Multifidelity forward-inverse workflow in Dakota. A demonstration in Wind Energy

D. Thomas Seidl[1], Gianluca Geraci[1], Friedrich Menhorn[2], Michael S. Eldred[1] and Ryan King[3]

[1]Sandia National Laboratories, Albuquerque NM
[2]Technical University of Munich, Garching Germany
[3]National Renewable Energy Laboratory, Golden CO

January 2021

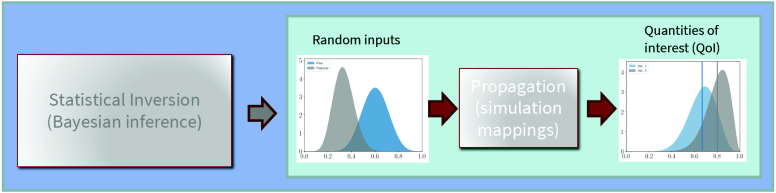
Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

Plan of the talk

- Uncertainty Quantification
- MF in UQ: motivation
- Multifidelity surrogate UQ
- Surrogate-based inverse UQ
- Inverse UQ demonstration for Wind
- Conclusions

Uncertainty Quantification

Uncertainty Quantification
Generalities



Uncertainty Quantification:

- **Forward UQ:** **random inputs are propagated** through the computer code and statistics of the QoI are evaluated
- **Inverse UQ:** measurements of the QoIs are used to **obtain data-informed distributions of the input** parameters

Forward and Inverse UQ are part of the same workflow:

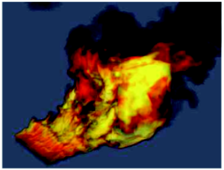
- Prior **distributions based on a priori knowledge**
- From observational data (experiments, etc.) **we can infer posterior distributions via Bayes rule**
- Use of **data can reduce uncertainty** in parameter to QoI mapping (priors are constrained)
- Design using **prior uncertainties can be overly conservative**
- Reduced uncertainty of **data-informed UQ can produce designs with greater performance**

Why multifidelity in Uncertainty Quantification?

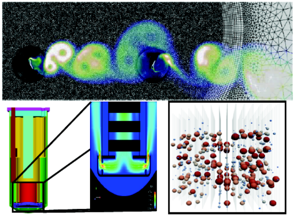
Uncertainty Quantification

DoE and DoD deployment activities

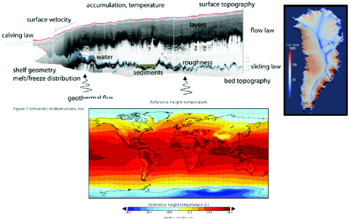
Stewardship (NNSA ASC)
Safety in abnormal environments



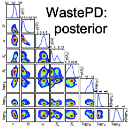
Energy (ASCR, EERE, NE)
Wind turbines, nuclear reactors

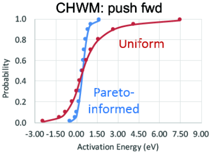


Climate (SciDAC, CSSEF, ACME)
Ice sheets, CISM, CESM, ISSM, CSDMS



Addtnl. Office of Science:
(SciDAC, EFRC)
Comp. Matls: waste forms / hazardous matls (WastePD, CHWM)
MHD: Tokamak disruption (TDS)





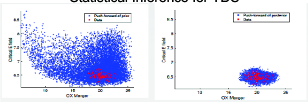


Figure: Courtesy of Mike Eldred

High-fidelity state-of-the-art modeling and simulations with HPC

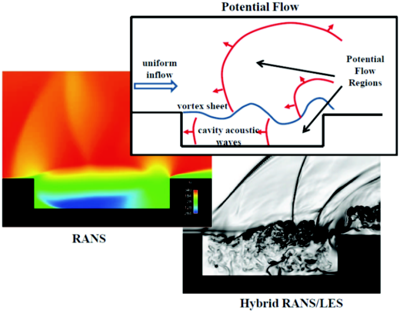
- **Severe** simulations **budget constraints**
- **Significant dimensionality** driven by model complexity

