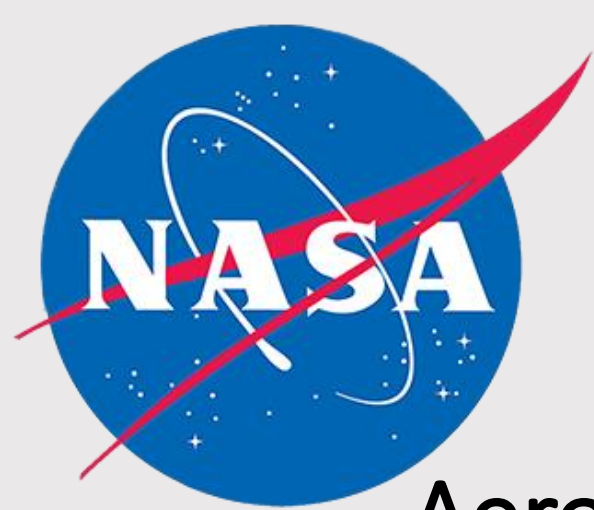




AeroFusion: Data Fusion and Uncertainty Quantification for Entry Vehicles

Steven Snyder

NASA Langley Research Center

NASA AeroFusion-MLUQ Early Career Initiative

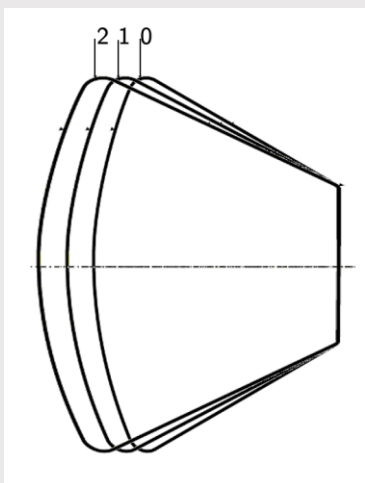
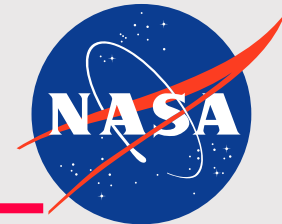
CME NASA Symposium

2023 August 14



*This material is a work of the U.S. Government
and is not subject to copyright protection in the United States.*

Aerodynamic Modeling Process



Define Flight Conditions

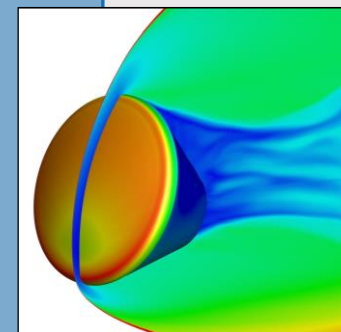
flight window
vehicle outer mold line
quantities of interest

minimize complexity

Generate Aerodynamic Data

computational fluid dynamics
wind tunnel testing
ROMs, machine learning

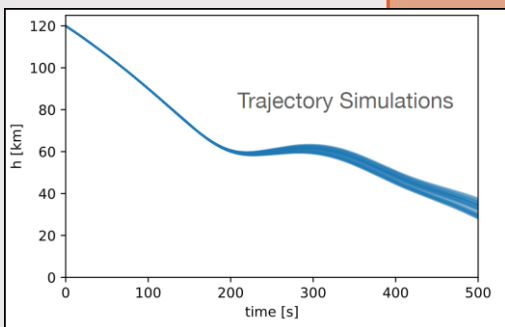
maximize informativity



Use Database

trajectory simulation
loads analysis
systems engineering

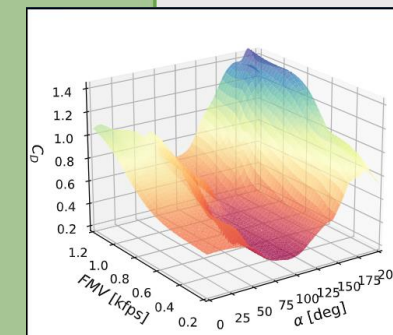
maximize relevance



Organize Database

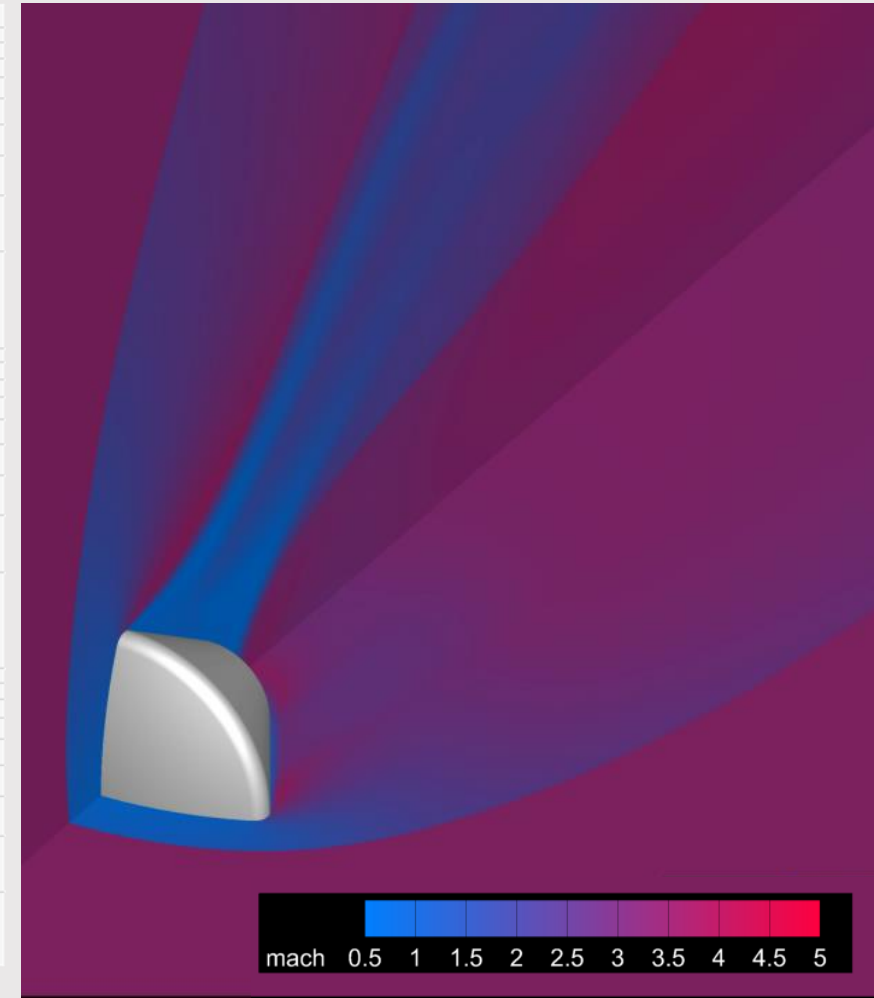
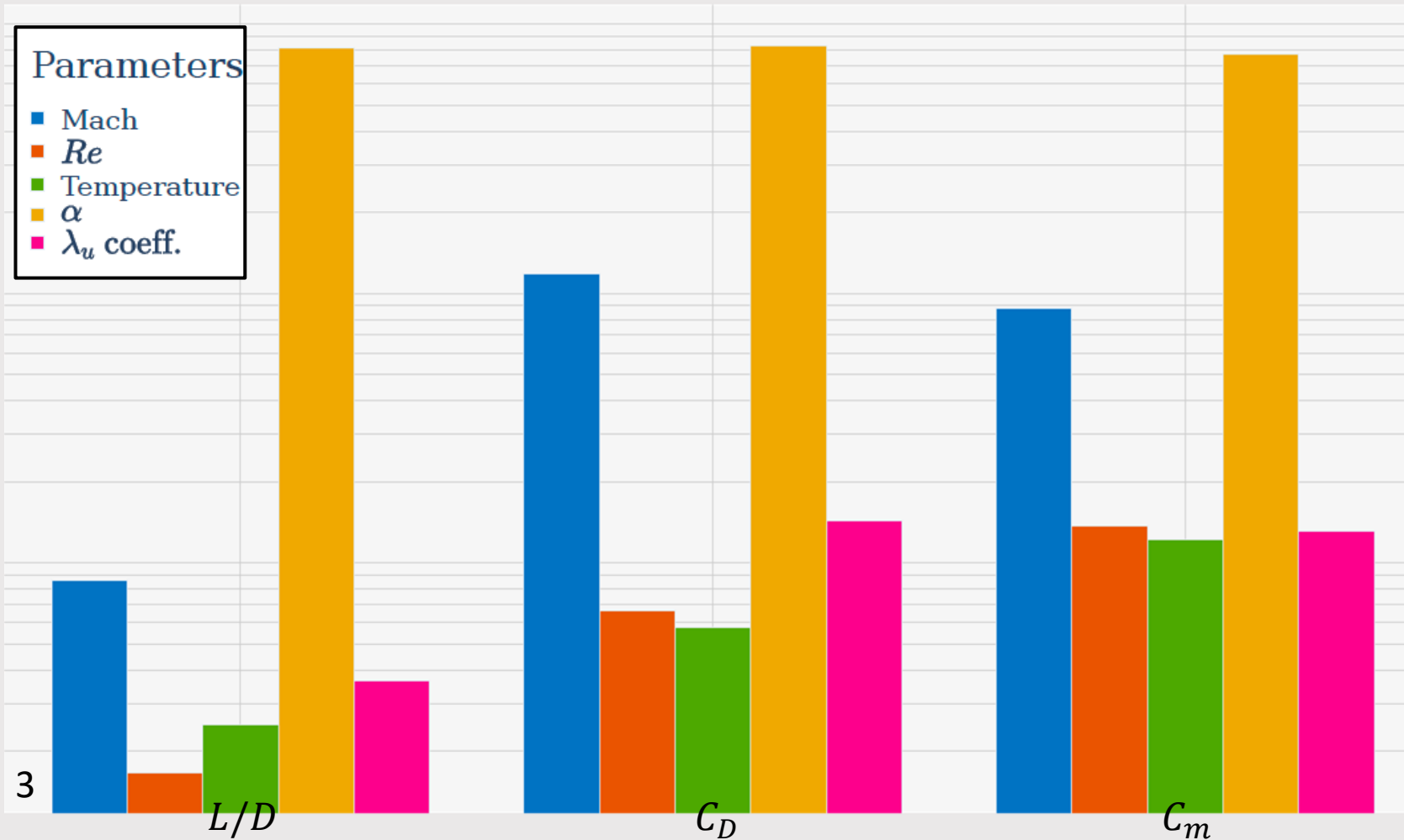
data fusion
surrogate modeling
delivery format

