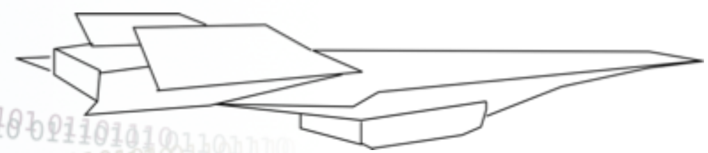
Application of Adaptive Sampling to Advance the Metamodeling and Uncertainty Quantification Process

Erik L. Axdahl and Robert A. Baurle
Hypersonic Airbreathing Propulsion Branch
NASA Langley Research Center



Approved for Public Release

DATAWorks 2018 Workshop
20—22 March 2018
Springfield, VA

Motivation and Rationale

Is Big data always better?
What did you give up for the knowledge you gained?

- Why converge the model of Subsystem A to +/- 0.1% error when +/- 5% would have sufficed?
  - What if Subsystem B needed more resources [people, cluster time, tunnel time] that could have been a better allocation of resources?
  - *So track both model convergence and the top down requirements of the project.*

- Why predefine your design space sampling scheme before collecting any data?
  - What if Subsystem C dedicated a significant number of samples that did not contribute significantly to the variability of the model?
  - *So understand the error properties of the model and how to use that information to choose the next best set of points to sample.*

State-of-the-art R&D doesn't go as planned…so at least be responsive as you collect your data.

Adaptive Sampling Benefits

How does adaptive sampling buy its way onto a project?

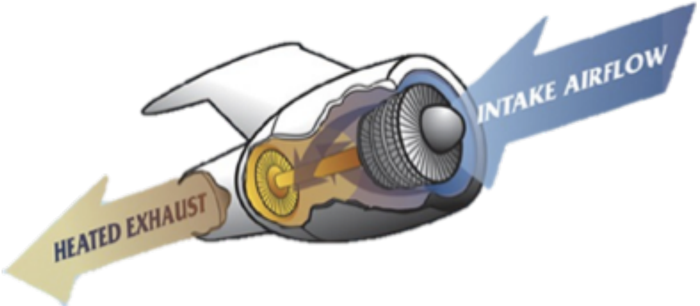
- **Better team communication**
  - *Having more frequent conversations about how a dataset is evolving avoids a "throw it over the fence" mentality.*

- **Improved metamodel confidence**
  - *Treat your metamodel similarly to CFD, and check to see if it is asymptotically converging.*

- **Catch issues earlier on**
  - *Unusual or unexpected behavior, if nonphysical, can be corrected for future data samples.*

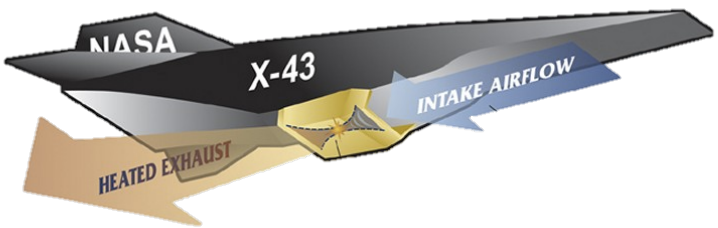
- **Better allocation of resources**
  - *Stop collecting subsystem data when it no longer improves the fidelity of overall system predictions.*
  - *Improve the efficiency of very high intensity tasks, such as UQ.*

Application: High Speed Air Breathing Propulsion



Turbojet: Mach 0 → 2-3

Low to Moderate Airframe Integration



Ramjet: Mach 2 → 5

DMRJ: Mach 4 → 8
Scramjet: Mach 8+

Our present regime of interest

High Airframe Integration

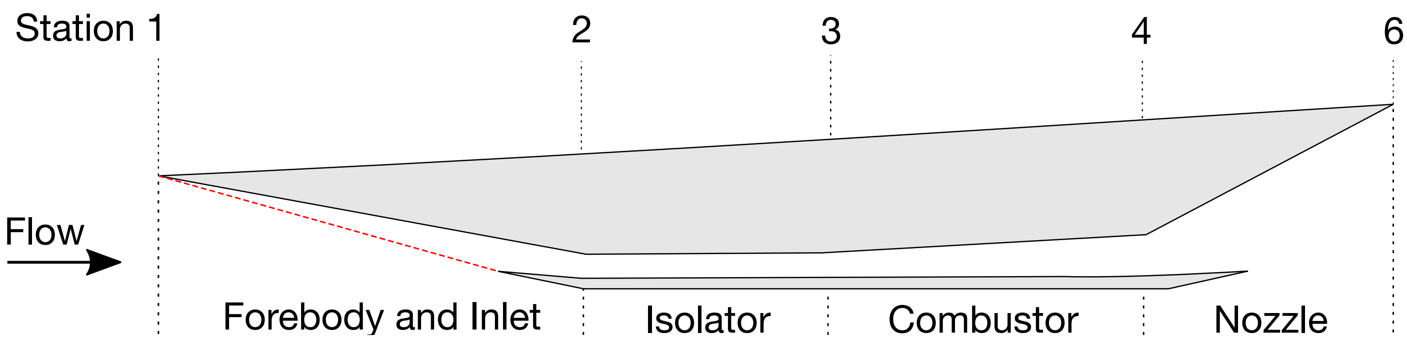
Predicting the performance of integrated scramjet and DMRJ systems requires efficient data collection, metamodeling, and UQ because:

1. Point (numerical or experimental) simulations are expensive due to high enthalpy facilities or need for highly resolved, reacting CFD.
2. As airbreathing systems scale, they quickly outpace the size of existing ground test facilities, requiring predictive flight CFD to be validated against representative component ground testing.

Two examples of applying the adaptive sampling methodology in remainder of presentation

Example 1: Uninstalled Air-Specific Impulse

As a model problem, the maximum uninstalled air-specific impulse performance of a hydrogen-fueled supersonic combustion propulsion cycle was evaluated over a Mach number range of 6 → 12 and an inlet kinetic energy efficiency range of 0.93 → 0.97. All other cycle parameters were held to fixed values.



$$I_a = \frac{Fn}{\dot{W}_a} = \frac{F_6 - F_1 - p_1(A_6 - A_1) - \cancel{F_{add}}}{g_0\, \dot{m}_{air}}$$

neglect

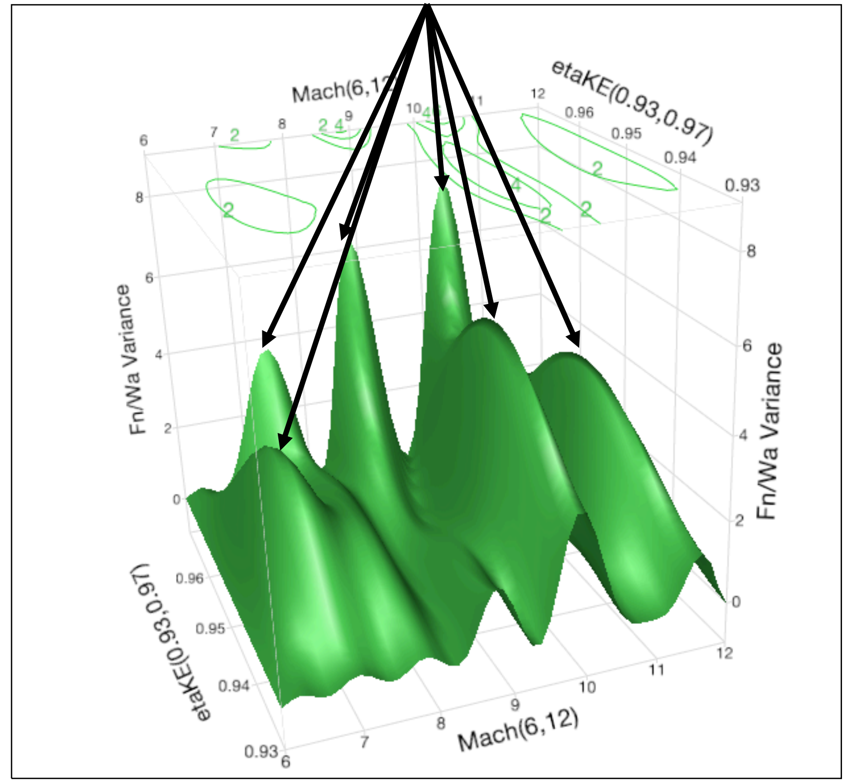
This is a challenging design space to model because it is ridged due to not being able to use a stoichiometric fuel-to-air ratio everywhere.

Adaptive Sampling Methodology

The adaptive sampling methodology assumes that a metamodel is used in tandem with the data collection.

1. Construct metamodel using the existing collected data.
2. Use the model variance to search for the next best point(s) to sample. For multiple samples at once, rank variance optima in descending order.
3. Perform data collection at the selected points in the design space.
4. Refit the metamodel and compute the RMSE to assess metamodel convergence.
5. Repeat steps 2 → 4 until convergence or end of budget/time.