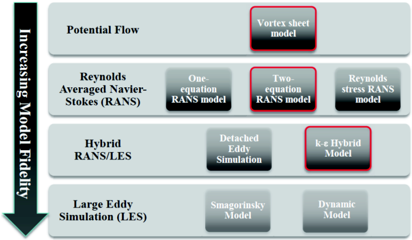
Uncertainty Quantification

Rich set of modeling choices – Discretization Vs Fidelity

Multi-fidelity: several accuracy levels available

- Physical models (Laminar/Turbulent, Reacting/non-reacting, viscous/inviscid...)
- Numerical methods (high/low order, Euler/RANS/LES, *etc*...)
- Numerical discretization (fine/coarse mesh...)
- Quality of statistics (long/short time history for turbulent flow...)





Relationships amongst models can be difficult to anticipate

- **Hierarchical** relationships usually correspond to modeling choices like, *e.g.* discretization
- However, **peer** relationships are often observed in the presence of physical approximations

MF Uncertainty Quantification

Sampling, Surrogate based methods and challenges

MF Sampling methods

- Derive directly from **Monte Carlo**
- Exploit **correlation** among model outputs
- Are built to obtain an **estimator variance reduction**
- Ex: MLMC (Giles, 2015), MFMC (Peherstorfer *et al.*, 2016), MLMF (Geraci *et al.*, 2015) ACV (Gorodetsky *et al.*, 2020)

MF Surrogate methods

- Provide an approximation of the *input-output* mapping
- Achieve **rapid error decrease** (as the amount of data increases), provided that the *input-output* mapping is smooth
- Can be **easily integrated** in inverse and OUU workflows
- Ex: Co-Kriging (Gradiet and Garnier, 2014) and stage-fitting (Liu *et al.* 2018)

Surrogate-based Multifidelity

Seminal idea

Decreasing 'complexity' for the discrepancy function

- The concept of **multifidelity** has been known/exploited in the optimization community for decades
- One of the first applications of this concept in UQ:

  Ng and **Eldred**. *Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation.* In 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, 2012.

The **main idea** is quite simple and effective: Can you use a **LF model to capture most of the response** and use only **fewer HF evaluations to correct** it?

An illustrative demonstration:

$$Q_{HF} = \exp -0.05\xi^2 \cos 0.5\xi \boldsymbol{-0.5} \exp \boldsymbol{-0.02(\xi - 5)^2}$$

$$Q_{LF} = \exp -0.05\xi^2 \cos 0.5\xi$$

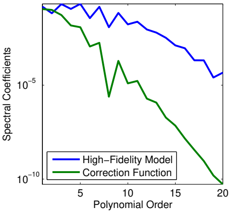


FIGURE: Spectral content

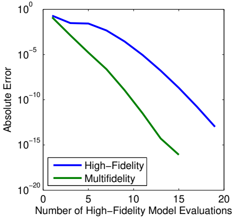


FIGURE: Error (Mean)

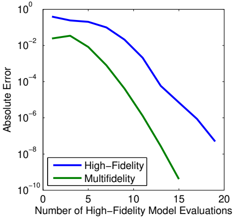


FIGURE: Error (Mean)

'Complexity' of a function

Order, sparsity, low-rank structure...

The original idea was based on the following assumptions:

- the LF model is able to capture the high frequencies of the response
- only the low-order terms are included in the discrepancy term → **few evaluations of the discrepancy are needed** to build the response for the discrepancy

In many **practical applications**:

- the **LF model only captures low-order effects**
- however the discrepancy term can have a structure that we can still exploit

Two possible structures that we can exploit are:

- **Sparsity** → Compressed sensing: orthogonal matching pursuit (OMP), basis pursuit denoising (BPDN), least angle regression (LARS), least absolute selection and shrinkage operator (LASSO)...
- **Low-rank** → Functional Tensor-Train decomposition (TT)

Exploiting favorable function's structures
Three main strategies – Hierarchical relationship among models

In order we have tried **several approaches**:

1. **Optimal resources allocation** (direct extension of MLMC concepts to surrogates)