maximize coverage

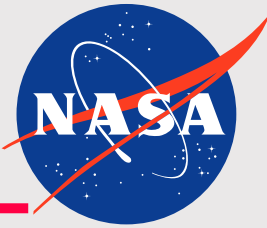


Sensitivity Analysis

How do we define our parameter space?

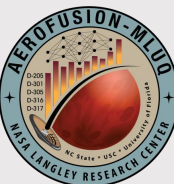


The Cost of More Data



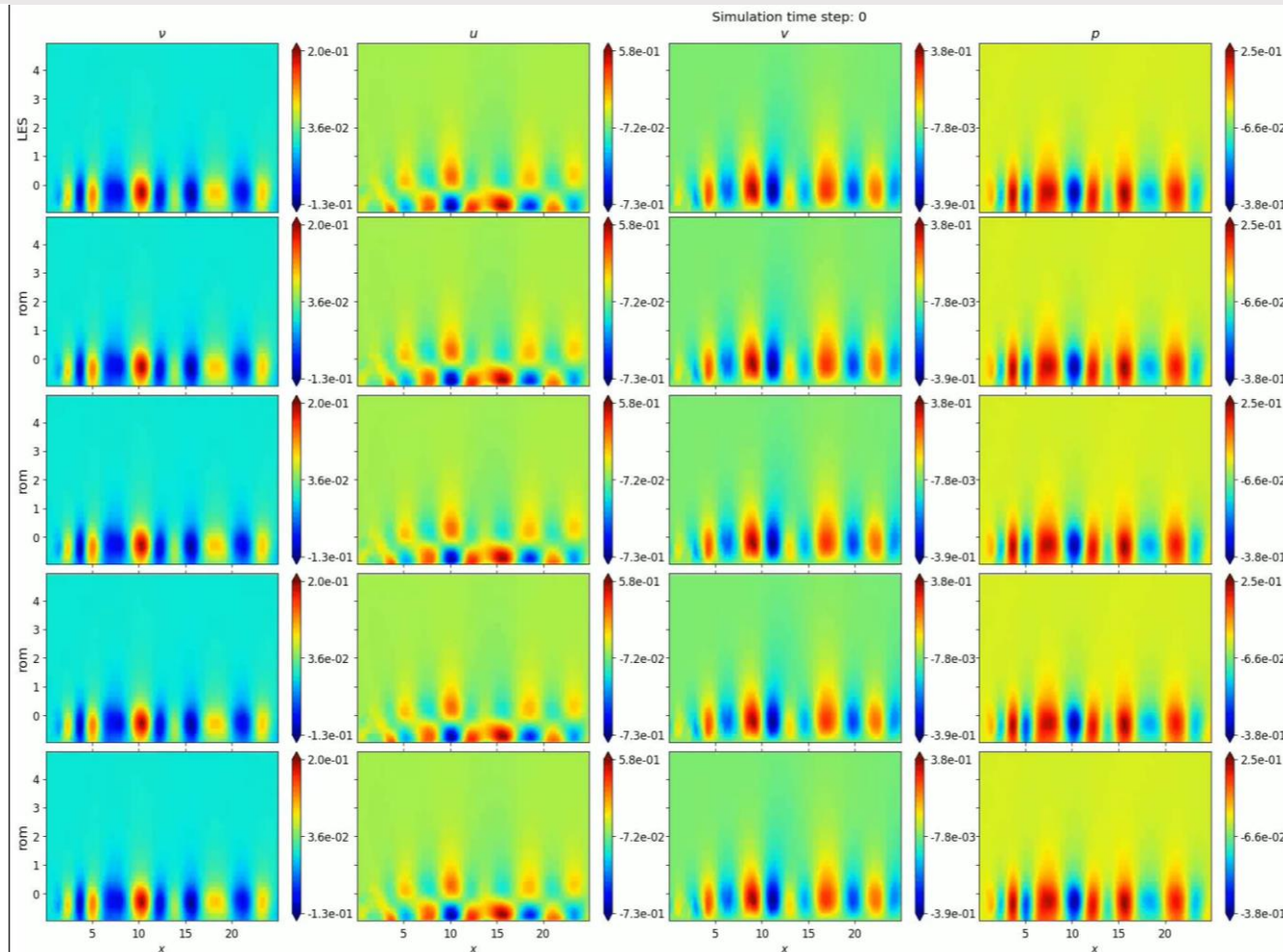
Why should we consider ROMs?

- CFD can produce informative data about our vehicle without ever having a physical test article
- One downside to CFD is the high computational cost and run time
- Reduced-order models (ROMs) can be developed to reduce the computational cost once some preliminary data is acquired
 - CFD cost: hours/days/weeks
 - ROM cost: seconds/minutes
- ROMs are less inherently trustworthy than CFD
- Trade off between fidelity and computational cost



Compressible ROM Formulation

Can we compute ROMs over compressible flow regimes?