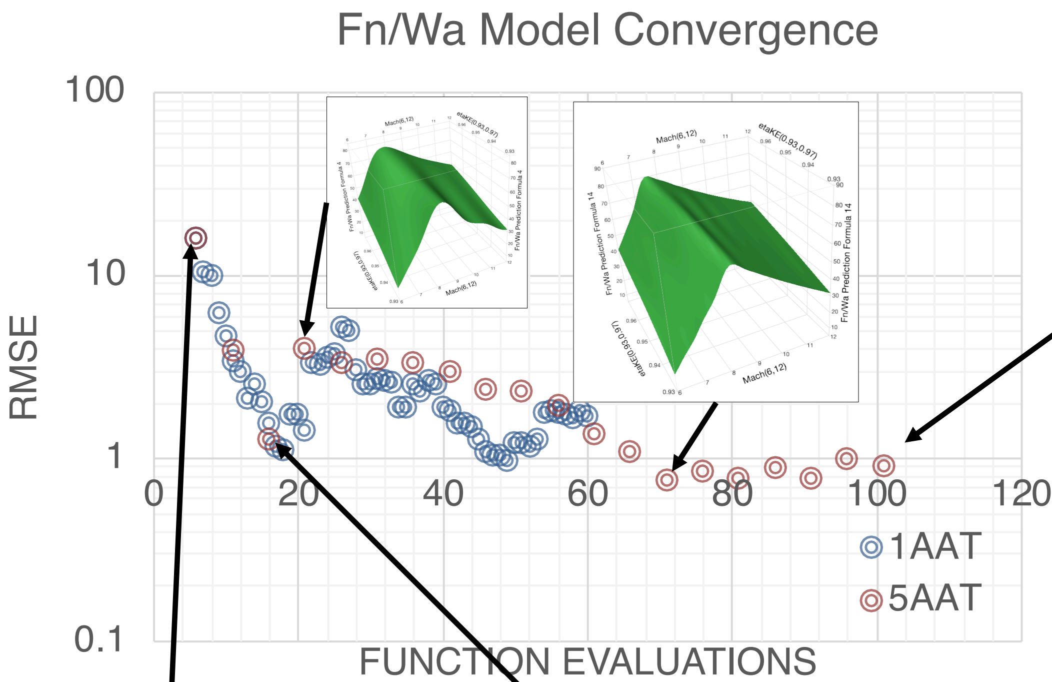
Candidate Sampling Locations

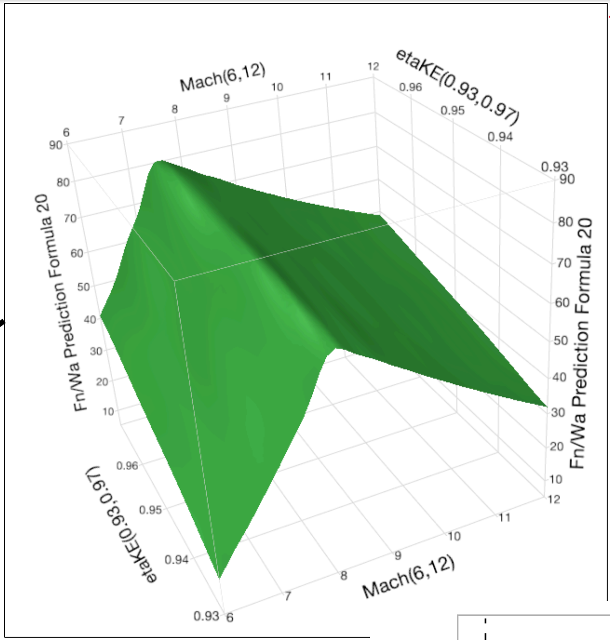


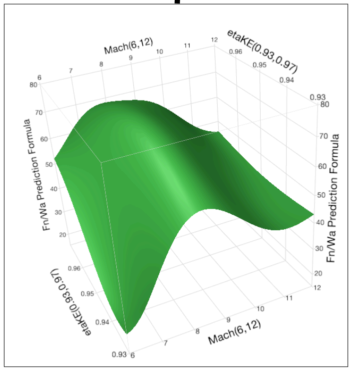
$$RMSE = \sqrt{\frac{\sum_i^N \left(R_{jackknife,i} - R_i\right)^2}{N}}$$

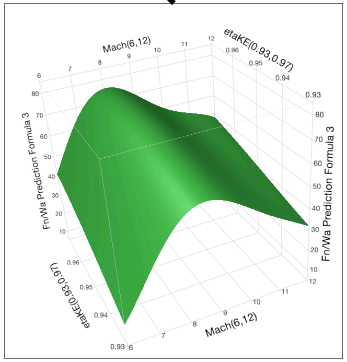
The following analyses use a Gaussian Process (Kriging) metamodel form to model the raw data

Uninstalled Fn/Wa Metamodel Evolution



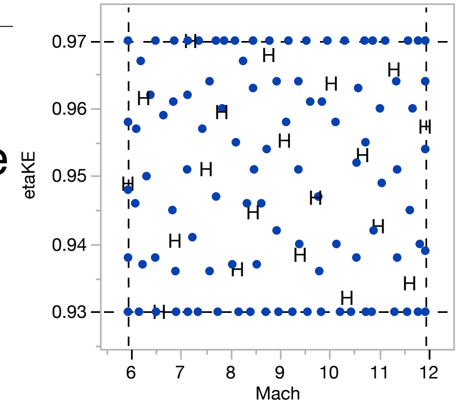






Metamodel begins to take shape after 21 samples and converge after 71 samples.

