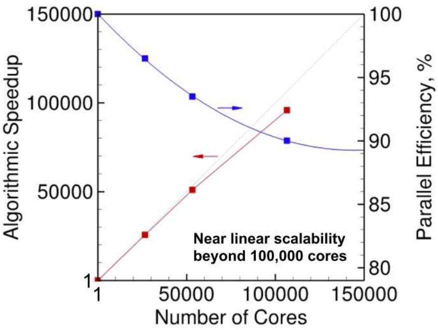




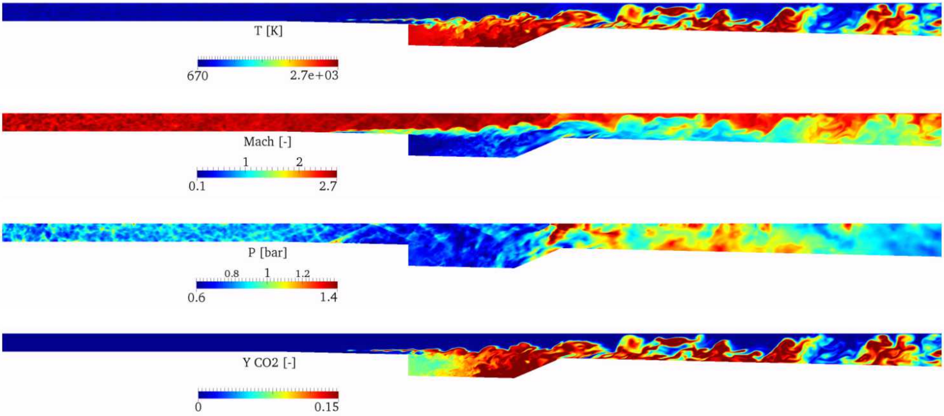
LES Performed using RAPTOR Code Framework

Joe Oefelein – Sandia National Labs. – now at Georgia Tech

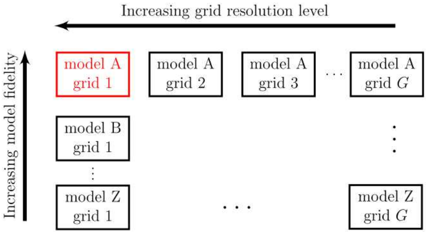
- **Theoretical framework … (Comprehensive physics)**
  - **Fully-coupled, compressible conservation equations**
  - **Real-fluid equation of state (high-pressure phenomena)**
  - **Detailed thermodynamics, transport and chemistry**
  - **Multiphase flow, spray**
  - **Dynamic SGS modeling (No Tuned Constants)**
- **Numerical framework … (High-quality numerics)**
  - **Staggered finite-volume differencing (non-dissipative, discretely conservative)**
  - **Dual-time stepping with generalized preconditioning (all-Mach-number formulation)**
  - **Detailed treatment of geometry, wall phenomena, transient BC's**
- **Massively-parallel … (Highly-scalable)**
  - **Demonstrated performance on full hierarchy of HPC platforms (e.g., scaling on ORNL CRAY XK7 TITAN architecture shown below)**
  - **Selected for early science campaign on next generation SUMMIT platform (ORNL Center for Accelerated Application Readiness, 2015 – 2018)**



Instantaneous Flow Structure — z-inj-cut – 3D d16



Multilevel and multifdelity forms



Telescopic sum:

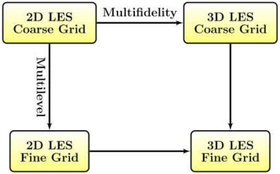
$$f_L(\lambda) = f_0(\lambda) + \sum_{\ell=1}^{L} f_{\Delta_\ell}(\lambda)$$

- ℓ indicates different grid levels or fidelity of models
- Δ_ℓ indicates difference between models ℓ and $\ell - 1$

Function approximation:

$$f_L(\lambda) \approx \hat{f}_L(\lambda) = \hat{f}_0(\lambda) + \sum_{\ell=1}^{L} \hat{f}_{\Delta_\ell}(\lambda)$$

High-D – ML/MF UQ Results



Two model forms and two mesh discretization levels

- Model form: 2D (LF) and 3D (HF) LES
- Meshes: $d/8$ and $d/16$

The jet-in-crossflow problem (24 inputs):
Five QoIs extracted over a plane at $x/d = 100$.