LES

$$\lambda_{\min}^* = -0.001$$

$$\lambda_{\min}^* = -0.01$$

$$\lambda_{\min}^* = -0.05$$

$$\lambda_{\min}^* = -0.1$$

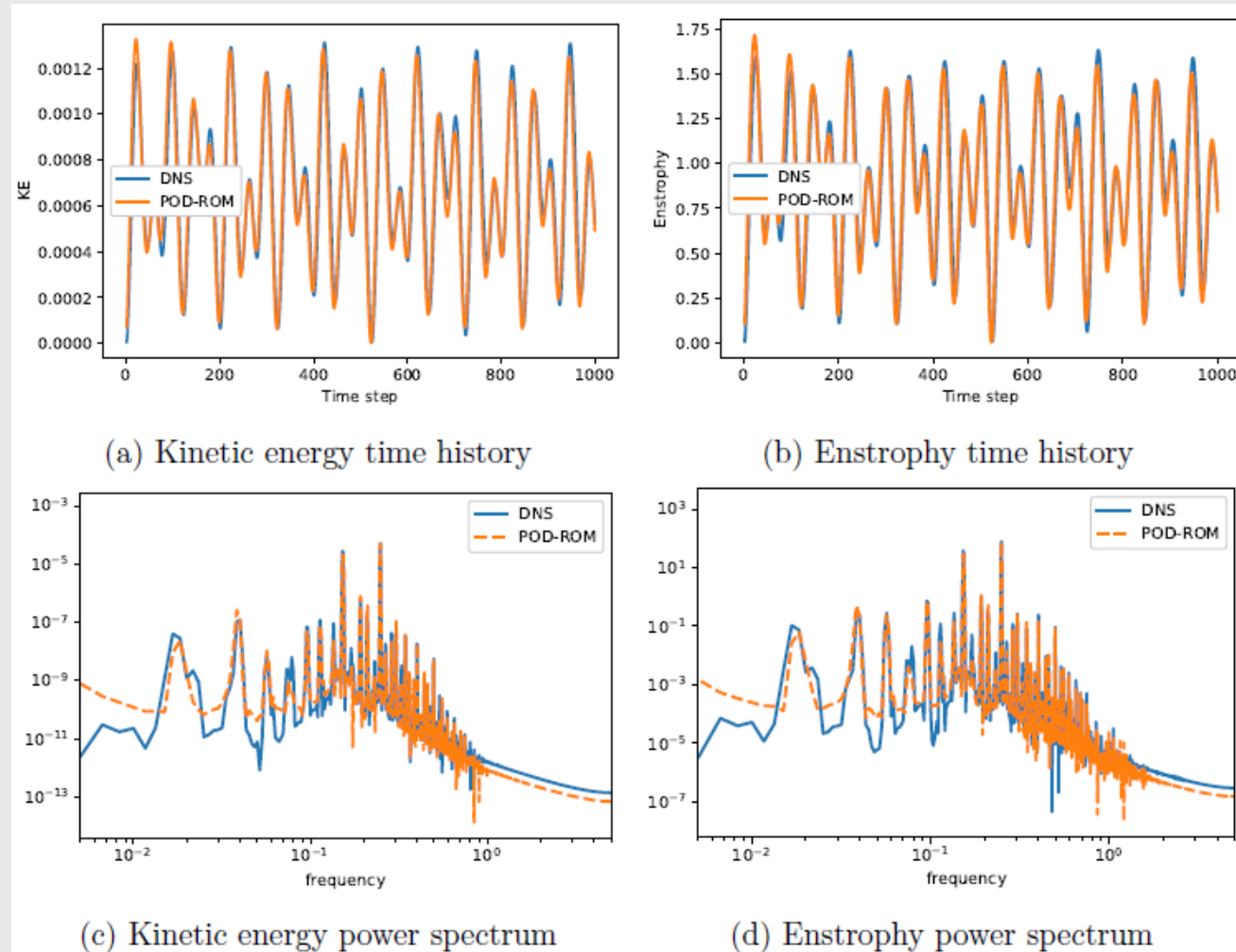
- If the hyperparameter is too close to zero, the ROM becomes unstable and diverges.
- There is a sweet spot where the ROM qualitatively matches the LES “truth”.
- If the hyperparameter is too negative, the ROM is overdamped and dissipates, deviating from the LES “truth”.

Interpolated ROMs

Can we generate ROMs where we do not have CFD results?

ROM constructed based on data at Reynolds number of 15k, 16k, 17k, and 18k.