Exploiting favorable function's structures

Three main strategies – Hierarchical relationship among models

In order we have tried **several approaches**:

1. **Optimal resources allocation** (direct extension of MLMC concepts to surrogates)
2. Exploiting **Restricted Isometry Property** (RIP)

Exploiting favorable function's structures

Three main strategies – Hierarchical relationship among models

In order we have tried **several approaches**:

1. **Optimal resources allocation** (direct extension of MLMC concepts to surrogates)
2. Exploiting **Restricted Isometry Property** (RIP)
3. **Greedy Multilevel Refinement**

Exploiting favorable function's structures

Strategy 1: Extending the MLMC sampling approach to surrogates

Main idea: Parametrize the variance of the recovered discrepancy term

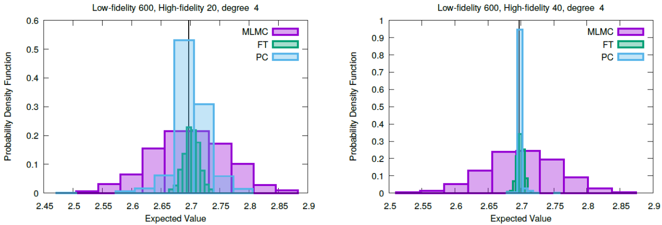
$$\mathbb{V}ar\left[\hat{Y}_\ell\right] = \frac{\mathbb{V}ar\left[Y_\ell\right]}{\gamma N^{\mathbf{k}}} \rightarrow N_\ell = \sqrt[k]{\frac{\sum_{q=0}^{L} \sqrt[k+1]{\mathbb{V}ar\left[Y_q\right] \mathcal{C}_q^k}}{\gamma \varepsilon^2/2}} \sqrt[k+1]{\frac{\mathbb{V}ar\left[Y_\ell\right]}{\mathcal{C}_\ell}}$$

Notes:

- γ and k can be obtained as by-product of the k-fold cross-validation process
- this approach can be extended to level-dependent parameters, *i.e.* γ_ℓ and k_ℓ (slightly different closed form solution)

Findings:

- **Abrupt transition** in both sparse and low-rank recovery **does not allow to efficiently estimate the parameters** and exploit the faster convergence



(a) $N_{low} = 600$, $N_{high} = 20$ and $deg = 4$

(b) $N_{low} = 600$, $N_{high} = 40$ and $deg = 4$

Exploiting favorable function's structures

Strategy 2: Restricted Isometry Property (RIP) from *Jakeman, Narayan, Zhou, 2016*

Main idea: Address/Avoid abrupt transition by **ensuring enough samples** for accurate recovery

$$\text{RIP}: \quad N_\ell \geq s_\ell L_\ell \log^3 (s_\ell) \log (C_\ell)$$

where

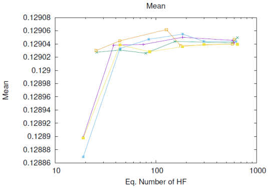
- s_ℓ is the sparsity, *i.e.* number of non-zero coefficients
- L_ℓ is the mutual coherence, *i.e.* if a_i are the normalized $(a_i^T a_i = 1)$ columns of the matrix A then $L = \max \left| a_i^T a_j \right|$ for $i \neq j$
- C_ℓ is the cardinality of the dictionary

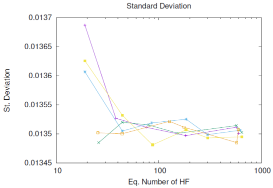
Algorithm:

- Start with pilot sample to estimate sparsity at each level ℓ
- Number of samples is increased to allow the recovery

Findings:

- **RIP is quite conservative** and it is likely to overshoot so it is necessary to add a constraint on the profile $\rightarrow$ very difficult to handle the feedback





Exploiting favorable function's structures
Strategy 3: Greedy Multilevel refinement

Main issues discovered with strategy #1 and #2 are:

- **Difficult to estimate** a trend
- **Difficult to handle** the sample allocation in order to avoid overshooting

