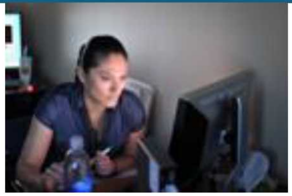
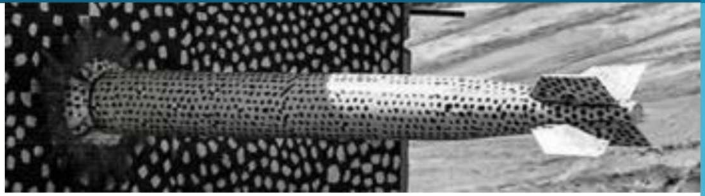
SAND2020-5082C

Full Function Sampling of Uncertain Correlations





ASME Verification and Validation Symposium 2020
May 20-22, 2020, Online
VVS2020-8834

Kevin W. Irick, Jeff Engerer, Blake Lance,
Scott A. Roberts, & Ben Schroeder

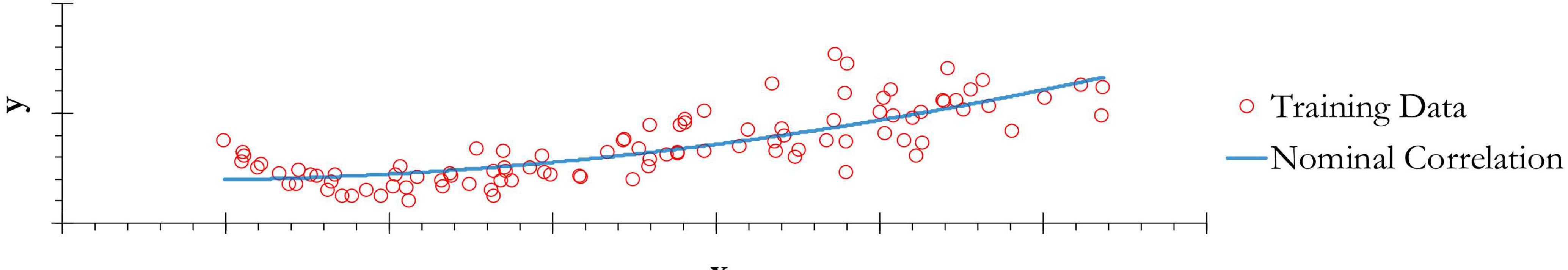
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Problem

Given set of training data
- Empirical or otherwise

Try to succinctly describe data by some correlation (or fit, model, equation, etc.)
- Does not perfectly describe data due to "model" form error and/or variation in data
- Want to ***use*** the correlation for subsequent calculations

How can we accurately & effectively account for uncertainty in the correlation's representation of the underlying data when using the correlation?



Example Training Data with Nominal Correlation

Introduction

Prediction Interval

- Range of values within which a future value is expected to fall (at some confidence level, e.g. 95% confidence)

F-test

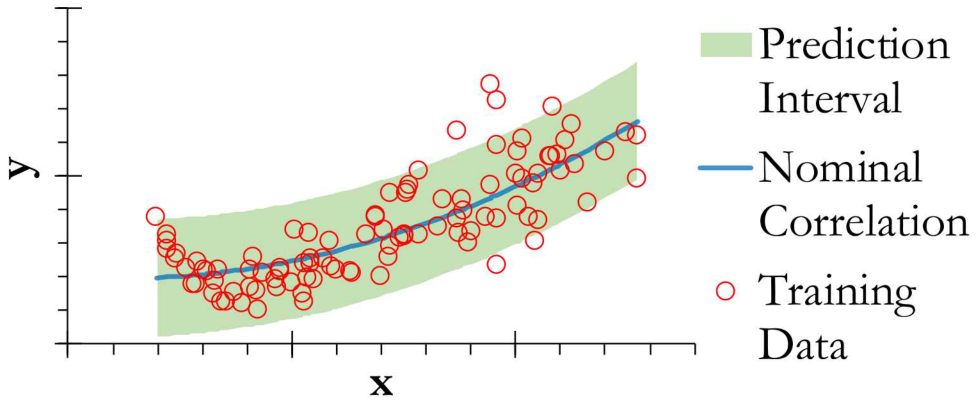
- Metric for statistical comparison of sample variances