ROM compared with direct numerical simulation at Reynolds number of 17.5k

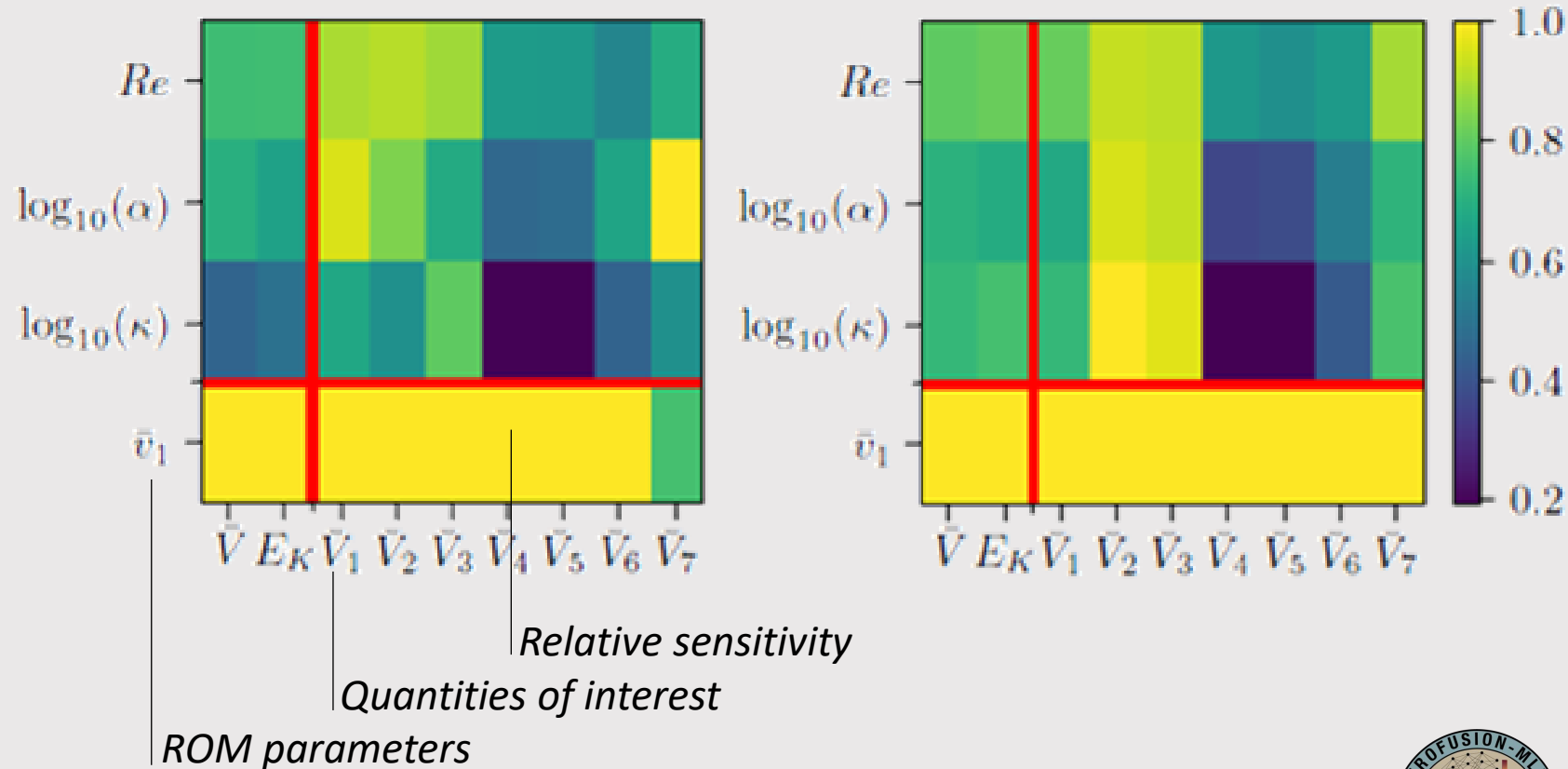


Good agreement in both the time and frequency domains

Perturbed ROMs

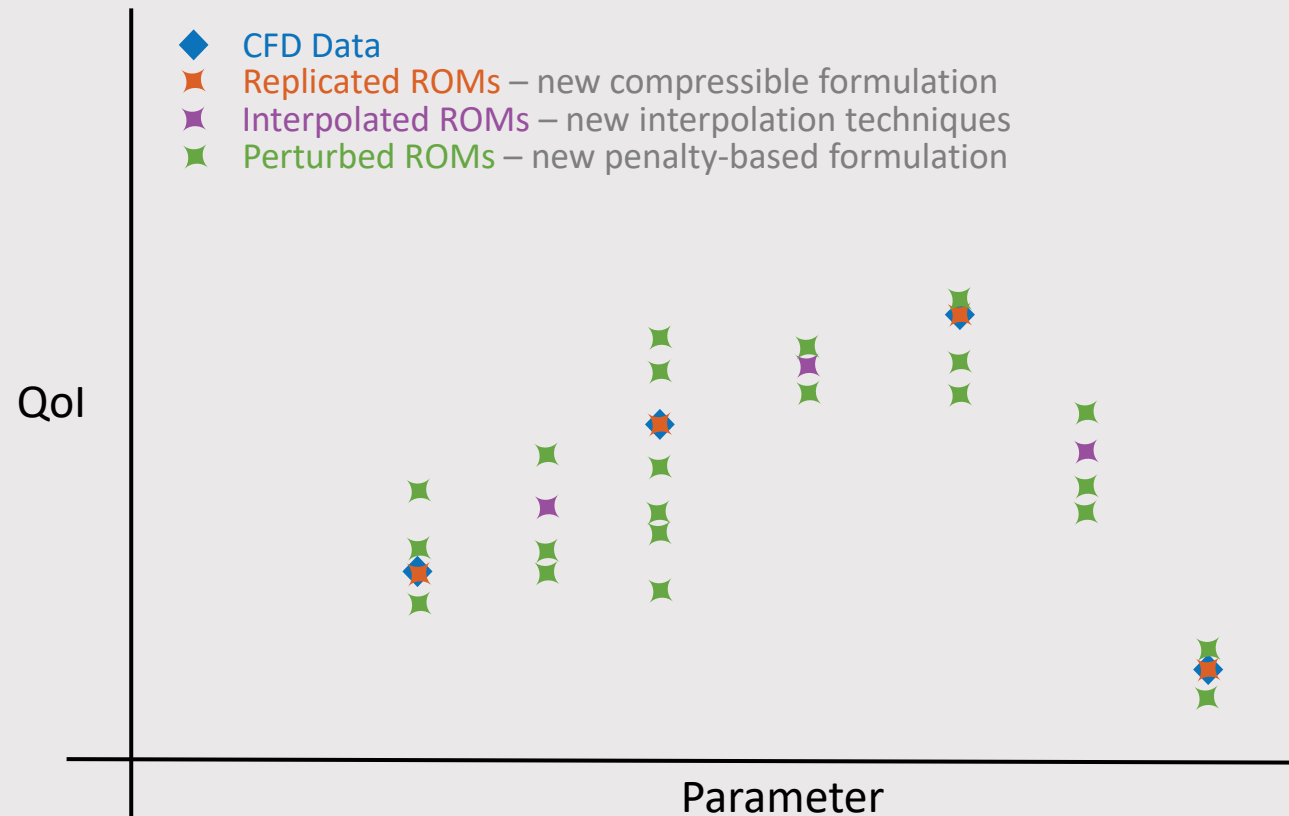
Can we examine how our ROMs perform under small perturbations?

- New boundary-penalty ROM model
- Can control boundary conditions for sensitivity analysis
- Can inform ROM uncertainty quantification



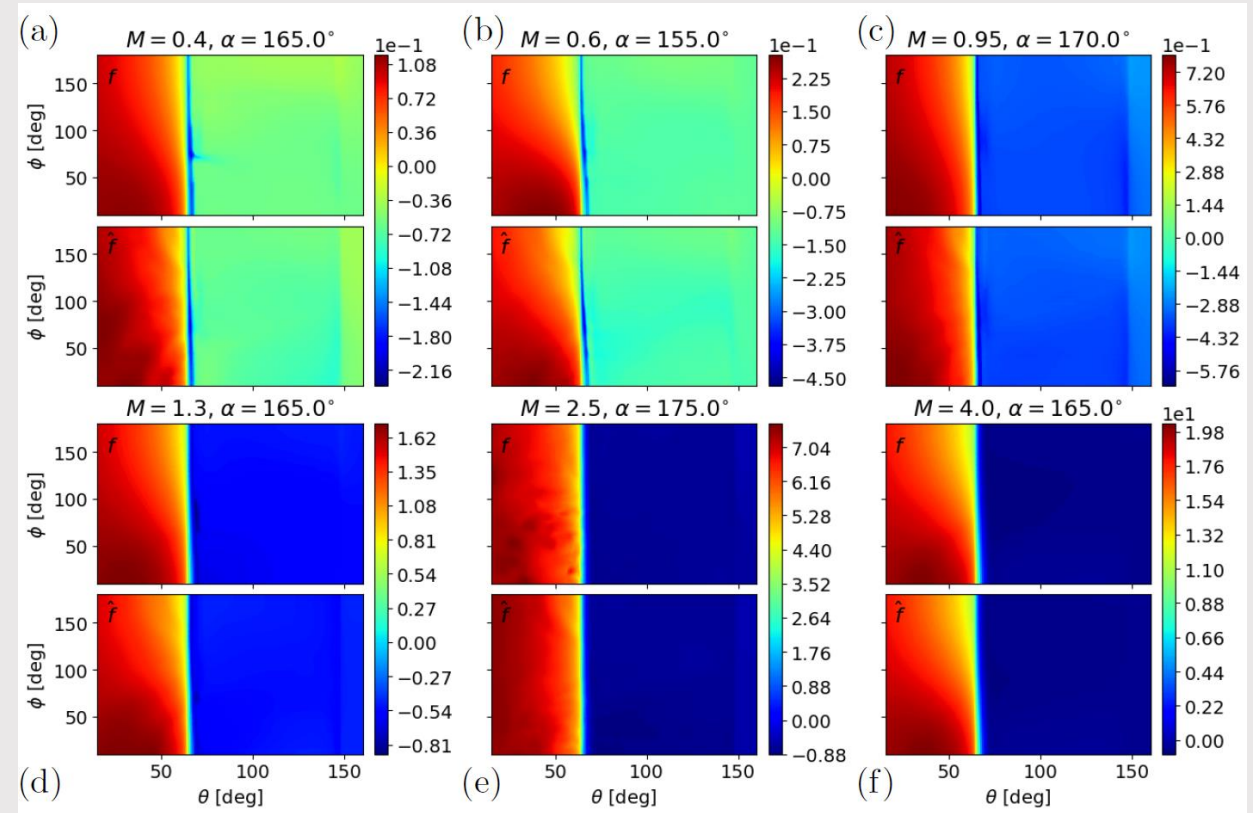
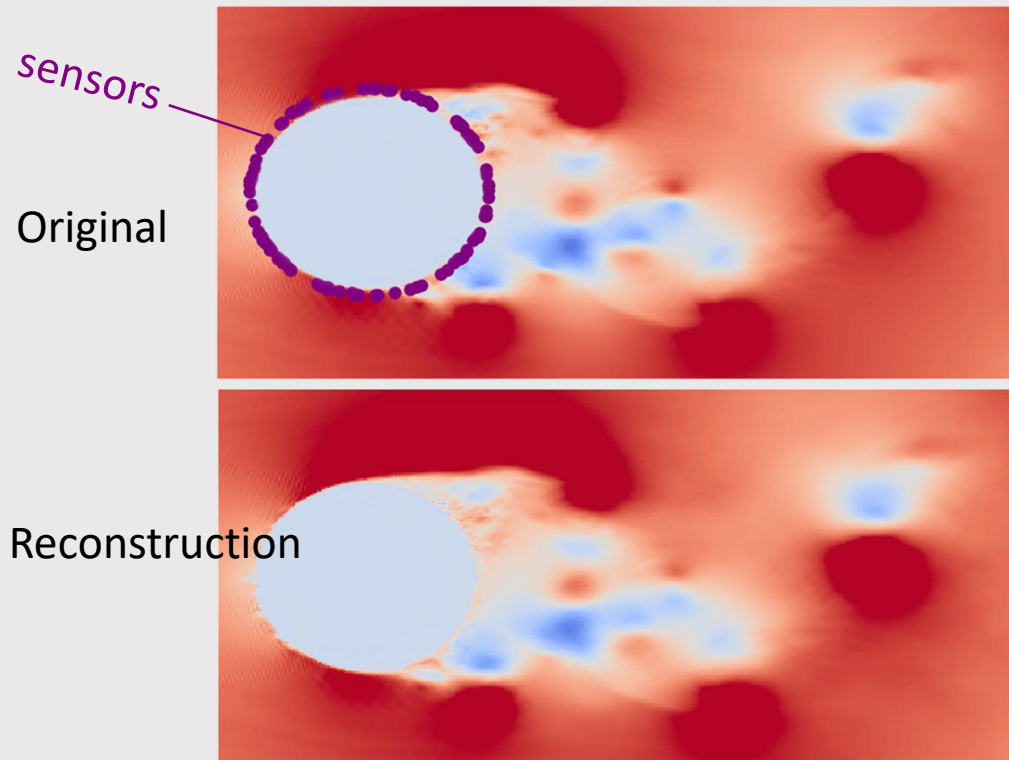
Incorporating Different ROM Techniques