Proposed solution: Greedy refinement - compete refinement candidates to **maximize induced change per unit cost**

Algorithm:

- One or more **candidates are generated** per each level
- The **impact of each candidate on the final QoIs statistics is evaluated** and normalized by the relative cost of level increment
- **Greedy selection** of the best candidate
- **Generation of new candidates** for the selected level

Greedy Multilevel refinement

Level candidate generators – Current options in Dakota

- **Uniform refinement:** coarse-grained refinement with one expansion order / grid level candidate per model level
  - Tensor / sparse grids: projection PCE and nodal/hierarchical SC
  - Regression PCE: least squares / compressed sensing using a fixed sample ratio
- **Anisotropic refinement:** coarse-grained refinement with one expansion order / grid level candidate per model level
  - Tensor / sparse grids: projection PCE and nodal/hierarchical SC
- **Index-set-based refinement:** fine-grained refinement with multiple index set candidates per model level; exponential growth in size of candidate set with dimension.
  - Generalized sparse grids: projection PCE and nodal/hierarchical SC
- **Basis selection:** coarse-grained refinement with a few expansion order frontier advancements per model level
  - Regression PCE

Greedy Multilevel refinement
CS/GSG – Statistics

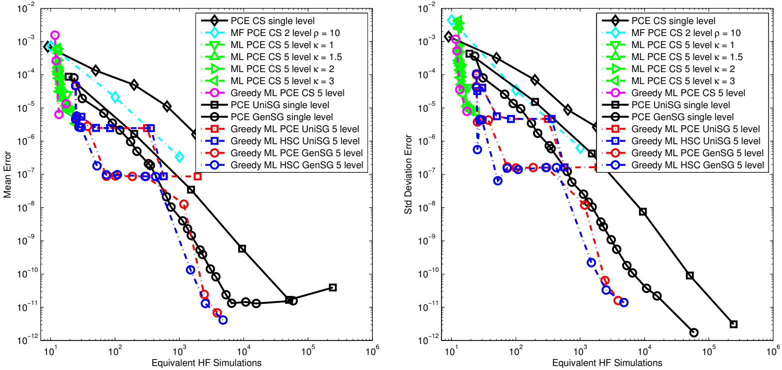


FIGURE: Convergence for greedy multilevel PCE comparing generalized sparse grids and compressed sensing.

Notes:

- The explicit nature of the sparse grid approaches allows for more precise convergence
- The CS-based approaches are hampered in accuracy by the large implicit system solve needed at the coarse level

Surrogate-based inverse UQ

Bayesian inversion

Generalities on the approach adopted in this work

Bayesian calibration

- Sandia's UQ software **Dakota** (see Dakota Theory Manual for more details)
- **Markov Chain Monte Carlo** for computing a sample-based **posterior distribution**
- We are interested in calibrating the parameters $\boldsymbol{\theta}$
- We assume that a surrogate for the computational model is available for the QoI: $\mathbf{q} = \mathbf{q}(\boldsymbol{\theta})$
- Reference data $\mathbf{d}$ are available

NOTES:

- From computational perspective it is more convenient to work with the negative log-likelihood

$$-\log \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{d}) = \frac{n}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{d}}| + \frac{1}{2}\boldsymbol{r}^T \boldsymbol{\Sigma}_{\boldsymbol{d}}^{-1}\boldsymbol{r}$$

- The term $\boldsymbol{r}^T \boldsymbol{\Sigma}_{\boldsymbol{d}}^{-1}\boldsymbol{r}$ is called **Misfit Function**
- Minimizing the **Misfit Function** corresponds to maximizing the **Likelihood**
- Maximizing the Likelihood (MLE) does not in general correspond to the Maximum A posteriori (MAP) point
- Posterior probability is analytically intractable and therefore MCMC is used to approximate it
- We use the QUESO library in Dakota to perform MCMC

Bayesian inversion

Why does having a surrogate help?

