

# Global and Subdomain Validation of Computer Models

*Rui Tuo*

*Department of Industrial & Systems Engineering*

*Texas A&M University*

Joint work with:

Chaoan Li

Department of Industrial & Systems Engineering

Texas A&M University

Xianyang Zhang

Department of Statistics

Texas A&M University



## Overarching goal



TEXAS A&M UNIVERSITY  
Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

- Computer simulations play a significant role in today's scientific discovery and engineering advancements.
- The validation of computer simulation models is a critical step in ensuring their reliability and accuracy before they are employed in practical applications.
- The goal of validation is to determine whether the computer input-output relation agrees with the *physical process* (expected potential physical outcome).
- In this work, we propose a rigorous frequentist statistical methodology to validate a computer model globally or on a subdomain.



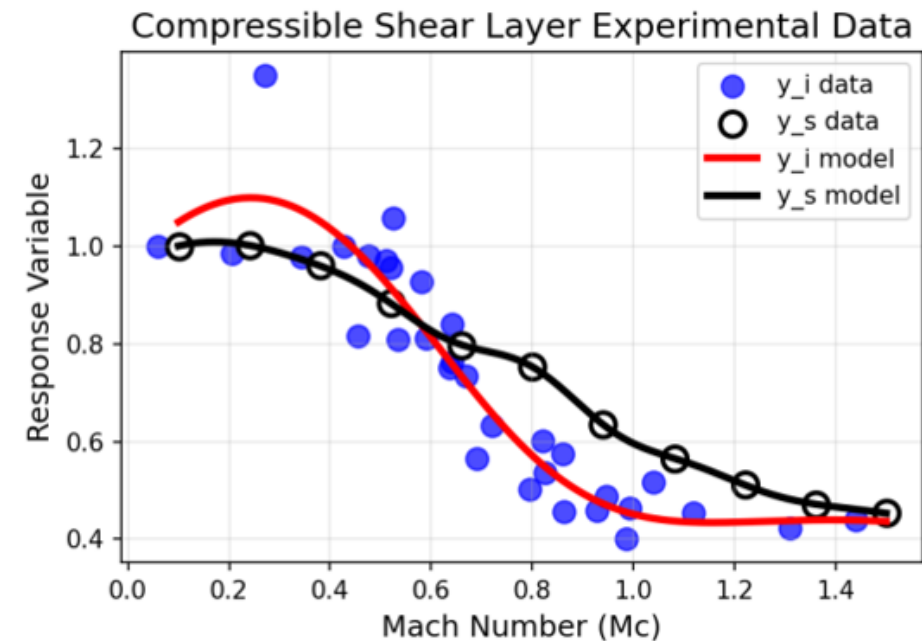
# Validation of computer models



TEXAS A&M UNIVERSITY

Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

- Example from (Oberkampf & Barone 2006).
- A free shear layer is formed when two streams with distinct velocities and temperatures separated by a thin splitter plate mix downstream of the splitter plate trailing edge.
- The data consists of a computational simulation dataset  $y_s$ , with 11 observations, and dataset  $y_i$ , with 32 physical observations.
- Input: the convective Mach number  $M_c$
- Output: Compressibility factor  $y$





## Some high-level questions



TEXAS A&M UNIVERSITY

Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

- Why a frequentist method?
  - Frequentist methods are generally perceived as more objective.
- Why subdomain tests?
  - It is possible that the computer model is accurate in some input areas but not as accurate elsewhere.
  - Learning computer model's region of inaccuracy can help understand the model's limitation and guide further improvements.



- In this work, our goal is to validate a known deterministic function

$$y^s: \Omega \rightarrow \mathbb{R},$$

with  $\Omega \subset \mathbb{R}^d$ , which can be either the computer model or a surrogate model. We will call  $y^s$  the computer model for simplicity.

- 

- The physical data follows the non-parametric model

$$y_i = f(x_i) + e_i$$

- $f$  is called the physical process.
- $x_i$ 's are i.i.d. random variables with probability density function  $p(x)$ .
- $e_i$ 's are i.i.d. random variables with mean zero and variance  $\sigma^2$ .



- **Statistical validation:** provide a  $p$ -value with statistical guarantee.
- Global validation: **validate whether  $y^s(x) = f(x)$  for all  $x \in \Omega$ .**
  - Some existing statistical methods can be adapted to conduct a test.
- Subdomain validation: **validate whether  $y^s(x) = f(x)$  for  $x \in \Omega_i \subset \Omega$ .**
  - No existing work in the original format.
  - Can be useful when there is a subdomain of particular interest, or we want to understand the reliable input domain of the computer model.



- Global validation

- $$H_0: f(x) = y^s(x) \text{ for all } x \in \Omega \leftrightarrow H_1: f(x) \neq y^s(x) \text{ for some } x \in \Omega.$$

- Subdomain validation is formulated as a multiple testing problem.

- Consider a partition  $\Omega = \bigcup_{i=1}^N \Omega_i$ .
- The goal is to control the individual error rate (IER) or the family-wise error rate (FWER) for the following set of hypotheses:

- $$H_{0i}: f(x) = y^s(x) \text{ for all } x \in \Omega_i \leftrightarrow H_{1i}: f(x) \neq y^s(x) \text{ for some } x \in \Omega_i.$$