How can we combine these ROM techniques?

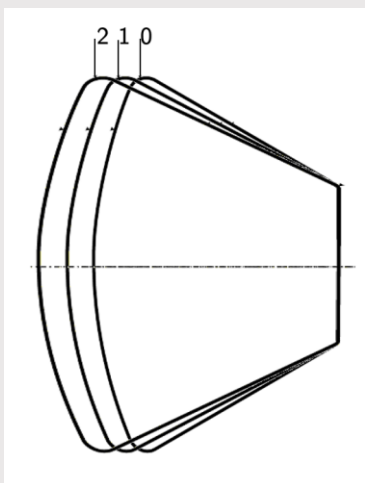
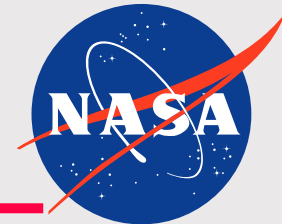


Data Reconstruction

Can we reconstruct full flow fields from sparse measurements?



Aerodynamic Modeling Process (Revisited)



Define Flight Conditions

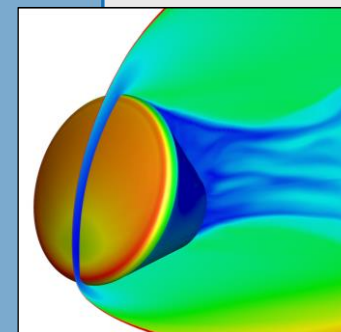
flight window
vehicle outer mold line
quantities of interest

minimize complexity

Generate Aerodynamic Data

computational fluid dynamics
wind tunnel testing
ROMs, machine learning

maximize informativity



Use Database

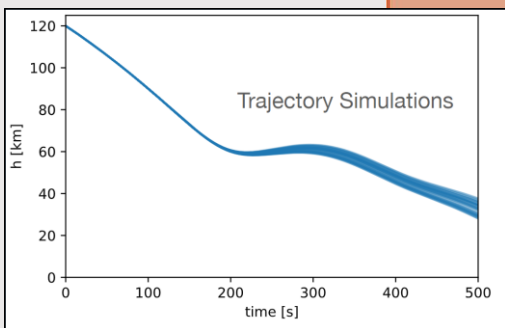
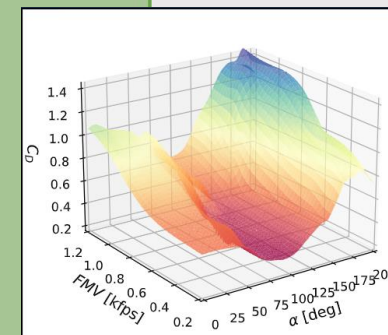
trajectory simulation
loads analysis
systems engineering

maximize relevance

Organize Database

data fusion
surrogate modeling
delivery format

maximize coverage

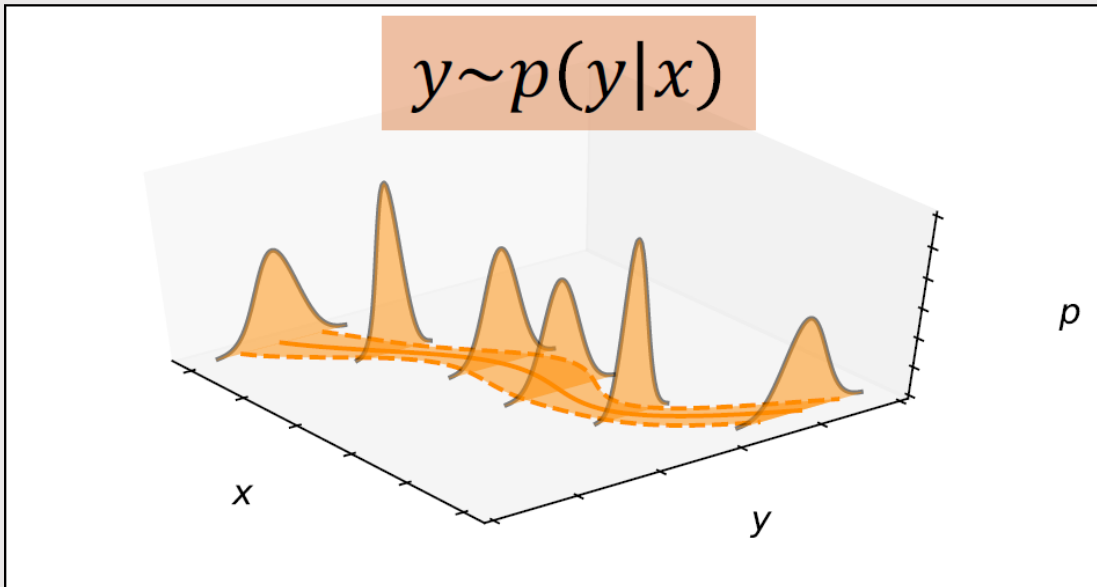


Machine Learning for Database Construction

AeroFusion Surrogate Modeling:

Once we have our data, how do we assemble the database?