- **MCMC requires a very large number of runs/evaluations** to converge
- Surrogates can provide:
  - Computing **local accurate proposal density** (by using Hessian information)
  - **Pre-solving for the MAP** in order to eliminate the initial burn-in phase

Computing a local accurate proposal density

- The MCMC proposal covariance to be the inverse of the Hessian of the negative log posterior

$$\nabla^2_{\boldsymbol{\theta}} \left[-\log(\pi_{\text{post}}(\boldsymbol{\theta})) \right] = \nabla^2_{\boldsymbol{\theta}} M(\boldsymbol{\theta}) - \nabla^2_{\boldsymbol{\theta}} \left[\log(\pi_0(\boldsymbol{\theta})) \right]$$

- A standard approximation is the multivariate normal (MVN) distribution with mean centered at the actual point in the chain and prescribed covariance

$$-\nabla^2_{\boldsymbol{\theta}} \left[\log(\pi_0(\boldsymbol{\theta})) \right] = \boldsymbol{\Sigma}_{\mathbf{0}}^{-1} \quad \rightarrow \quad \nabla^2_{\boldsymbol{\theta}} \left[-\log(\pi_{\text{post}}(\boldsymbol{\theta})) \right] = \nabla^2_{\boldsymbol{\theta}} M(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_{\mathbf{0}}^{-1} \left[\log(\pi_0(\boldsymbol{\theta})) \right]$$

- The Heassian of the Misfit Function can be computed through the surrogate model as

$$\nabla^2_{\boldsymbol{\theta}} M(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \boldsymbol{q}(\boldsymbol{\theta})^T \boldsymbol{\Sigma}_{\boldsymbol{d}}^{-1} \nabla_{\boldsymbol{\theta}} \boldsymbol{q}(\boldsymbol{\theta}) + \nabla^2_{\boldsymbol{\theta}} \boldsymbol{q}(\boldsymbol{\theta}) \cdot \left[\boldsymbol{\Sigma}_{\boldsymbol{d}}^{-1} \boldsymbol{r} \right].$$

Avoiding the burn-in phase

- **When a surrogate is available** the burn-in can be avoided by pre-solving for the MAP point using an optimizer to minimize the negative log posterior

$$\boldsymbol{\theta}_{MAP} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left[-\log(\pi_{\text{post}}(\boldsymbol{\theta})) \right]$$

Inverse UQ demonstration
Nalu/WindSE

Bayesian inversion

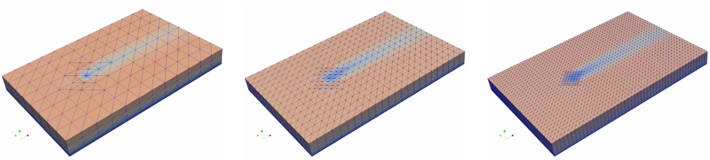
Wake characterization for a V27 rotor – Multilevel Problem setup

Reference data from Nalu

- 100 m $\times$110 m slice 5D downstream (135m)
- Data acquired each second for 10 minutes
- Reference data are averaged

Computational tool: WindSE

- Medium fidelity tool for 3D Reynolds-averaged Navier Stokes (RANS) simulations
- Turbines are represented by means of **non-rotating** actuator disks
- Turbulence closure via mixing length
- Based on FEniCS which enables easy user customization



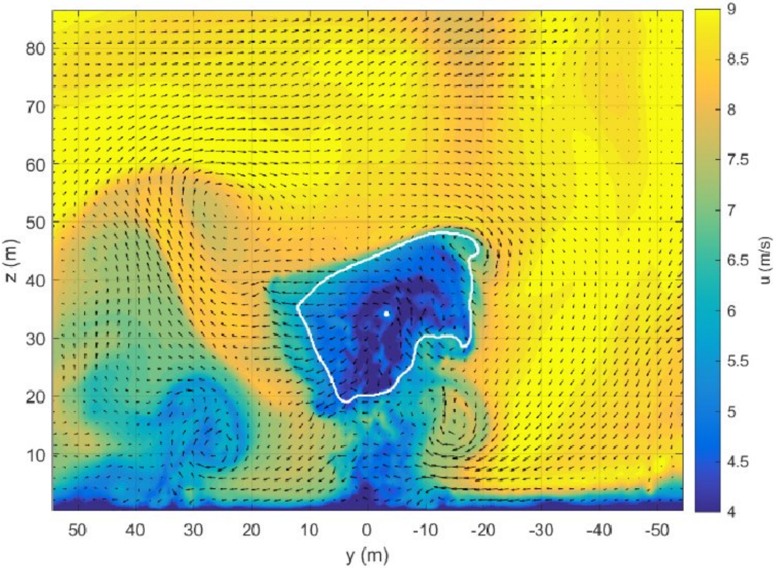
(a) Coarse: 5252 DoFs (b) Medium: 33428 DoFs (c) Fine: 228064 DoFs