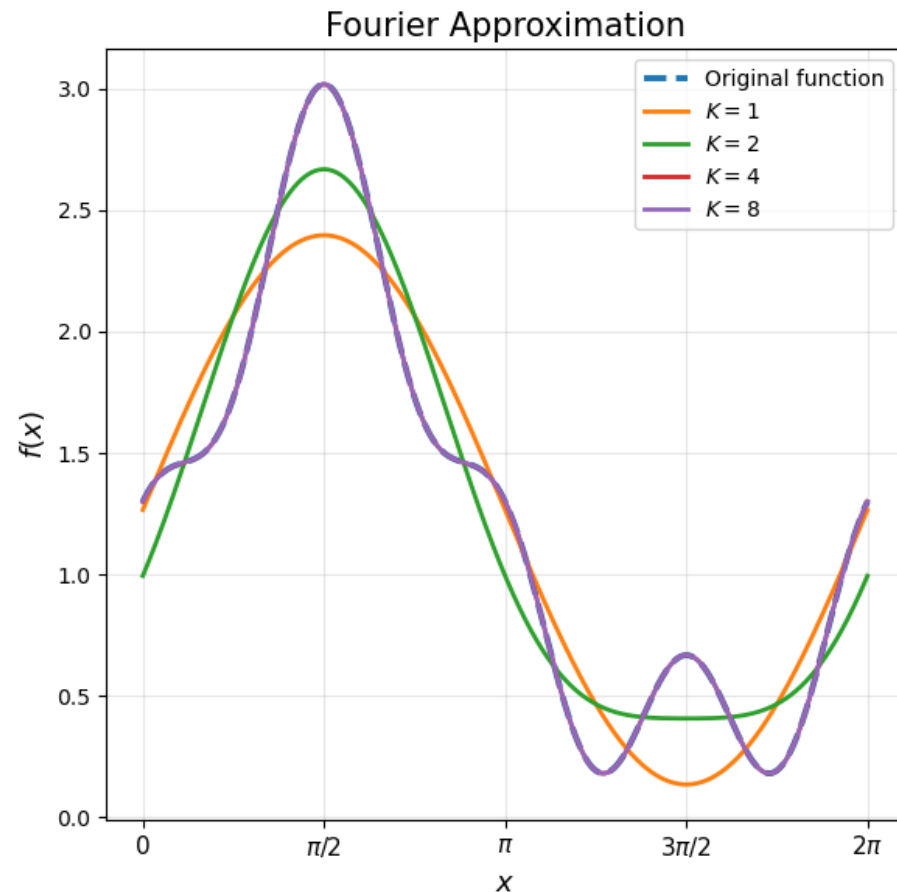
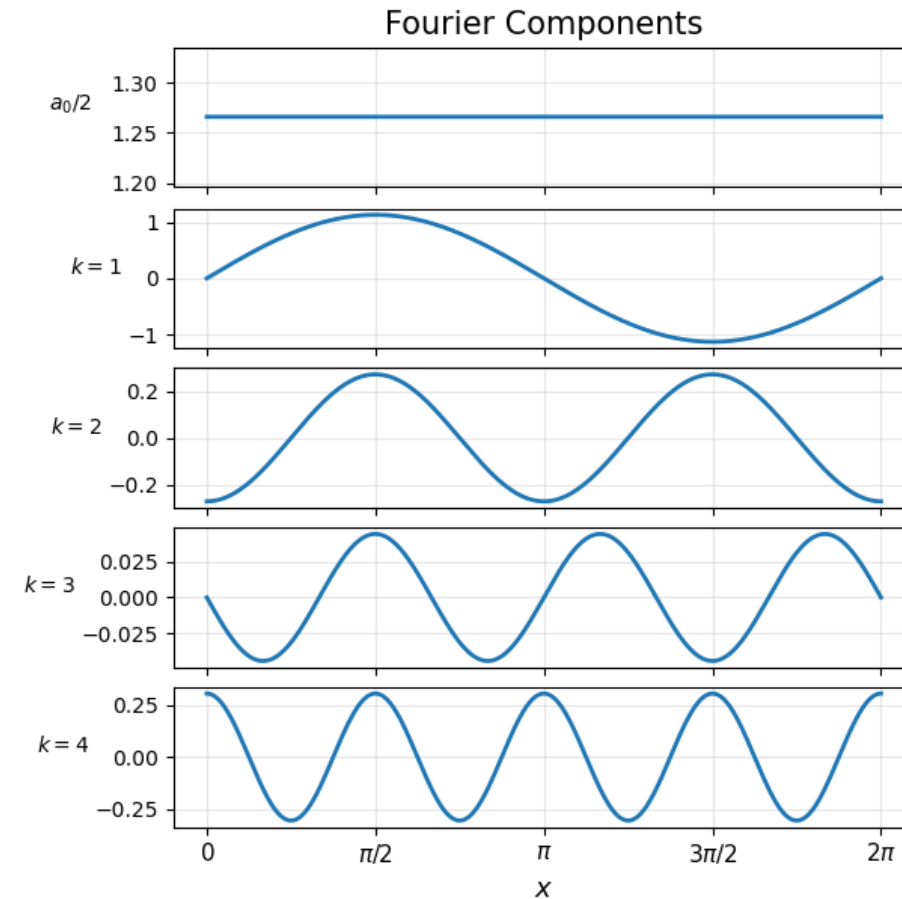


TEXAS A&M UNIVERSITY

Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

# Fourier expansion of a function

- We propose a novel statistical inference framework using the (generalized) Fourier coefficients.





## Proposed method: High level ideas

- Let  $\{h_j\}$  be an orthonormal basis of  $L_2(\Omega)$ . Then each  $g \in L_2(\Omega)$  admits

$$g = \sum_{j=1}^{\infty} g_j h_j,$$

where  $g_j = \int g(x) h_j(x) dx$  is called a generalized Fourier coefficient.

- **Fact:**  $g$  is zero if and only if all  $g_j$ 's are zero.

- Let  $\delta = f - y^s$  be the discrepancy process.

- Consider the generalized Fourier coefficients of  $\delta\sqrt{p}$ :

$$s^{(j)} := \int \delta(x) \sqrt{p(x)} h_j(x) dx.$$

- **We hope to:** Test  $s^{(j)} = 0$  for all  $j$  simultaneous.