- $\mathbb{E}_{y,t}$ stagnation pressure ($P_{0,mean}$)
- $\mathbb{E}_y$ RMS$_t$ stagnation pressure ($P_{0,rms}$)
- $\mathbb{E}_{y,t}$ Mach number (M_{mean})
- $\mathbb{E}_{y,t}$ turbulent kinetic energy (TKE$_{mean}$)
- $\mathbb{E}_{y,t}$ scalar dissipation rate (χ_{mean})

	2D	3D
$d/8$	1	204
$d/16$	25.5	1844

Relative computational cost for the model forms and discretization levels.

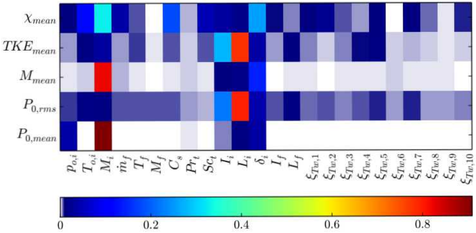
Optimize statistical accuracy given a limited number of high fidelity model evaluations by leveraging cheaper lower fidelity simulations.

Jet in crossflow problem: 24 parameters, 3^{rd}-order PCE

Parameter	Range	Description
	Inlet boundary conditions	
p_0	[1.406, 1.554] MPa	Stagnation pressure
T_0	[1472.5, 1627.5] K	Stagnation temperature
M_0	[2.259, 2.761]	Mach number
δ_a	[2, 6] mm	Boundary layer thickness
I_i	[0, 0.05]	Turbulence intensity magnitude
L_i	[0, 8] mm	Turbulence length scale
	Fuel inflow boundary conditions	
$\dot{m}_f$	[6.633, 8.107] $\times 10^{-3}$ kg/s	Mass flux
T_f	[285, 315] K	Static temperature
M_f	[0.95, 1.05]	Mach number
I_f	[0, 0.05]	Turbulence intensity magnitude
L_f	[0, 1] mm	Turbulence length scale
	Turbulence model parameters	
C_R	[0.01, 0.06]	Modified Smagorinsky constant
Pr_t	[0.5, 1.7]	Turbulent Prandtl number
Sc_t	[0.5, 1.7]	Turbulent Schmidt number
	Wall boundary conditions	
T_w	Expansion in 10 params of $\mathcal{N}(0,1)$	Wall temperature represented via Karhunen-Loève expansion

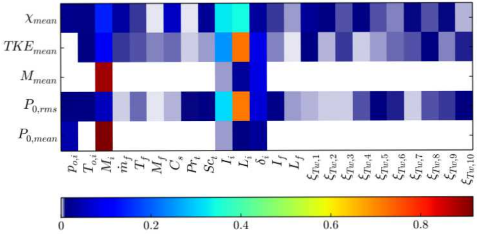
- QoIs computed at $x/d = 100$, averaged over (y, t)
- 2D runs: 1939 (coarse grid), 79 (fine grid)
- 3D runs: 46 (coarse grid), 11 (fine grid)

Unit problem: total sensitivity



Multilevel expansion of:

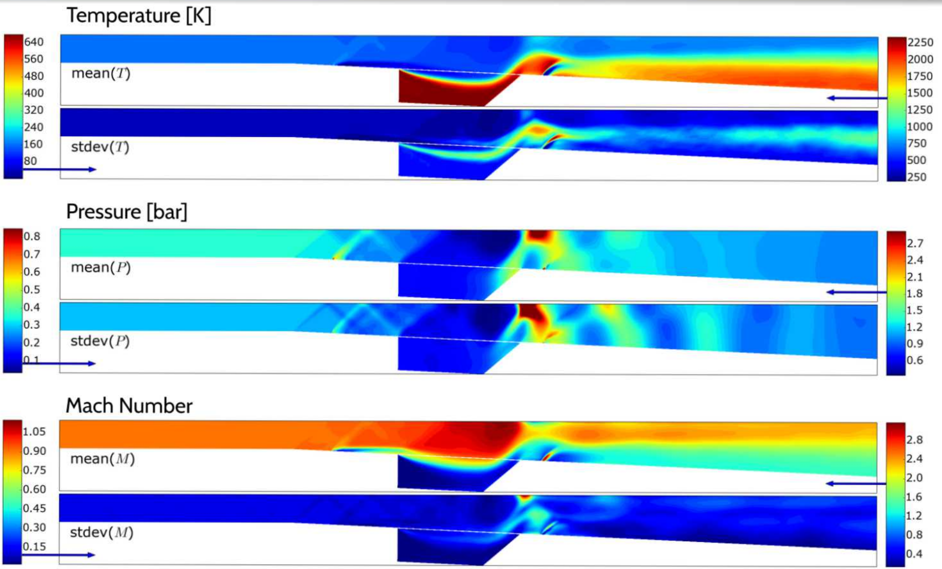
$$\widehat{f}_{2D,d/16} = \widehat{f}_{2D,d/8} + \widehat{f}_{\Delta_{2D,d/16-2D,d/8}}$$



Multifidelity expansion of:

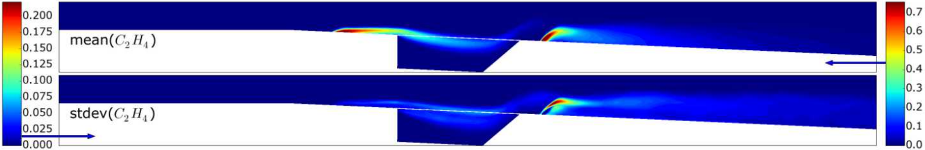
$$\widehat{f}_{3D,d/8} = \widehat{f}_{2D,d/8} + \widehat{f}_{\Delta_{3D,d/8-2D,d/8}}$$

MC-Predicted Uncertainty in Mean Flow Quantities – 3D

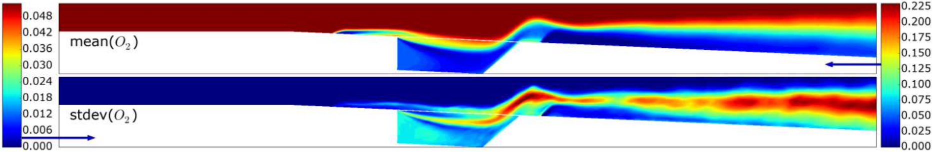


MC-Predicted Uncertainty in Mean Flow Quantities – 3D

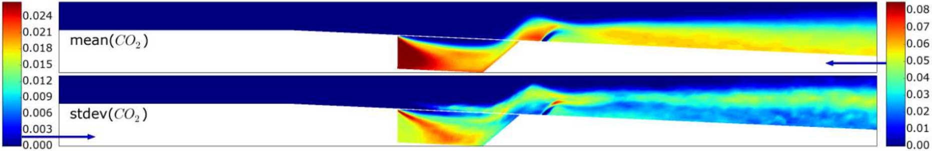
Ethane



Oxygen



Carbon Dioxide



Discussion and Closure

- Necessary workflow for UQ in large-scale computational models
  - Global sensitivity analysis to cut dimensionality, assisted by
    - Polynomial Chaos regression
    - ℓ_1-norm regularization / compressive sensing
    - Multilevel Monte Carlo & Multifidelity
  - Adaptive sparse quadrature forward UQ on reduced dimensional space
  - Resulting PC surrogate can be used in Bayesian inference on model parameters and optimization under uncertainty
- Other avenues to re-cast the problem in low-D:
  - Basis adaptation & active subspace methods
  - Manifold discovery, e.g. via Isomap or diffusion maps
  - Low rank tensor methods, etc
- Caution: Noisy computational QoIs due to finite averaging windows