Model Resolution	N_x	$N_y = N_z$	Cost (s)
Coarse	12	8	8.51
Medium	24	16	60.4
Fine	48	32	1270

Table: Multilevel model hierarchy unrefined grid discretization and simulation cost.

Bayesian inversion

Wake characterization for a V27 rotor – Reference HF (Nalu-Wind) snapshot



Bayesian inversion

Wake characterization for a V27 rotor – Nominal conditions

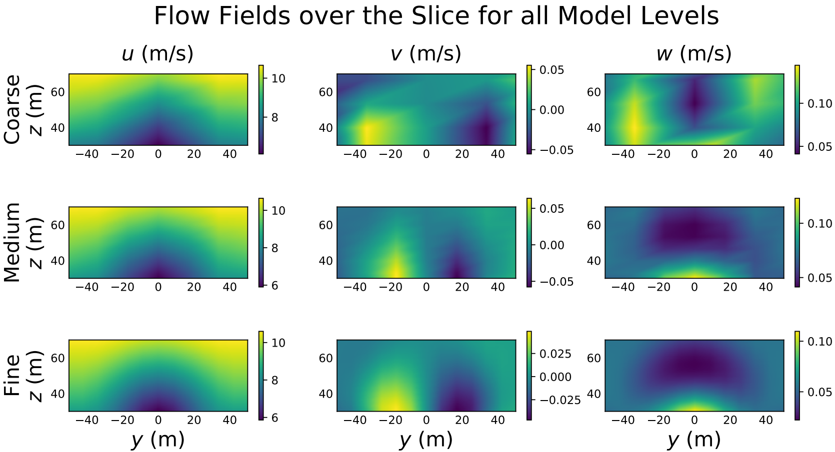


Figure: Nominal output for three velocity components u, v and w over all models.

Bayesian inversion

Wake characterization for a V27 rotor – Multilevel Surrogate Construction

Parameter	u_H ($\frac{m}{s}$)	α	θ_{wind} ($^\circ$)	Eff. Thickness (m)	Ax. Ind. Factor	ℓ_{max} (m)
Lower bound	8.25	0.02	-15	2.4	0.15	3.5
Upper bound	8.75	0.5	15	15	0.9	15

Table: Uniform parameter bounds for the forward and inverse UQ studies.

RANS data

- **First set of tests** demonstrated that the misfit between the **data was dominated by boundary layer data**
- We **truncated the spatial region of interest** to $30m < z < 70m$ (total of 131×161 points)
- The **total number of QoIs** to be considered is $31\,395$

Tolerance	5e-4		5e-5		5e-6	
PCE Type	SF	ML	SF	ML	SF	ML
Coarse Evaluations	N/A	129	N/A	409	N/A	1201
Medium Evaluations	N/A	53	N/A	137	N/A	601
Fine Evaluations	81	13	209	17	433	61
Equivalent Fine Evaluations	17		27		99	
ML Speedup	4.9		8.0		4.4	

Table: Number of model evaluations for SF (single high-fidelity) and ML (multilevel) PCEs for three tolerances. The construction of each ML PCE requires less than a quarter of the cost of the corresponding SF model.

NOTE: A cost saving factor $4 - 9$ can be obtained in this case!