# Nonparametric reconstruction with KRR

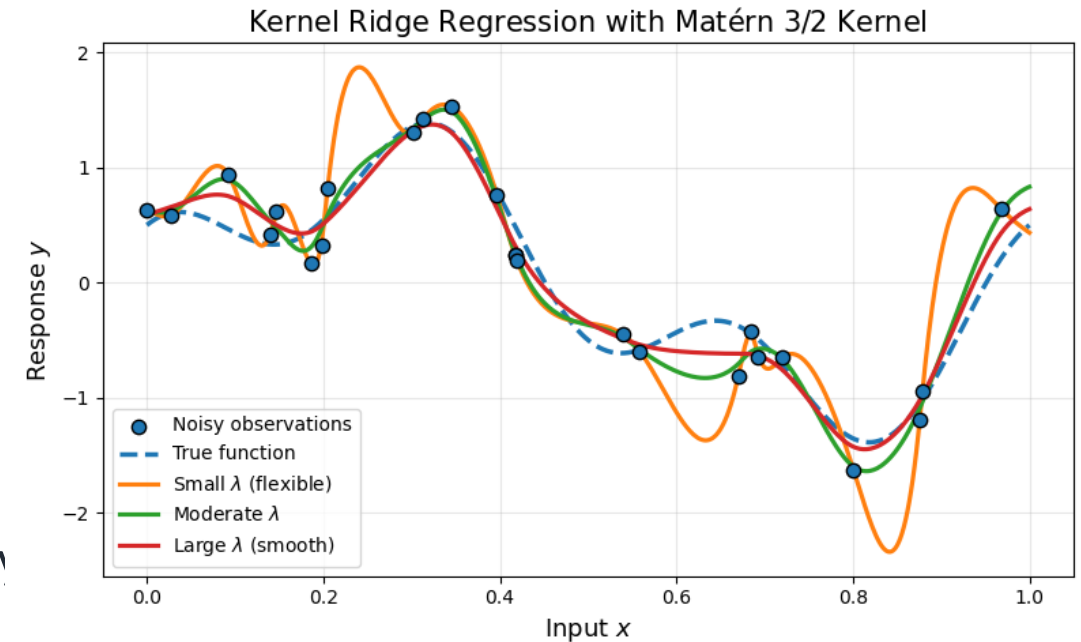
TEXAS A&M UNIVERSITY

Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

- We first reconstruct the discrepancy  $\delta = f - y^s$ .
- 
- Let  $z_i = y_i - y^s(x_i)$ . Use kernel ridge regression

$$\hat{\delta}_n := \operatorname{argmin}_{v \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^n (z_i - v(x_i))^2 + \lambda_n \|v\|_{\mathcal{H}}^2.$$

- $\mathcal{H}$ : the reproducing kernel Hilbert space generated by an isotropic Matérn kernel.
- $\lambda_n$ : a tuning parameter to balance the bias and the variance.
- Solution coincides with Gaussian process regression.





- For now, let us assume that the input density  $p$  is known. **This condition will be relaxed later.**

- We estimate  $s^{(i)}$  via a plug-in method

$$\hat{s}_n^{(j)} = \int \hat{\delta}_n(x) \sqrt{p(x)} h_j(x) dx .$$

- A known asymptotic result: for each  $j$ ,  $\sqrt{n} \left( \hat{s}_n^{(j)} - s^{(j)} \right)$  is asymptotically equivalent to  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{e_i h_j(x_i)}{\sqrt{p(x_i)}}$ .
  - A nice property: the leading terms for different  $j$ 's are uncorrelated.



- Therefore,  $\sqrt{n} \left( \hat{s}_n^{(j)} - s^{(j)} \right) \rightsquigarrow N(0, \sigma^2)$  for each  $j$ .
  - $\sigma^2$  is the variance of the random noise in the physical data, which can be consistently estimated from the residual of KRR.
  - This is a semi-parametric result: the  $n^{-1/2}$ -rate is attained despite the slower nonparametric convergence rate for  $\hat{\delta}_n - \delta$ .
- This result allows for a Wald-type test for  $H_0^{(j)}: s^{(j)} = 0 \leftrightarrow H_1^{(j)}: s^{(j)} \neq 0$ .
  - The asymptotic p-value is
$$2 \left( 1 - \Phi \left( \frac{\sqrt{n} \hat{s}_n^{(j)}}{\hat{\sigma}_n} \right) \right).$$



## A simultaneous convergence result

- Theorem 1 enables simultaneous inference for all  $s^{(j)}$ , on the premise that a sequence of decay coefficients  $\rho_j$  is involved.
- **Theorem 1.** Under regularity conditions (e.g.,  $\rho_j = o(\log^{-s}(j+1))$  for  $s > 1/2$ ), we have

$$\begin{aligned} & \sqrt{n} \left( \rho_1 \left( \hat{s}_n^{(1)} - s^{(1)} \right), \rho_2 \left( \hat{s}_n^{(2)} - s^{(2)} \right), \dots \right) \\ & \rightsquigarrow N \left( 0, \text{diag}(\rho_1^2 \sigma^2, \rho_2^2 \sigma^2, \dots) \right). \end{aligned}$$



- In light of Theorem 1, we propose the following testing method.

- Maximum modulus test statistic  $T = \max_{j \geq 1} \left| \frac{\sqrt{n} \rho_j \hat{s}_n^{(j)}}{\hat{\sigma}_n} \right|.$

- The asymptotic  $p$ -value is  $1 - \prod_{j=1}^{\infty} (2\Phi(T/\rho_j) - 1).$

- Under the global alternative hypothesis,  $T$  diverges at a  $\sqrt{n}$ -rate.



- When  $p$  is unknown, we may replace it by a density estimator  $\hat{p}_n$ .

- A natural estimator of  $s^{(j)}$  would be

$$\tilde{s}_n^{(j)} := \int \hat{\delta}_n(x) \sqrt{\hat{p}_n(x)} h_j(x) dx .$$

- However, semi-parametric theory fails to work this time. Due to the slow convergence rate of the density estimator,  $\tilde{s}_n^{(j)}$  does not converge to  $s^{(j)}$  at a  $n^{-1/2}$ -rate.