Different approaches to implementing correlation uncertainty in applications

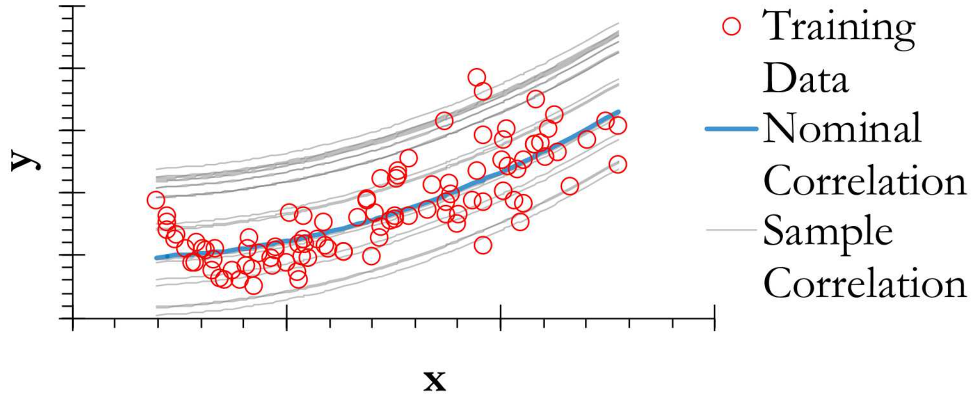
- Eyeballing
- Constant multiplier
- Constant offset
- Coefficient sampling
- These aren't necessarily problems…until they are

Goal

- Sample a random realization of the uncertain correlation
  - Respect form (e.g., linear, logarithmic, sinusoidal, etc.) of nominal correlation
  - Respect statistical trends in data without bias
- Sample the entire correlation holistically
  - Perhaps more "abstractly" than typical approaches



Prediction Interval Example

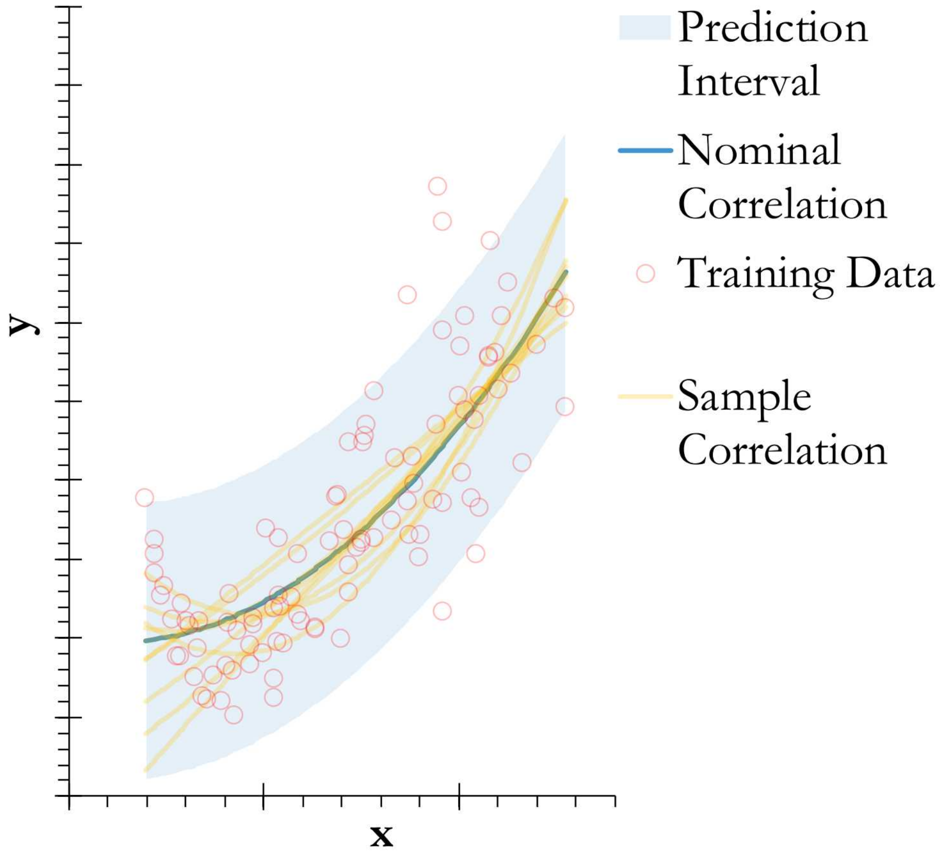


Example Uncertainty Implementation with Bias

Procedure Overview

1. Gather training data
2. Define correlation form
3. Obtain nominal correlation
4. Determine prediction interval on nominal correlation
5. Select random numbers
6. Determine input values
7. Determine control response values
8. Determine target response
9. Find sample coefficients
10. Repeat steps 5 through 9 for multiple samples
11. Perform F-test