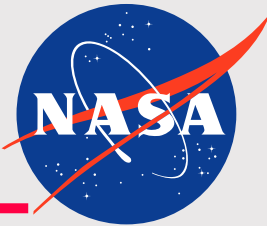
From data, learn a surrogate
conditional probability distribution



Many possible methods...

- Gaussian process regression
- Polynomial regression
- Generalized polynomial chaos
- Bayesian NNs
- Dropout NNs
- Conditional generative NNs
- Structured covariance Gaussian NNs

Gaussian Process Regression (GPR)



What is GPR?

Any finite set of index points $\{x_1, \dots, x_n\}$ represents a multivariate Gaussian distribution for function values.

prior

$$f(x) \sim p(f|x) = \mathcal{GP}(\mu, k)$$

$$\Rightarrow \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu(x_1) \\ \mu(x_2) \\ \mu(x_3) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & k(x_1, x_3) \\ k(x_2, x_1) & k(x_2, x_2) & k(x_2, x_3) \\ k(x_3, x_1) & k(x_3, x_2) & k(x_3, x_3) \end{bmatrix} \right)$$

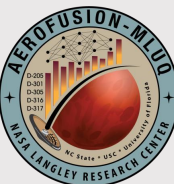
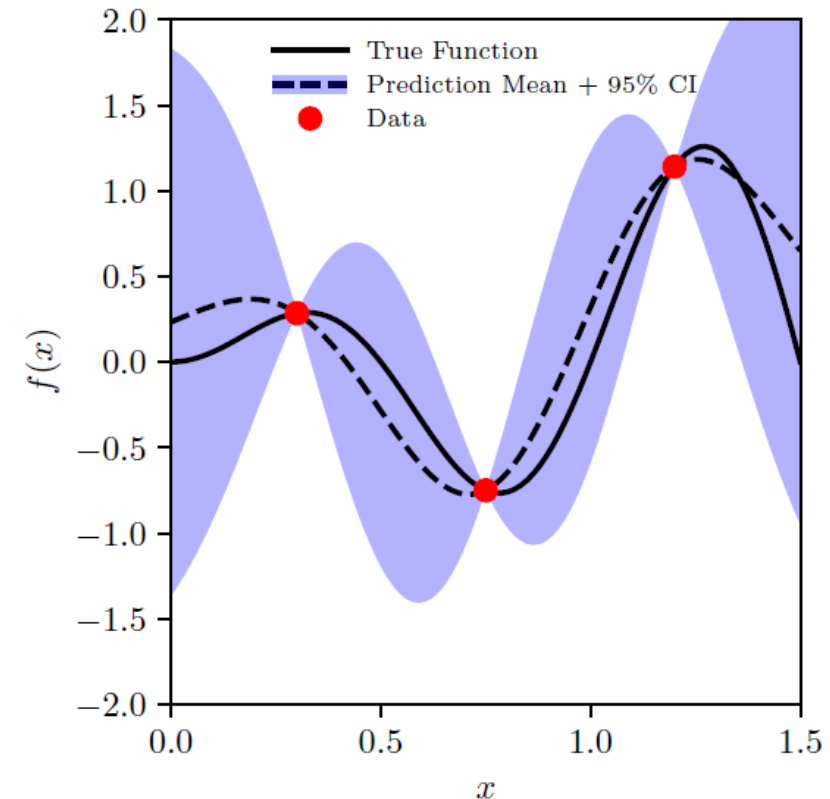
Regression performed by conditioning the distribution on data $\mathcal{D} = (X, y)$

posterior

$$\hat{f}(x) \sim p(f|\mathcal{D}, x) = \mathcal{GP}(\hat{\mu}, \hat{k})$$

$$\hat{\mu}(x) = \mu(x) + k(x, X) k(X, X)^{-1} (y - \mu(X))$$

$$\hat{k}(x, x') = k(x, x') + k(x, X) k(X, X)^{-1} k(X, x')$$



Multi-hierarchy GPR

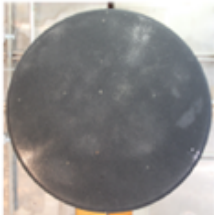
Can we account for multiple effects without a clear hierarchy of fidelity?

Easier to categorize as nominal and off-nominal than to establish a hierarchy of fidelity levels

Asymmetric, Smooth



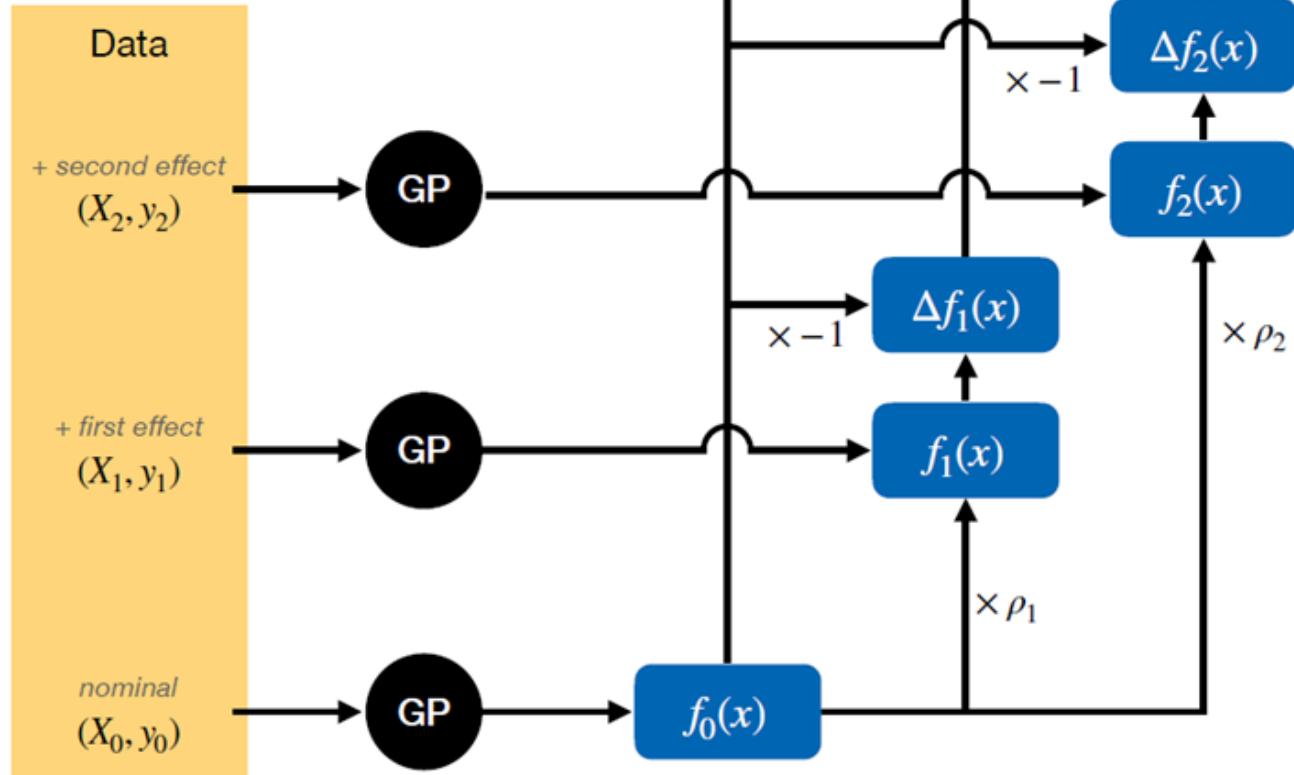
Symmetric, Rough



Symmetric, Smooth



Orion heatshield models used in 133-CA test campaign in National Transonic Facility