- Define

$$d_j = \int [\hat{\delta}_n(x) - \delta(x)] \sqrt{\hat{p}(x)} h_j(x) dx .$$

- **Theorem 2.** Under regularity conditions, we have

$$\sqrt{n}(\rho_1 d_1, \rho_2 d_2, \dots) \rightsquigarrow N \left( 0, \text{diag}(\rho_1^2 \sigma^2, \rho_2^2 \sigma^2, \dots) \right) .$$

- Theorem 2 justifies the testing method on the previous slide with estimated density.
  - The test is consistent: under  $H_1$ , the test statistic diverges at a  $\sqrt{n}$ -rate.



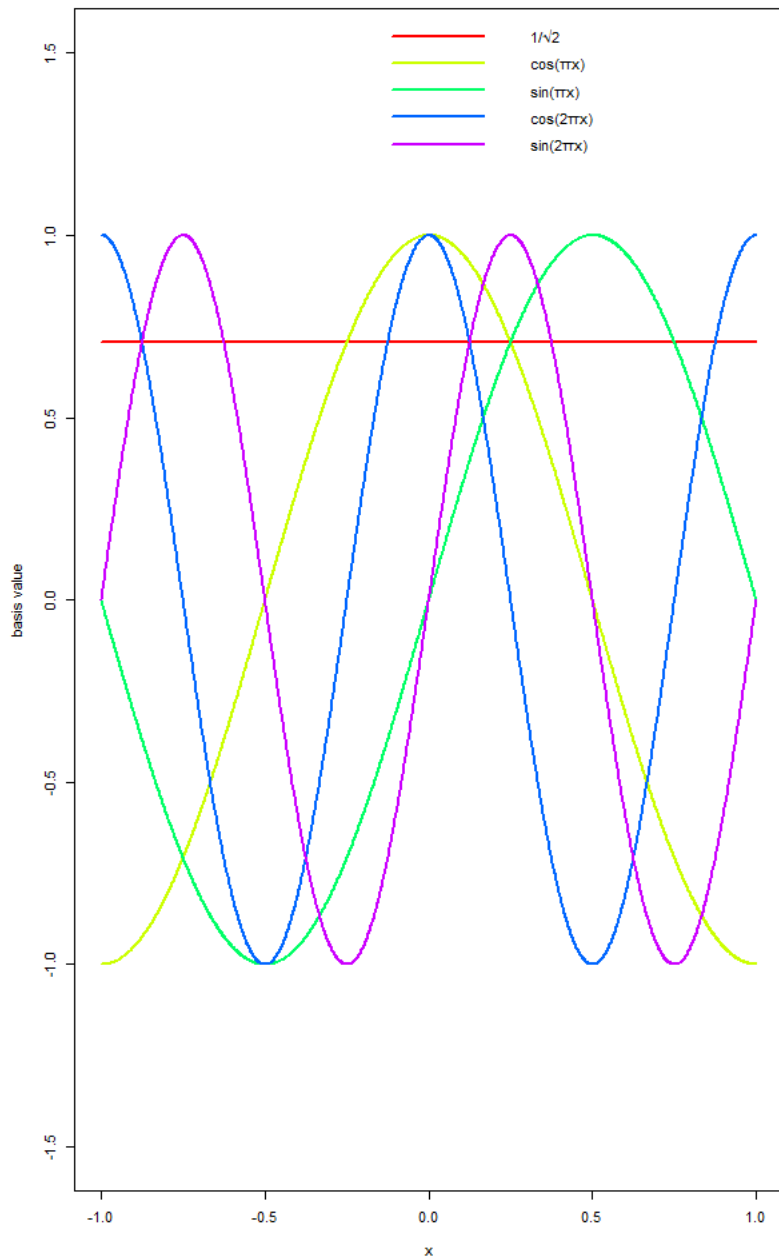
TEXAS A&M UNIVERSITY

Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

## Proposed method: Subdomain test and $p$ -value

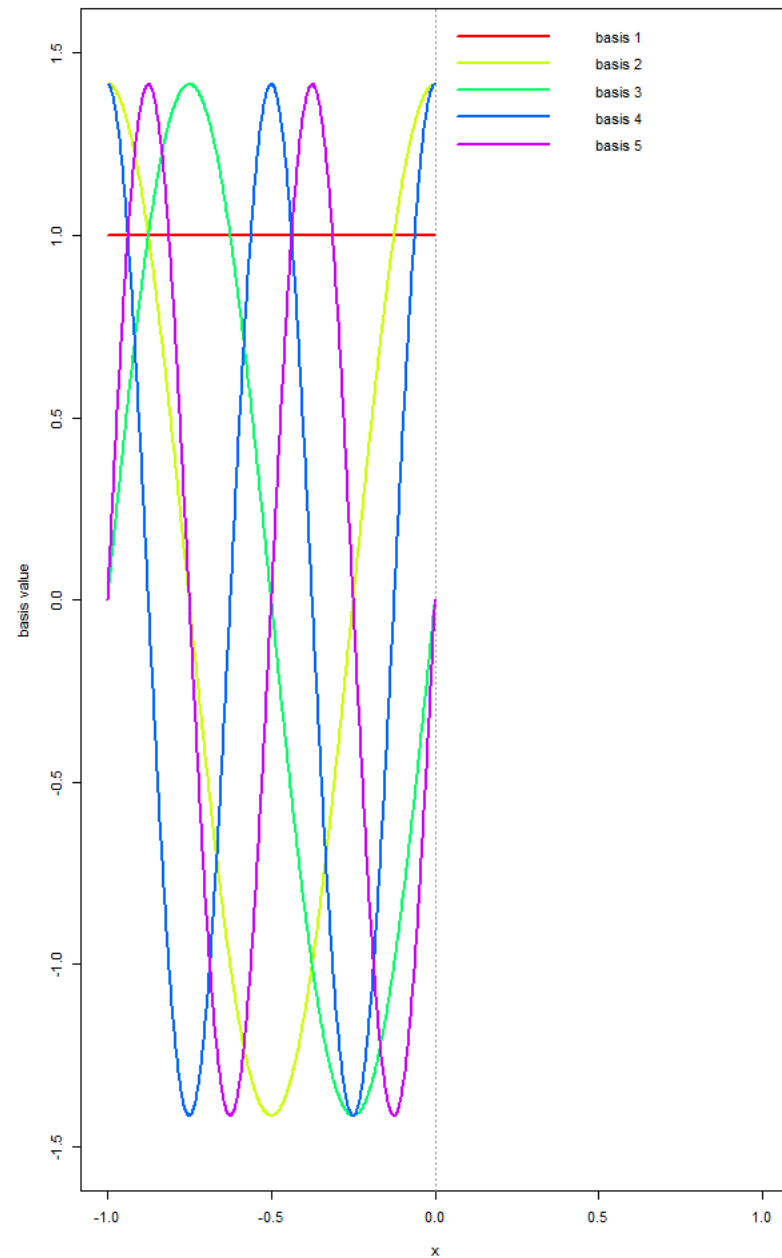
- For the subdomain problem, we pick an orthonormal basis  $\{h_{ij}\}_{j=1}^{\infty}$  for each  $\Omega_j$ .
- Regard each  $h_{ij}$  as a local function supported within  $\Omega_i$ .