Bayesian inversion

ML Forward propagation – A priori surrogate is not able to reproduce the data

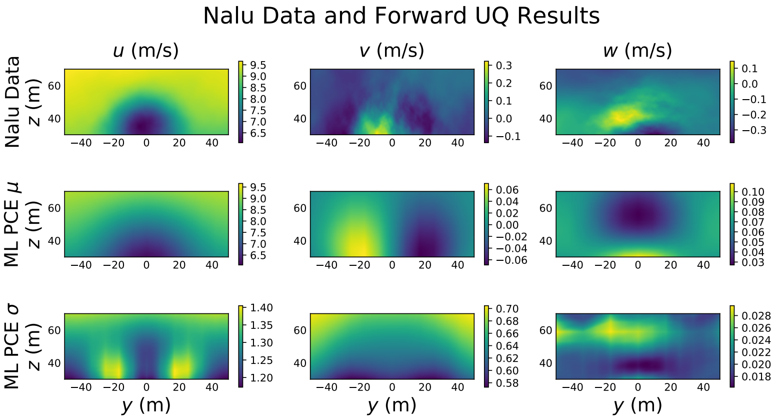
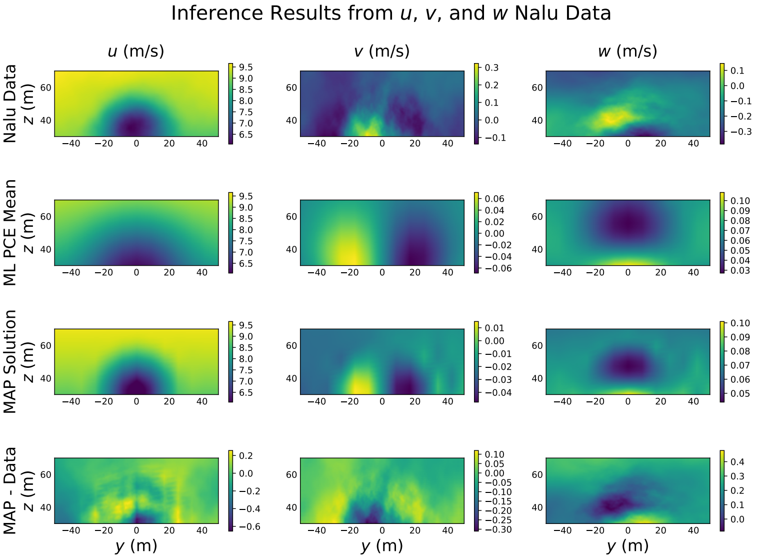


Figure: ML PCE built for all velocity components compared with the time-averaged Nalu slice data. The mean u component resembles the Nalu data but the other components do not due to the model error between WindSE and Nalu.