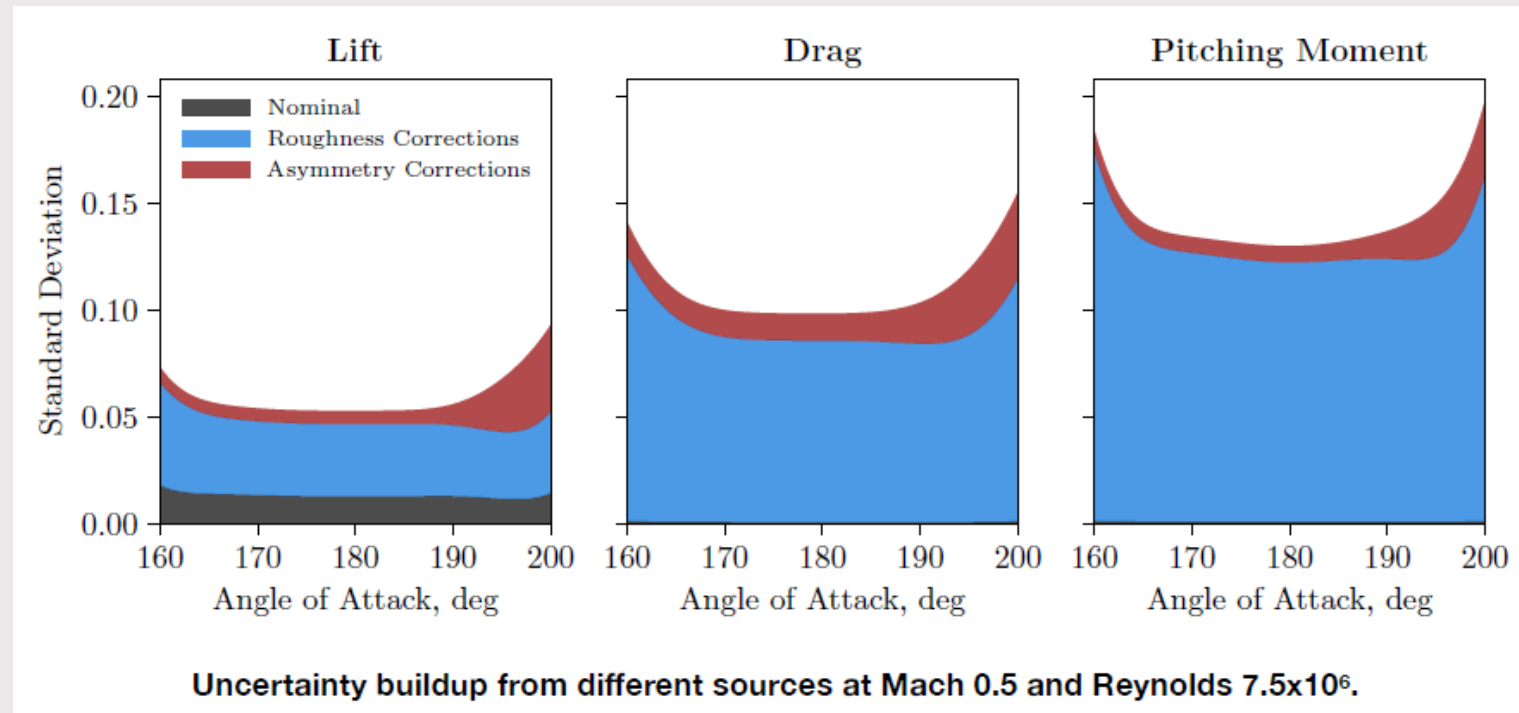
Correction functions account for change in nominal

Multi-fidelity hierarchy for each correction model

Separating Uncertainty Sources

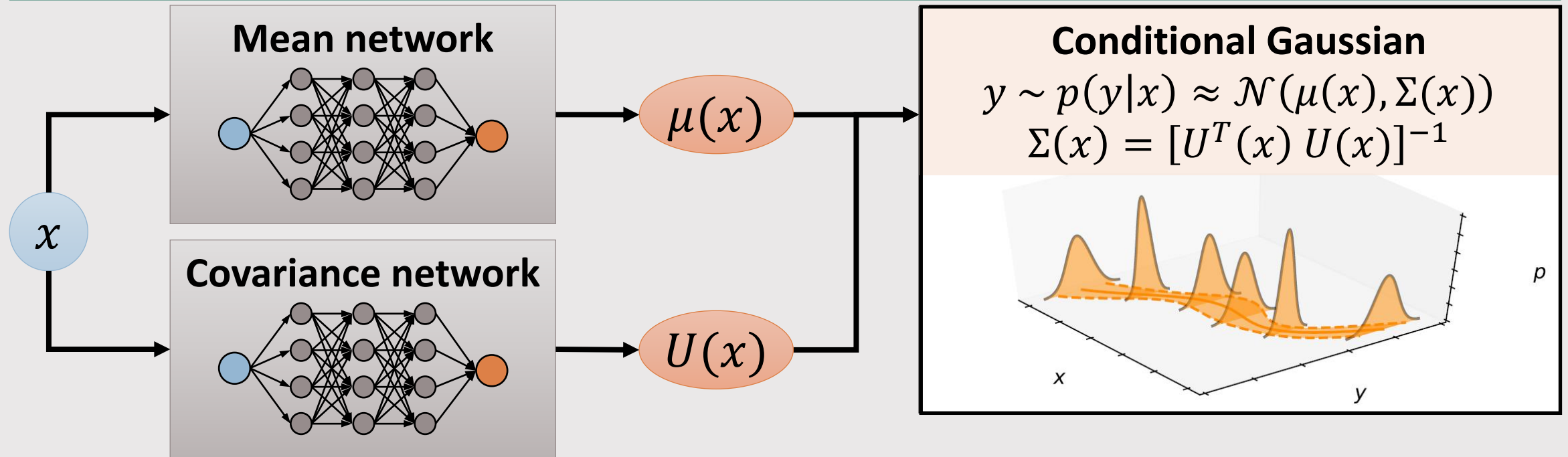
What does the multi-hierarchy structure buy us?

- Can attribute uncertainty from different sources to overall uncertainty in predictive distribution
- Only uncertainties present in data, does not include those associated with factors not considered here



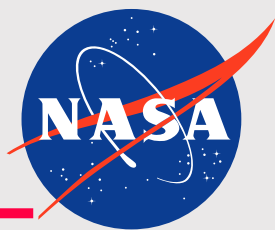
Structured Covariance Gaussian Networks (SCGNs)

Can we achieve similar performance to GPR but with better scalability?