Heavy data collection along the max and min etaKE boundaries.

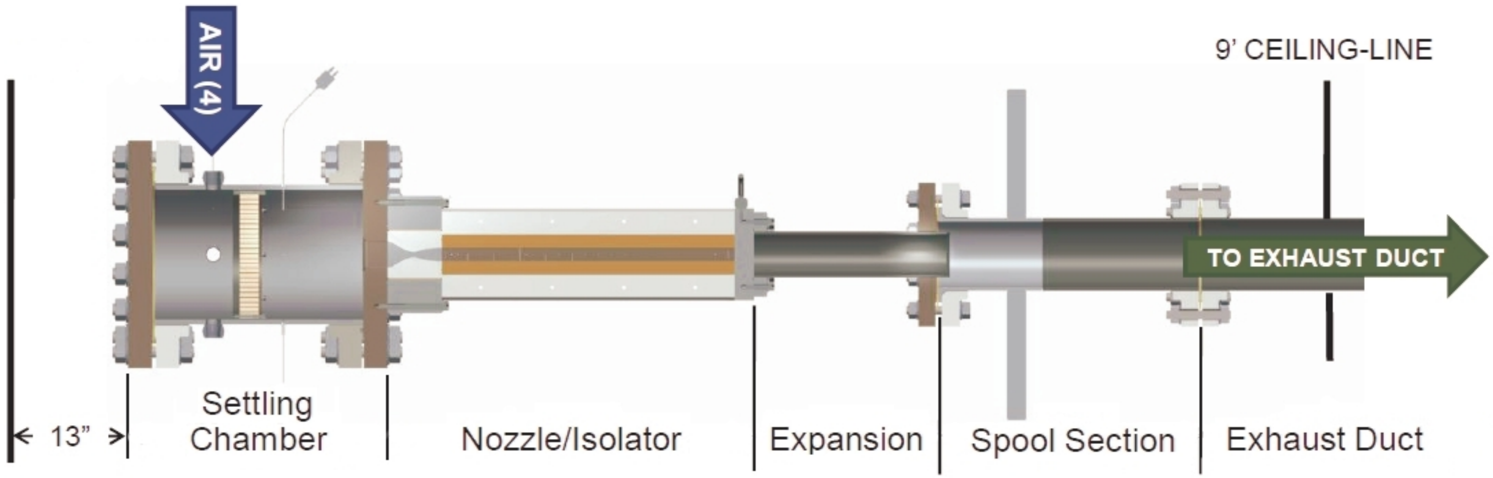


▪ = Model Sample

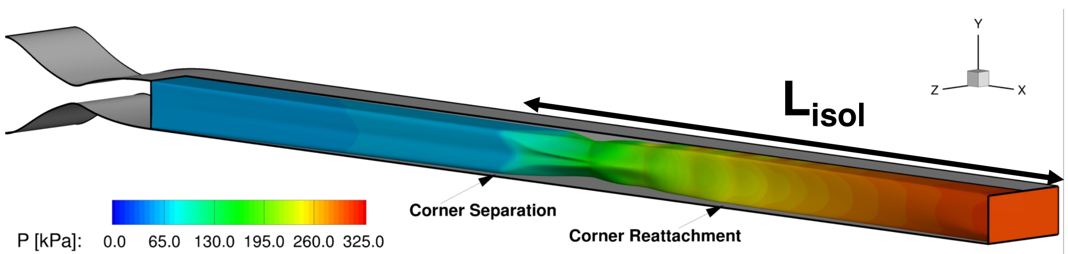
H = Holdback Sample

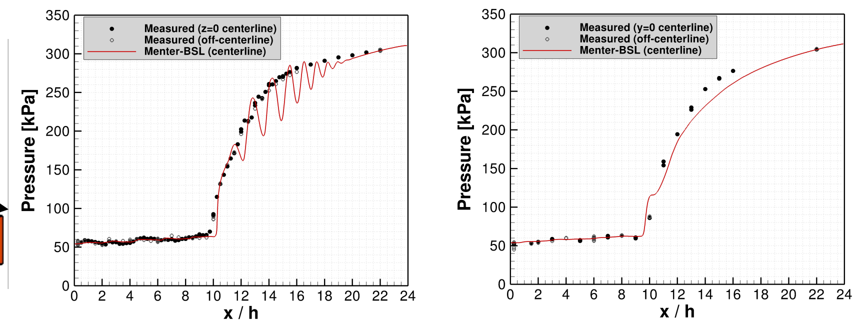
Collecting one sample at a time (1AAT) competitive with collecting five at a time (5AAT)

Example 2: Isolator Dynamics Research Laboratory (IDRL)



The IDRL is an experimental apparatus at NASA Langley Research Center (LaRC) designed to be a validation testbed for numerical simulations of separated boundary layer flows in DMRJ isolators.