- Test statistic  $T_i = \frac{\sqrt{n}}{\hat{\sigma}_n} \max_{j \geq 1} \left| \rho_j \int (\hat{f}_n - y^s)(x) \sqrt{\hat{p}_n(x)} h_{ij}(x) dx \right|$ .
- Asymptotic  $p$ -value under individual error rate (IER):  
 $1 - \prod_{j=1}^{\infty} (2\Phi(T_i/\rho_j) - 1)$ .
- Use the Bonferroni correction to control the family-wise error rate (FWER).

Orthonormal basis on [-1,1]



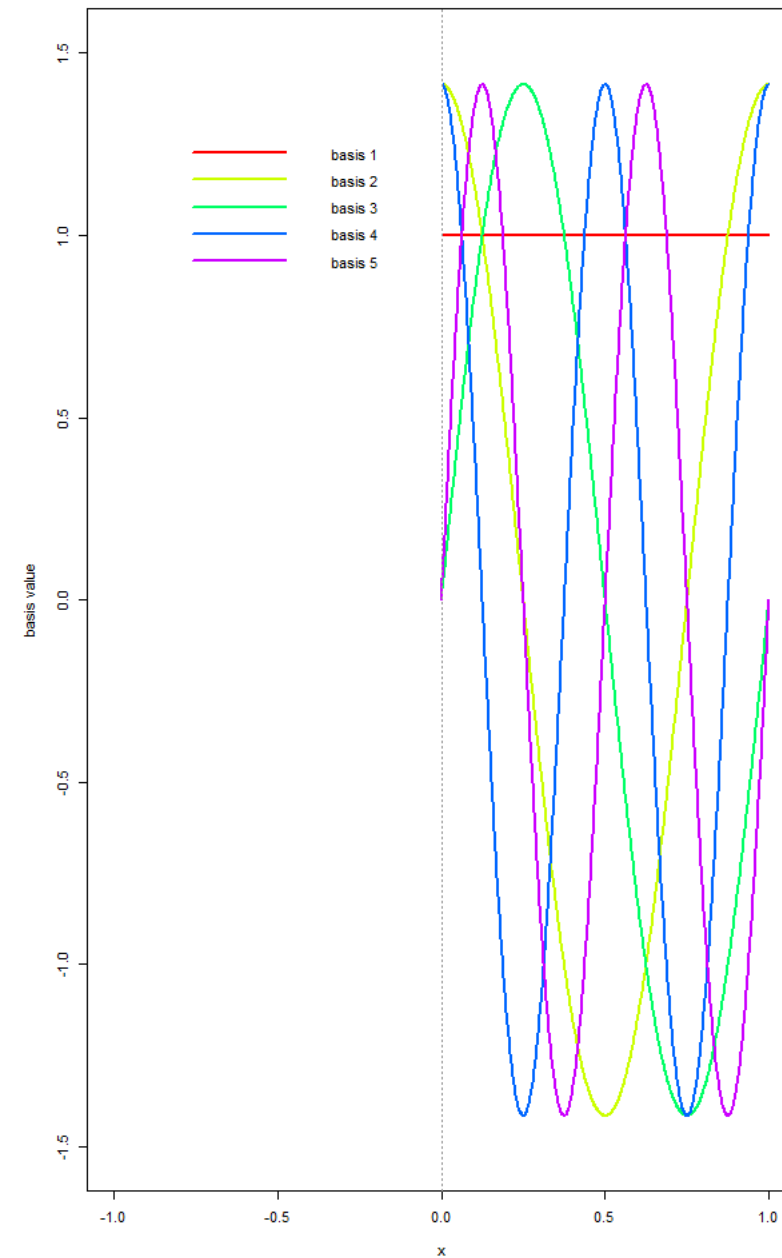
10/22/2026

Orthonormal basis on [-1,0]



UQSay #102

Orthonormal basis on [0,1]





## Comparison with benchmark: Hypotheses

- We compare the proposed method with the established nonparametric test of [Neumeyer & Dette \(2003\)](#). Examine the rejection probabilities of the tests based on our global test and their statistics  $K^{(1)}$  and  $K^{(2)}$ .
- We consider six pairs of hypotheses. In each test:  $H_0: f_1(x) = f_2(x)$ ;  $H_1: f_2(x) = f_1(x) + \text{perturbation}$ .