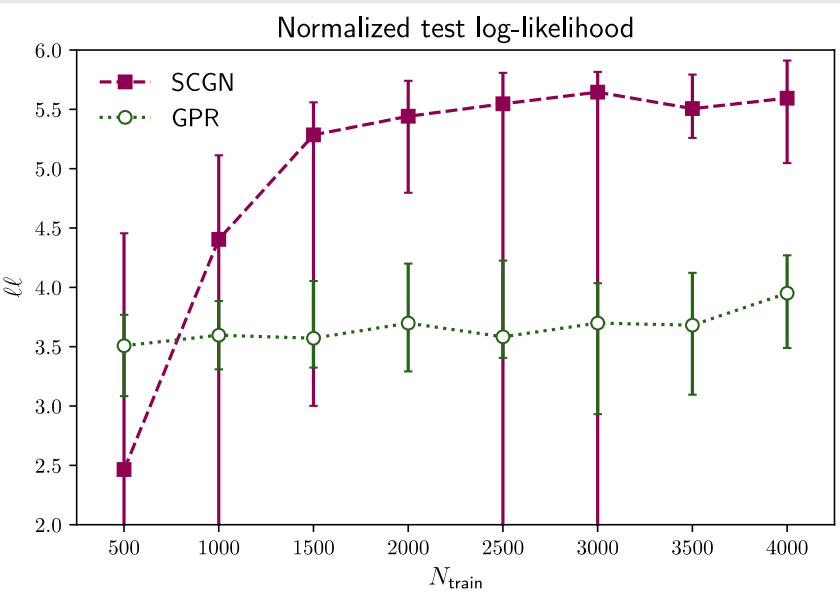
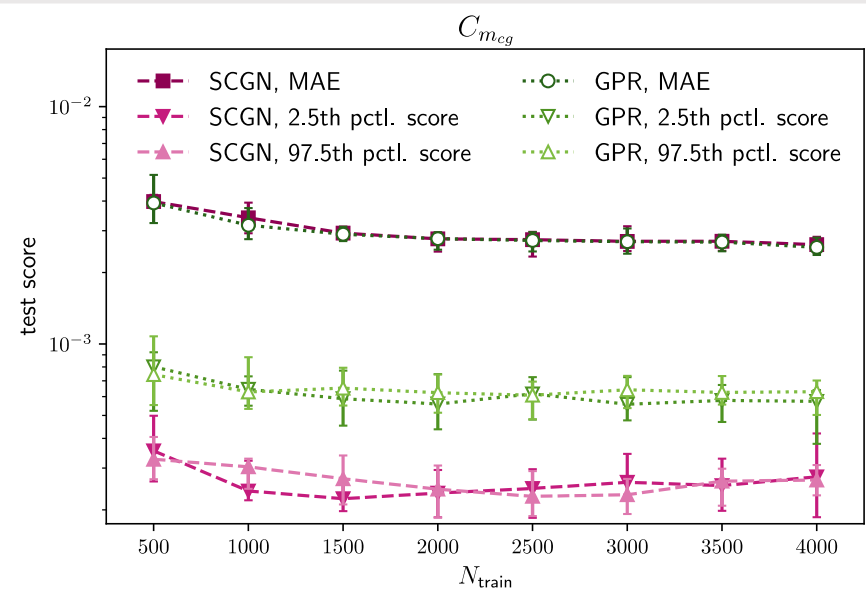


- $U(x)$ is an **upper triangular matrix** $\Rightarrow$ **Cholesky decomposition** of inverse covariance
- Take exponent of diagonal to make positive $\Rightarrow$ **positive definite** covariance
- Proposed for UQ of image generation NNs in *Structured uncertainty prediction networks* (Dorta et al., *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018)

Comparison of SCGN and GPR

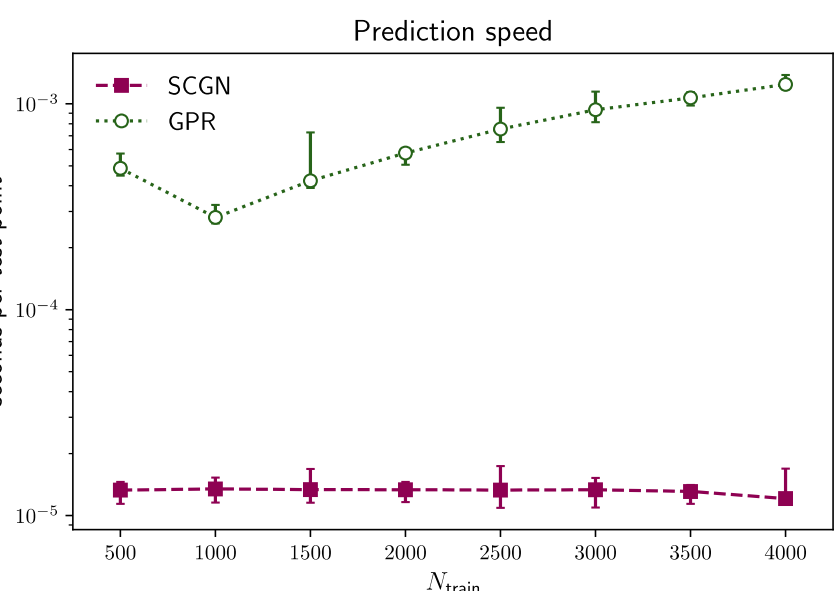
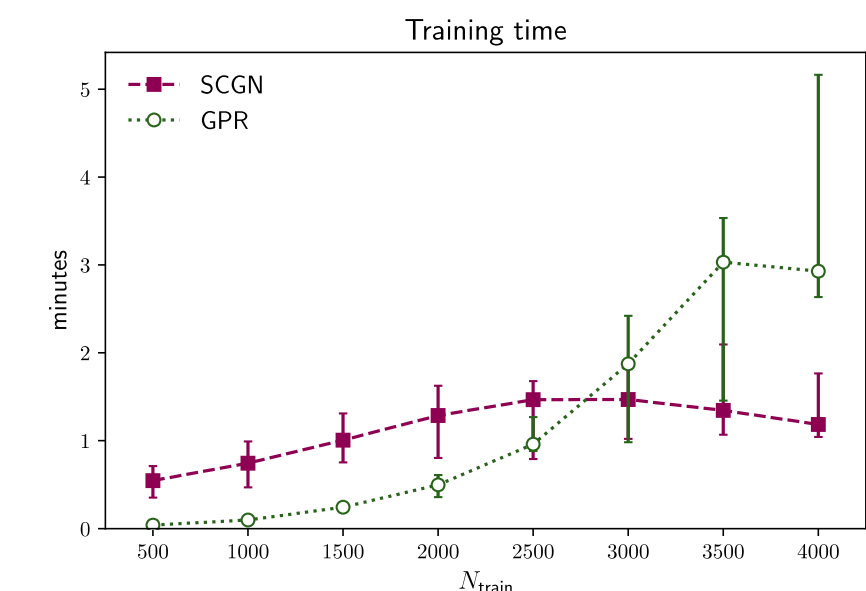


Similar performance on maximum absolute error



SCGN shows higher log likelihood

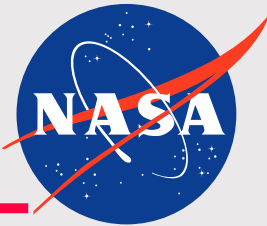
SCGN scales better for large datasets



SCGN has better prediction speed

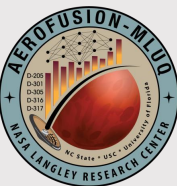


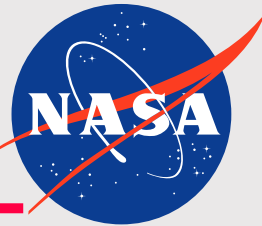
Summary



What have we discussed today?

- Data Generation
 - CFD sensitivity analysis – where to collect data
 - ROMs – nominal values and uncertainties
 - Compressible ROMs – nominal possible for complex flows
 - Interpolated ROMs – nominal can be computed between CFD test points
 - Perturbed ROMs – assessing the impact of perturbations
 - Reconstructing data – learning more from our measurements
- Database Assembly
 - Multi-hierarchy GPR – isolating uncertainty structures
 - SCGN – addressing scalability for large data sets





Acknowledgments

- Funded by the Early Career Initiative (ECI) within NASA’s Space Technology Mission Directorate (STMD)
- Contributors and Collaborators:

NASA Langley Research Center	University of Southern California	North Carolina State University	University of Florida
Steven Snyder (Lead)	Ivan Bermejo-Moreno (PI)	Ralph Smith (PI)	Alina Zare (PI)
TJ Wignall (Lead)	Donya Ramezani	Chris Edwards	Yury Lebedev
Michael Lee (University Coordinator)	Jonas Buchmeier	Harley Hanes	Matthew Schrank
Brendon Colbert		Walker Powell	
Justin Green			
Tenavi Nakamura-Zimmerer			
JB Scoggins			
Mary Stringer			
R. Anthony Williams			
Interns: A. Cabrera, J. Katona, S. Kumar			