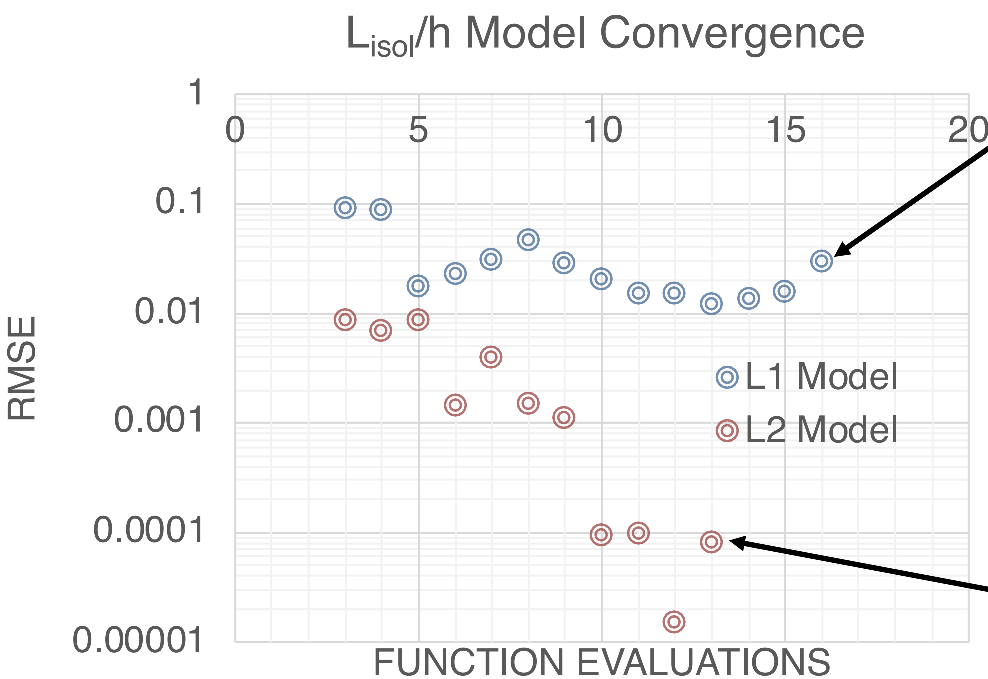


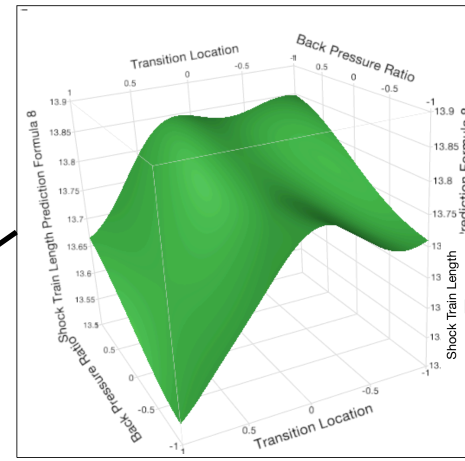


High fidelity simulations were carried out using VULCAN-CFD, developed and maintained at LaRC's Hypersonic Airbreathing Propulsion Branch (HAPB).

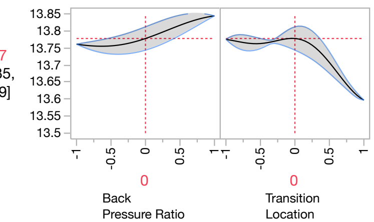
Isolator Separation Length Metamodel Evolution

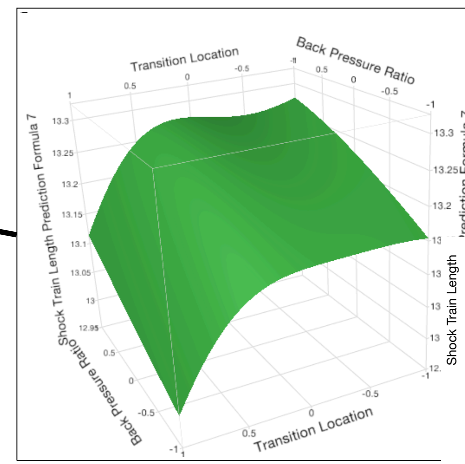
L_{isol}/h = f(Backpressure ratio, turbulent transition location, grid level)