BEST EXPLAINED BY EXAMPLE



End Result of Sampling Procedure

Procedure Walkthrough with Example – Step 1

Gather Training Data

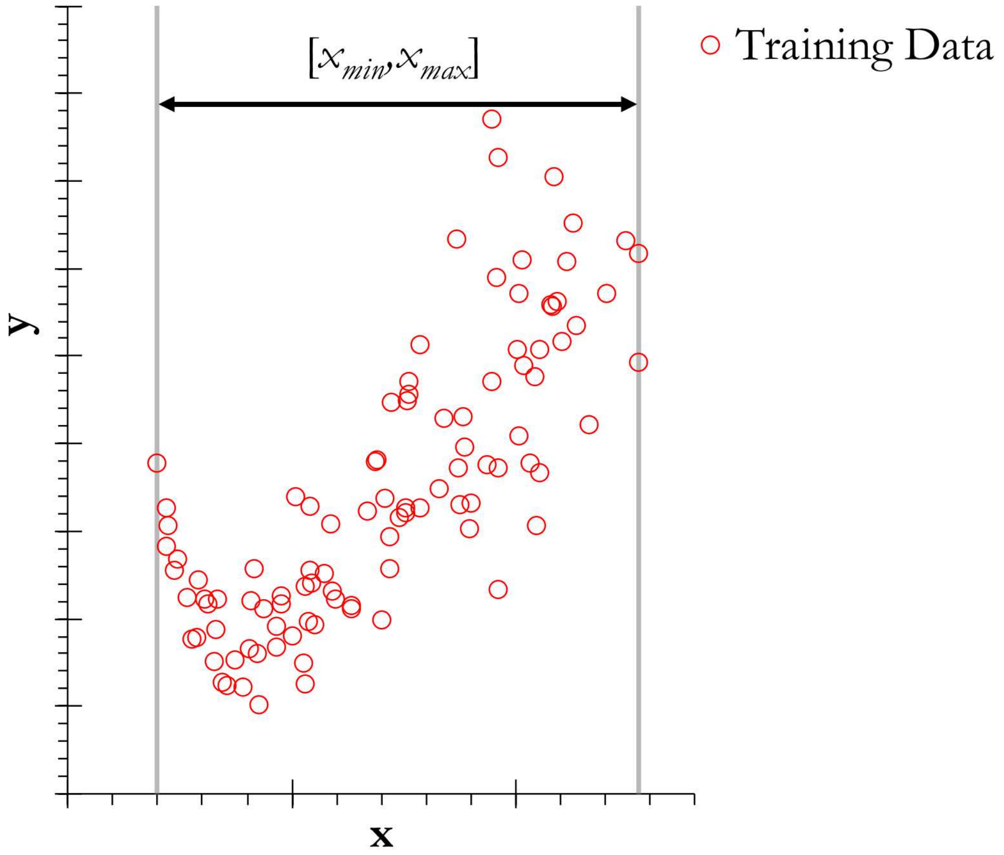
- Should span range of use where possible

N_{trn} training points

$\boldsymbol{x}_{trn}$ set of training independent variables

$\boldsymbol{y}_{trn}$ set of training dependent variables

$[x_{min}, y_{min}]$ application range



Training Data & Range

Procedure Walkthrough with Example – Steps 2,3

Define Correlation Form
- Selected parameterized function
- Does not need to be polynomial

$$y = f_{corr(\boldsymbol{a},x)}$$

y dependent variable
x independent variable
$\boldsymbol{a}$ set of correlation coefficients
N_a number of correlation coefficients
f_{corr} correlation function operator