| Name                        | $f_1(x)$       | $f_2(x)$              | Perturbation   |
|-----------------------------|----------------|-----------------------|----------------|
| Constant vs Linear          | 1              | $1 + x$               | $x$            |
| Exp vs Exp + Linear         | $e^x$          | $e^x + x$             | $x$            |
| Sine vs Sine + Linear       | $\sin(2\pi x)$ | $2 \sin(2\pi x) + x$  | $x$            |
| Constant vs Constant + Sine | 1              | $1 + 2\sin(2\pi x)$   | $\sin(2\pi x)$ |
| Exp vs Exp + Sine           | $e^x$          | $e^x + 2\sin(2\pi x)$ | $\sin(2\pi x)$ |
| Sine vs Double Sine         | $\sin(2\pi x)$ | $2\sin(2\pi x)$       | $\sin(2\pi x)$ |



TEXAS A&M UNIVERSITY

Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

## Comparison with benchmark: Other details

- For each hypothesis scenario, rejection probabilities are estimated from 1000 simulation trials at sample sizes  $n \in \{25, 50\}$  and significance levels  $\alpha \in \{0.025, 0.05, 0.1\}$ .
- In each trial, the design points  $\{x_i\}_{i=1}^n$  are sampled i.i.d from a truncated normal distribution with mean  $\mu = 0.5$ , standard deviation  $\sigma_0 = 0.2$ , and restricted to the interval  $[0,1]$ .
- The function values  $f_1(x)$  and  $f_2(x)$  are corrupted by additive noise, where  $e_1 \equiv 0$  and  $e_2 \sim \mathcal{N}(0, 0.5^2)$ , sampled independently.
- We use shaded cells to highlight the values closest to  $\alpha$  without exceeding it under  $H_0$ , and closest to 1 under  $H_1$ .



## Comparison with benchmark: Results I

- When the perturbation is linear, our test statistical power is similar to the benchmark and slight better in most cases.

| Scenario            | $n$ | $\alpha$ | Type I Error ( $H_0$ ) |           |             | Power ( $H_1$ ) |           |             |
|---------------------|-----|----------|------------------------|-----------|-------------|-----------------|-----------|-------------|
|                     |     |          | $K^{(1)}$              | $K^{(2)}$ | Global Test | $K^{(1)}$       | $K^{(2)}$ | Global Test |
| Constant Vs Linear  | 25  | 0.025    | 0.027                  | 0.038     | 0.020       | 0.949           | 0.948     | 0.976       |
|                     | 25  | 0.050    | 0.055                  | 0.067     | 0.040       | 0.975           | 0.975     | 0.989       |
|                     | 25  | 0.100    | 0.114                  | 0.133     | 0.074       | 0.988           | 0.993     | 0.998       |
|                     | 50  | 0.025    | 0.029                  | 0.045     | 0.014       | 0.998           | 1.000     | 1.000       |
|                     | 50  | 0.050    | 0.060                  | 0.077     | 0.035       | 0.999           | 1.000     | 1.000       |
|                     | 50  | 0.100    | 0.120                  | 0.131     | 0.060       | 0.999           | 1.000     | 1.000       |
| Exp Vs Exp Linear   | 25  | 0.025    | 0.031                  | 0.030     | 0.023       | 0.936           | 0.956     | 0.960       |
|                     | 25  | 0.050    | 0.067                  | 0.062     | 0.037       | 0.968           | 0.973     | 0.981       |
|                     | 25  | 0.100    | 0.122                  | 0.123     | 0.078       | 0.987           | 0.987     | 0.994       |
|                     | 50  | 0.025    | 0.038                  | 0.025     | 0.025       | 1.000           | 0.999     | 1.000       |
|                     | 50  | 0.050    | 0.067                  | 0.057     | 0.034       | 1.000           | 1.000     | 1.000       |
|                     | 50  | 0.100    | 0.122                  | 0.124     | 0.075       | 1.000           | 1.000     | 1.000       |
| Sine Vs Sine Linear | 25  | 0.025    | 0.026                  | 0.023     | 0.022       | 0.889           | 0.903     | 0.952       |
|                     | 25  | 0.050    | 0.051                  | 0.040     | 0.044       | 0.950           | 0.948     | 0.974       |
|                     | 25  | 0.100    | 0.090                  | 0.074     | 0.075       | 0.976           | 0.977     | 0.989       |
|                     | 50  | 0.025    | 0.018                  | 0.013     | 0.019       | 0.998           | 0.995     | 1.000       |
|                     | 50  | 0.050    | 0.043                  | 0.029     | 0.042       | 1.000           | 0.999     | 1.000       |
|                     | 50  | 0.100    | 0.086                  | 0.070     | 0.096       | 1.000           | 1.000     | 1.000       |



## Comparison with benchmark: Results II

- When the perturbation is sine, our test shows significant higher statistical power than the benchmark.