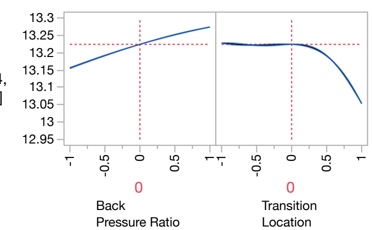


Medium Grid (L1) Model





Coarse Grid (L2) Model

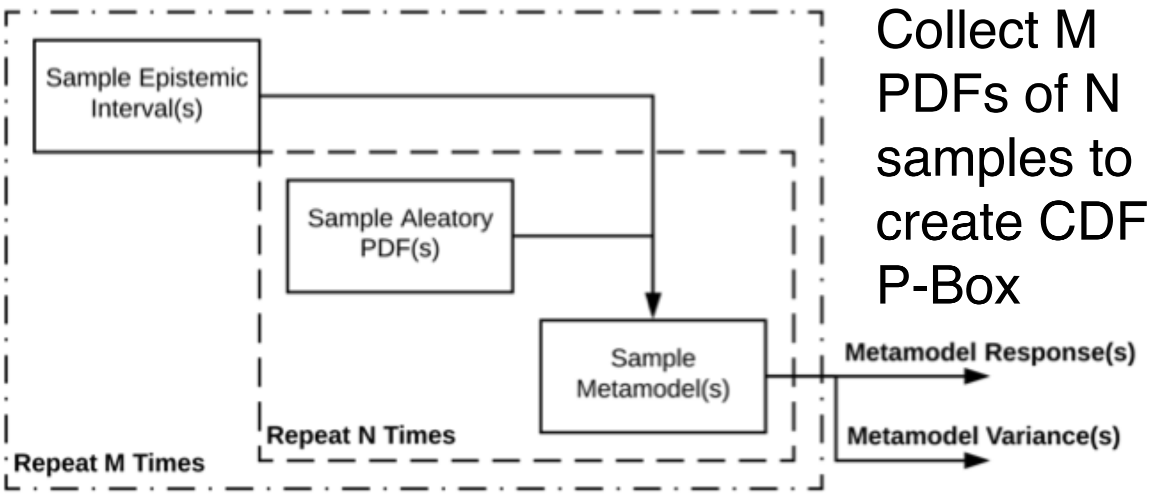


- The coarse grid model is converged over two orders of magnitude RMSE.
  - *Additional cases would not be helpful*.
- The fine grid model is not converged, but metamodel prediction interval can be quantified for UQ propagation.

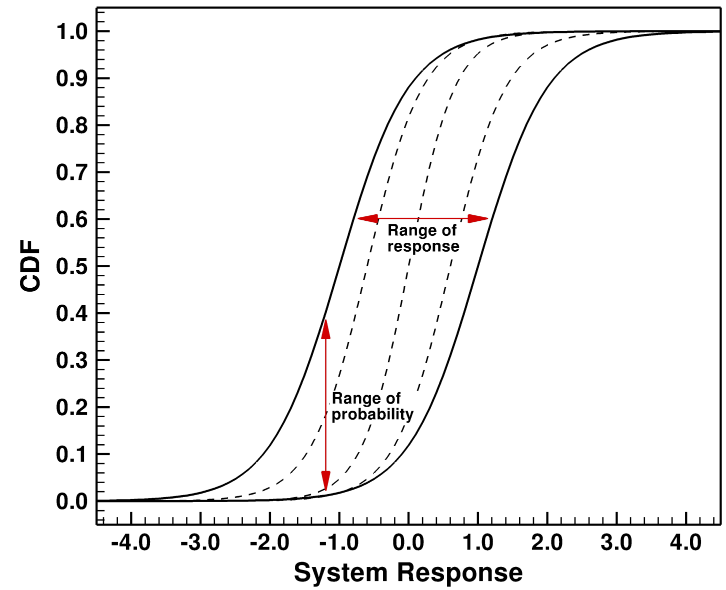
Uncertainty Propagation using Metamodels

Isolator simulation is a good uncertainty quantification test problem due to the sensitivity of the **length of the separated isolator flow** to quantities such as back pressure, grid level, turbulence model, etc.

Monte Carlo uncertainty propagation directly using CFD is too expensive. Medium grid level is order of magnitude more expensive than coarse grid level. Use metamodels to alleviate computational expense.