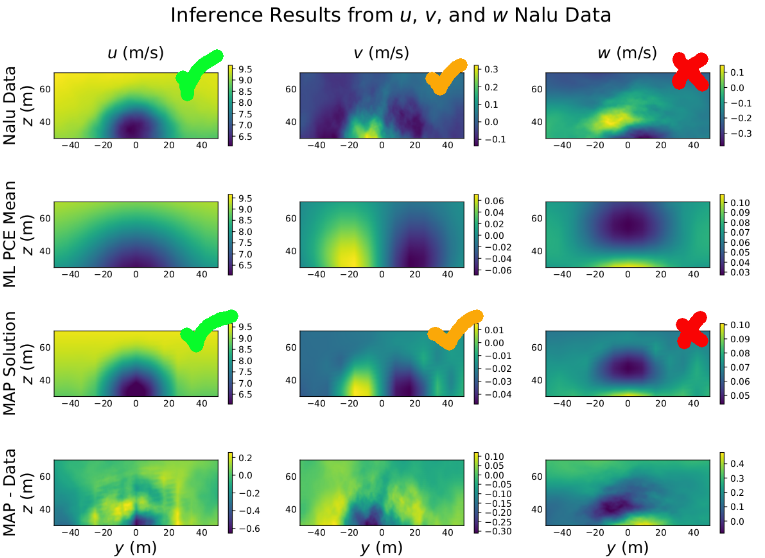
Bayesian inversion

Wake characterization for a V27 rotor – Multilevel Surrogate after calibration



Bayesian inversion

Wake characterization for a V27 rotor – Multilevel Surrogate after calibration



Bayesian inversion
Posterior distribution

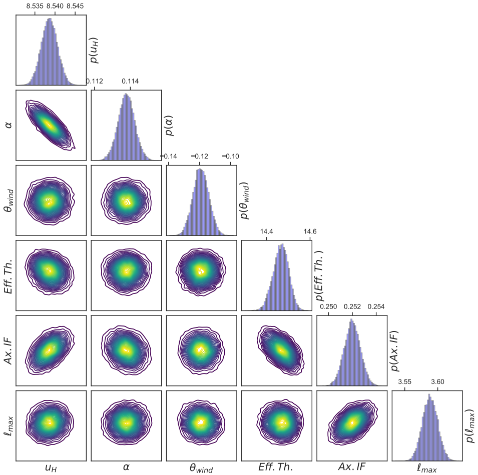


FIGURE: Visualization of the six-dimensional posterior distribution obtained through emulator-based inference from *all* velocity components. Marginal distributions are shown as histograms and pairwise joint distributions are displayed as contour plots.

Bayesian inversion
Posterior distribution

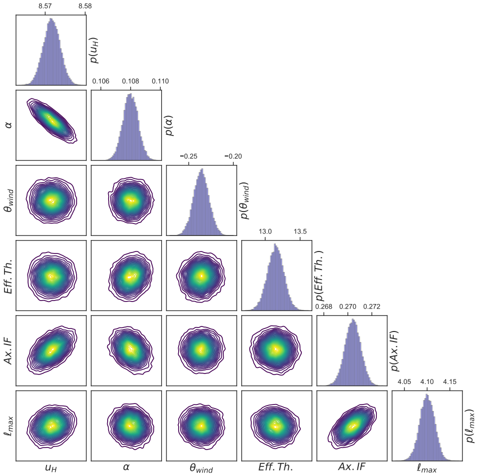


FIGURE: Visualization of the six-dimensional posterior distribution obtained through emulator-based inference from u data only. Marginal distributions are shown as histograms and pairwise joint distributions are displayed as contour plots.

Conclusions

Concluding Remarks
Still an active research area

Summary:

- Well designed Multifidelity estimators can **accelerate the construction of surrogates**
- Multifidelity surrogates can **accelerate the inverse UQ workflow**
- NOTE: Multifidelity surrogate-based **inversion cannot compensate for missing physics** – Non-rotating Actuator Disk in RANS Vs LES

(Incomplete) list of references:

- G. Geraci, M.S. Eldred, A.A. Gorodetsky & J.D. Jakeman, Recent advancements in Multilevel-Multifidelity techniques for forward UQ in the DARPA Sequoia project. *AIAA Scitech 2019 Forum*
- D.T. Seidl, G. Geraci, R. King, F. Menhorn, A. Glaws and M.S. Eldred, Multifidelity strategies for forward and inverse uncertainty quantification of wind energy applications. *AIAA Scitech 2020 Forum*
- A.A. Gorodetsky, J.D. Jakeman, G. Geraci, M. Eldred, MFNets: Multi-fidelity data-driven networks for Bayesian learning and prediction. *International Journal for Uncertainty Quantification*, Vol. 10(6), pp. 595–622, 2020.
- A.A. Gorodetsky, J.D. Jakeman, G. Geraci, MFNets: Learning network representations for multifidelity surrogate modeling *Journal of Computational Physics*, Under reviw, 2020. `https://arxiv.org/pdf/2008.02672.pdf`

THANKS!

Acknowledgements

- Alan Hsieh, David Maniaci, Thomas Herges, Myra Blaylock and Brent Houchens from Sandia National Laboratories for the insightful discussions and for providing the Nalu-Wind data
- Cosmin Safta (SNL) for providing the plotting tools for the posterior probability
- DOE EERE through the A2e program

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA-0003525.