| Scenario            | $n$ | $\alpha$ | Type I Error ( $H_0$ ) |           |             | Power ( $H_1$ ) |           |             |
|---------------------|-----|----------|------------------------|-----------|-------------|-----------------|-----------|-------------|
|                     |     |          | $K^{(1)}$              | $K^{(2)}$ | Global Test | $K^{(1)}$       | $K^{(2)}$ | Global Test |
| Constant Vs Sine    | 25  | 0.025    | 0.045                  | 0.030     | 0.028       | 0.422           | 0.312     | 0.983       |
|                     | 25  | 0.050    | 0.069                  | 0.073     | 0.048       | 0.580           | 0.547     | 0.994       |
|                     | 25  | 0.100    | 0.138                  | 0.131     | 0.087       | 0.750           | 0.759     | 0.996       |
|                     | 50  | 0.025    | 0.037                  | 0.038     | 0.022       | 0.747           | 0.863     | 1.000       |
|                     | 50  | 0.050    | 0.063                  | 0.074     | 0.037       | 0.869           | 0.947     | 1.000       |
|                     | 50  | 0.100    | 0.120                  | 0.139     | 0.079       | 0.955           | 0.983     | 1.000       |
| Exp Vs Exp Sine     | 25  | 0.025    | 0.032                  | 0.029     | 0.015       | 0.385           | 0.305     | 0.990       |
|                     | 25  | 0.050    | 0.048                  | 0.072     | 0.037       | 0.519           | 0.513     | 0.995       |
|                     | 25  | 0.100    | 0.101                  | 0.137     | 0.077       | 0.684           | 0.749     | 0.998       |
|                     | 50  | 0.025    | 0.024                  | 0.018     | 0.022       | 0.726           | 0.839     | 1.000       |
|                     | 50  | 0.050    | 0.042                  | 0.053     | 0.047       | 0.867           | 0.944     | 1.000       |
|                     | 50  | 0.100    | 0.084                  | 0.123     | 0.081       | 0.956           | 0.989     | 1.000       |
| Sine Vs Double Sine | 25  | 0.025    | 0.019                  | 0.014     | 0.028       | 0.261           | 0.154     | 0.979       |
|                     | 25  | 0.050    | 0.034                  | 0.031     | 0.049       | 0.383           | 0.306     | 0.989       |
|                     | 25  | 0.100    | 0.080                  | 0.084     | 0.090       | 0.563           | 0.532     | 0.995       |
|                     | 50  | 0.025    | 0.022                  | 0.019     | 0.019       | 0.606           | 0.718     | 1.000       |
|                     | 50  | 0.050    | 0.033                  | 0.040     | 0.045       | 0.756           | 0.862     | 1.000       |
|                     | 50  | 0.100    | 0.074                  | 0.083     | 0.094       | 0.899           | 0.953     | 1.000       |



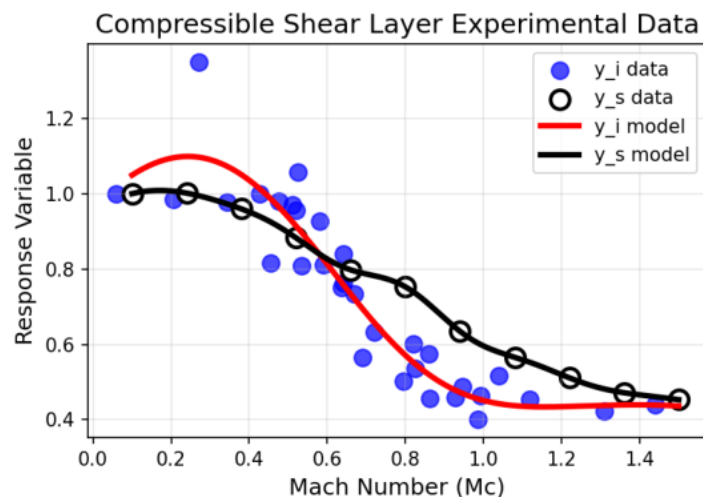
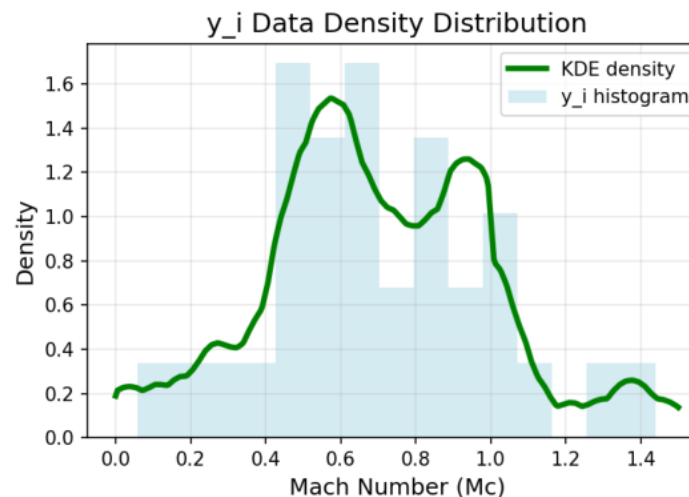
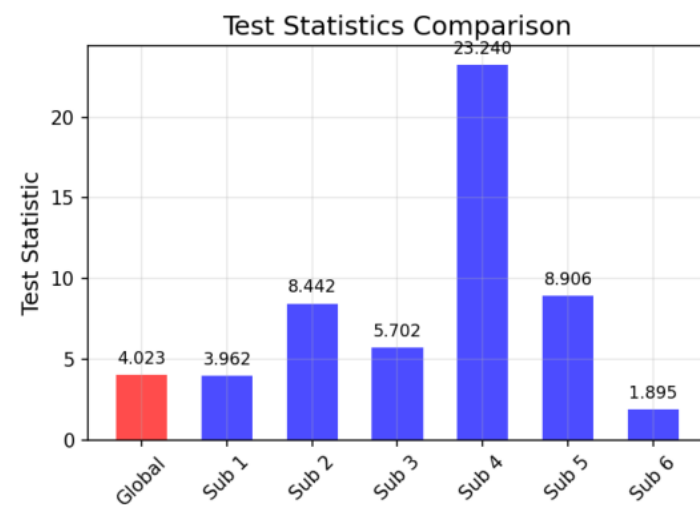
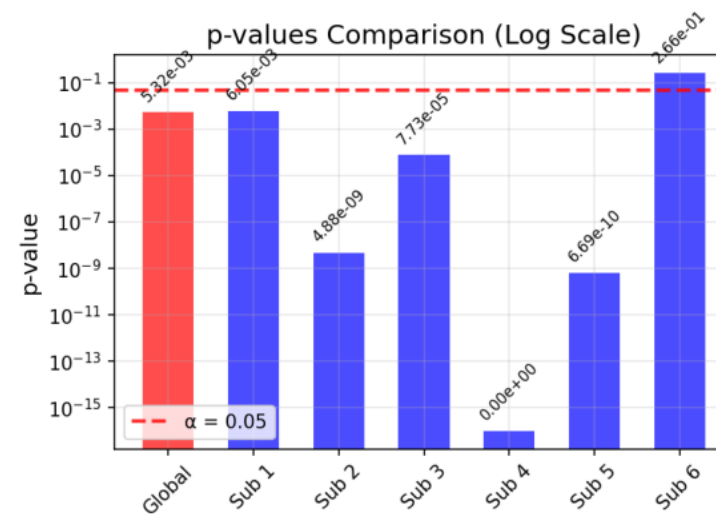
- The shear layer experiment investigates the relationship between the compressibility factor  $\Phi$  and the convective Mach number  $M_c$ .
- [Oberkampf & Barone \(2006\)](#) proposed a validation metric and analyzed the data using confidence interval techniques.
- Subsequently, [Wang, Chen, & Tsui \(2009\)](#) applied a Bayesian approach based on Gaussian process modeling to further analyze the same dataset.
- We apply our testing procedure both globally over the full domain  $[0, 1.5]$  and locally over six contiguous subdomains of length 0.25, resulting in one global test and six subdomain tests.
- Additionally, we compute the corresponding  $p$ -value over the full domain using the test statistic introduced by [Neumeyer & Dette \(2003\)](#).