Collect M PDFs of N samples to create CDF P-Box

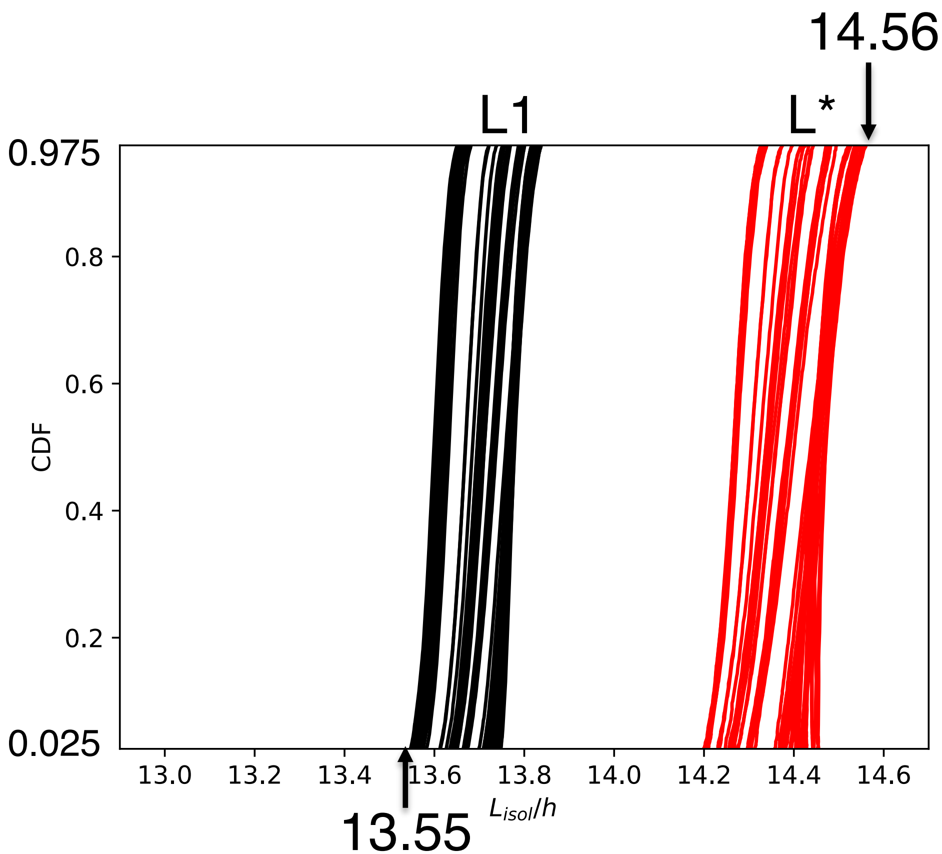


Lisol/h = f(Aleatory, Epistemic) = f(Backpressure ratio, transition location, grid level)

For IDRL UQ background, see R.A. Baurle, E.L. Axdahl, "Uncertainty Quantification of CFD Data Generated for a Model Scramjet Isolator Flowfield," JANNAF, Dec 2017.

Isolator Separation Length Uncertainty Propagation

Before drawing final P-Box, compute new CDF (L*) that takes discretization into account via grid convergence index (more details in backup). The L* and L1 families form the new P-Box.



Taking asymptotic grid convergence into account yields an **Lisol/h** 95% confidence interval of **[13.55, 14.56]** (~ 1 duct height)

Summary and Conclusions

Adaptive sampling has been demonstrated for two example problems in high speed air breathing propulsion.

- Using adaptive sampling with a convergence metric allows for a measure of confidence of how well saturated a model is.

- Tracking convergence of parallel models indicates where future resources should be focused.

- Adaptive sampling can be used to “batch” a group of promising cases, instead of simply relying on additional cases having good space filling properties.

- Adaptive sampling along with an appropriate metamodel form does a good job of resolving nonlinearities and near-discontinuities in design spaces.

- Adaptive sampling may be able to make UQ problems more tractable during the design process.

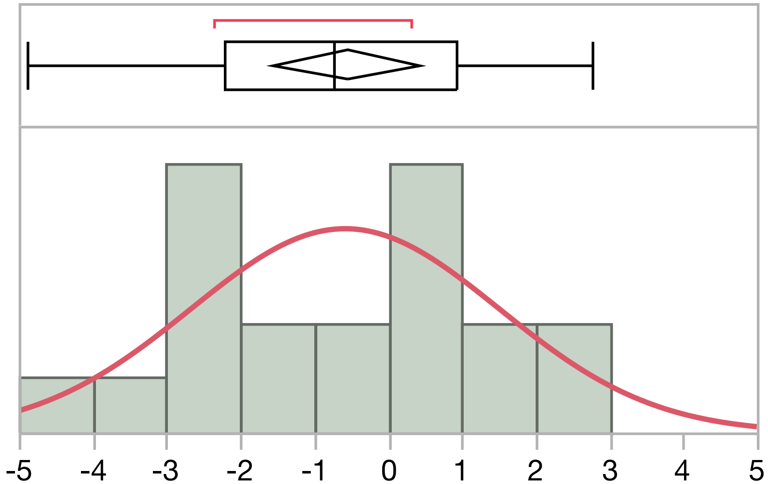
Thank You!

Erik L. Axdahl, Ph.D.
Research Aerospace Engineer
Hypersonic Airbreathing Propulsion Branch
NASA Langley Research Center
Office: (757) 864-8318
email: Erik.L.Axdahl@NASA.gov

Backup

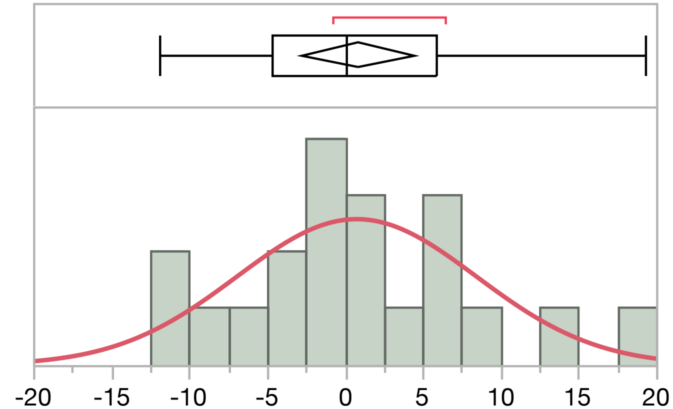
Fn/Wa Model Response Error

How do the 20 holdback points not used to fit the Kriging metamodel compare against the model predictions?



Normal(-0.5773,2.08873)
MRE within +/- 5%

Aside:
How does a least squares fit perform?



Normal(0.77918, 7.68911)
MRE within +/- 20%
Data has 5/101 violations of 95% PI

Is this bad? Maybe not as long as your report the model's prediction interval along with the mean model.

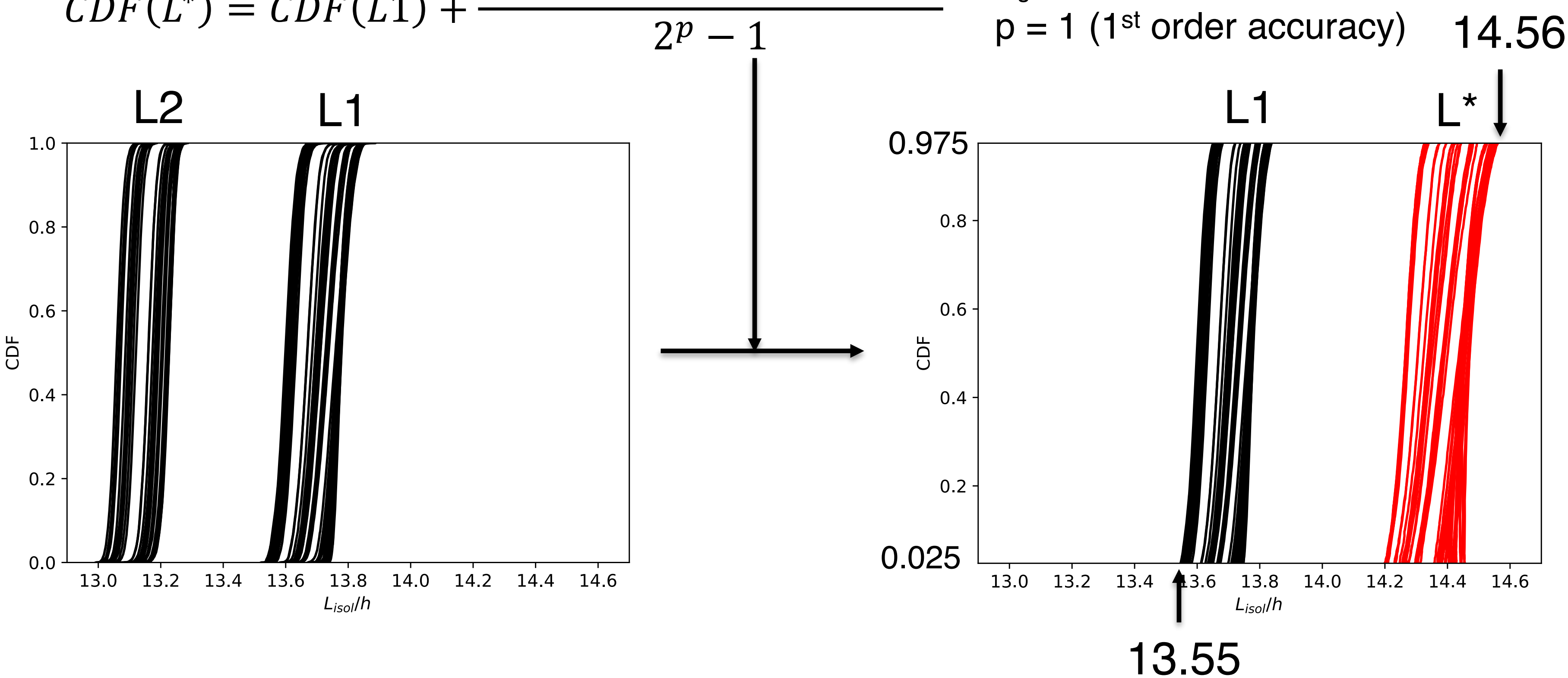
As a deliverable, always report your model error properties along with the model itself.

Isolator Separation Length Uncertainty Propagation

Before drawing final P-Box, determine new CDF that takes discretization into account via grid convergence index. The new CDF (L^*) and L1 form the new P-Box.

$$CDF(L^*) = CDF(L1) + \frac{F_s(|CDF(L1) - CDF(L2)|)}{2^p - 1}$$

$F_s = 1.25$
$p = 1$ (1st order accuracy)



Taking asymptotic grid convergence into account yields an **Lisol/h** 95% confidence interval of **[13.55, 14.56]** (~ 1 duct height)