# Shear layer experiment: Result

TEXAS A&M UNIVERSITY

Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering





TEXAS A&M UNIVERSITY

Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

## Shear layer experiment: Comparison

- Global-level comparison ( $\alpha = 0.05$ )

| Our Test       | Neumeyer & Dette $K^{(1)}$ | Neumeyer & Dette $K^{(2)}$ | Oberkampf & Barone (2006) | Wang et al. (2012) |
|----------------|----------------------------|----------------------------|---------------------------|--------------------|
| Reject (0.005) | Accept (0.165)             | Accept (0.440)             | Reject                    | Reject             |

- Subdomain-level comparison ( $\alpha = 0.05$ )

| Subdomain ( $M_c$ range)       | Our Test     | Wang et al. (2009)                      | Oberkampf & Barone (2006) |
|--------------------------------|--------------|-----------------------------------------|---------------------------|
| [0.5, 0.75]<br>(includes 0.65) | Reject $H_0$ | Accept at 0.65                          | Accept at 0.65            |
| [0.75, 1.0]<br>(includes 0.8)  | Reject $H_0$ | Reject at 0.8                           | Reject at 0.8             |
| [1.0, 1.25]<br>(includes 1.2)  | Reject $H_0$ | Result uncertain<br>(insufficient data) | Reject at 1.2             |



## Validation after model calibration



TEXAS A&M UNIVERSITY

Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

- Calibration is the process to adjust parameters in the computer model so that the computer output matches the physical process better.
- If the computer model is calibrated using physical data, the uncertainty of the physical data should be reflected in the limit distribution of the physical process.
- Here we only consider a simple case: the same dataset is used for both calibration and validation.
- Idea: We still use a similar test statistic, but pursue its new limit distribution



- We consider the  $L_2$  calibration framework (Tuo & Wu 2014), with the goal of estimating

$$\theta^* := \operatorname{argmin} \|f - y^s(\cdot, \theta)\|_{L_2}.$$

- The optimal estimator is

$$\hat{\theta} := \operatorname{argmin} \|\hat{f} - y^s(\cdot, \theta)\|_{L_2}.$$

- It has the asymptotic leading term

$$V^{-1} \sum_{i=1}^n e_i \frac{\partial y^s}{\partial \theta}(x_i, \theta^*),$$

with  $V := \mathbb{E} \left[ \frac{\partial^2}{\partial \theta \partial \theta^T} (f(x_1) - y^s(x_1, \theta^*))^2 \right]$  being the Fisher information matrix.



TEXAS A&M UNIVERSITY

Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

## Test statistics and asymptotic distribution

- Test statistic

$$\begin{aligned}\hat{s}_n^{(j)} &:= \int \left( \hat{f}(x) - y^s(x, \hat{\theta}) \right) \sqrt{\hat{p}_n(x)} h_j(x) dx \\ &= \int \left( \hat{f}(x) - y^s(x, \theta^*) \right) \sqrt{\hat{p}_n(x)} h_j(x) dx + \\ &\quad \int \left( y^s(x, \theta^*) - y^s(x, \hat{\theta}) \right) \sqrt{\hat{p}_n(x)} h_j(x) dx.\end{aligned}$$

- Both terms are asymptotically normal, but they are not independent.
- Consequently,  $\sqrt{n}(\rho_1 d_1, \rho_2 d_2, \dots)$  still converges to a normal distribution, but its covariance is no longer diagonal.
- The idea is to perform Gram-Schmidt orthogonalization first, then use the maximum modulus statistic.



- Some may argue that the computer model does not have to give the same output as the reality. It is fit for use as long as the discrepancy is small enough. One possibility is to introduce a thresholding  $\Delta$ , and consider the test

$$H_0: \|\delta\| \leq \Delta \leftrightarrow H_1: \|\delta\| > \Delta,$$

or

$$H_0: \|\delta\| \geq \Delta \leftrightarrow H_1: \|\delta\| < \Delta.$$



## Conclusion



TEXAS A&M UNIVERSITY  
Wm Michael Barnes '64 Department of  
Industrial & Systems Engineering

- We propose a novel hypothesis testing methodology to validate a computer model globally or in a subdomain of the input space.
- Preprint available at [arXiv:2604.17144](https://arxiv.org/abs/2604.17144). A revised version will be released soon.

• *Thank you for attending the talk!*

Email: [ruituo@tamu.edu](mailto:ruituo@tamu.edu)