Obtain Nominal Correlation
- Determine coefficients for best representation of underlying training data
- Use favorite fitting function (method)

$$f_{nom}(x) = f_{corr}(\boldsymbol{a}_{nom}, x)$$

$\boldsymbol{a}_{nom}$ set of nominal correlation coefficients

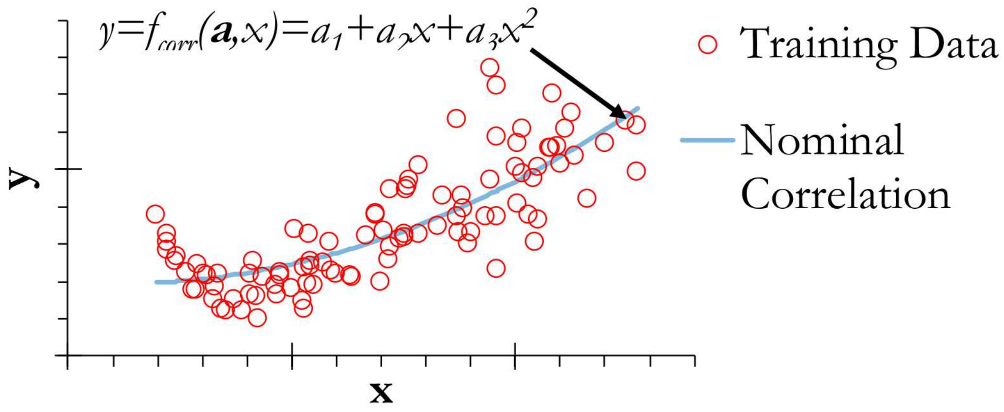
f_{nom} nominal correlation

Typical least squares objective function

$$f_o = \Sigma_{i=1}^{N_{trn}}\left(f_{nom}\left(x_{trn,i}\right) - y_{trn,i}\right)^2$$

Fitting method defined as function

$$f_{fit}(\boldsymbol{a}_{nom}, \boldsymbol{x}_{trn}, \boldsymbol{y}_{trn}) = \min_{\boldsymbol{a}_{nom}} f_o$$



Nominal Correlation Definition

Procedure Walkthrough with Example – Step 4

Determine Prediction Interval

- Describes possible range of future *y*-values about the nominal correlation (within statistical confidence level)
- Consider this as being a normal distribution of response probability about the nominal function
  - Where the true response is expected

$$y \in [f_{nom}(x) - PI(x,\alpha), f_{nom}(x) + PI(x,\alpha)]$$

PI prediction interval amplitude

a confidence level

$$PI(x,\alpha) = t(0.5(1+\alpha), N_{trn} - 2)m_{PI}(x)$$

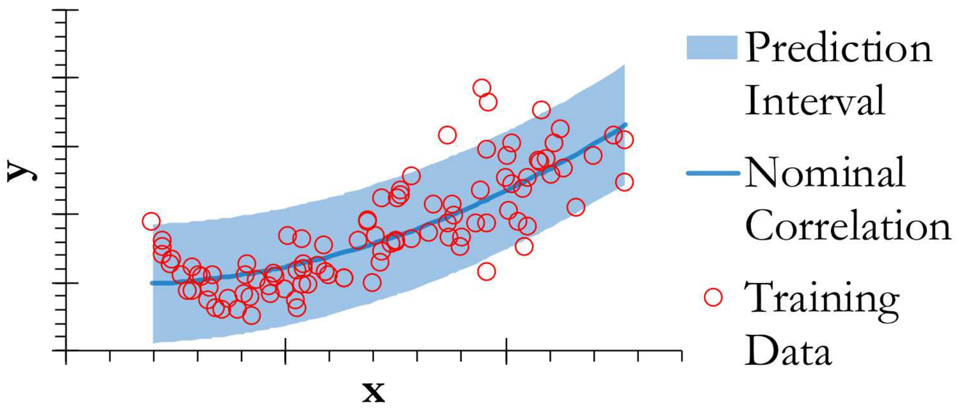
t Student's t-statistic

m_{PI} prediction interval deviation rate

$$m_{PI}(x)^2 = \left(\Sigma_{i=1}^{N_{trn}}\left(f_{nom}(x_{trn,i}) - y_{trn,i}\right)\right) * \cdots$$

$$\left(1 + \frac{1}{N_{trn}} + \frac{\left(x - x_{trn,avg}\right)^2}{\Sigma_{i=1}^{N_{trn}}\left(x_{trn,i} - x_{trn,avg}\right)^2}\right)$$

$x_{trn,avg}$ average of training independent variables



Nominal Correlation Prediction Interval

Procedure Walkthrough with Example – Steps 5,6

Select Random Number

r_1 random variable from U(0,1)

Determine Input Values

- N_a "control" points required to capture the nominal correlation form & shape
- 1 "target" point required to force the correlation to sample the uncertain space
- Points evenly distributed across range using random seed
  - Recommended to retain nominal correlation shape during uncertain space sampling

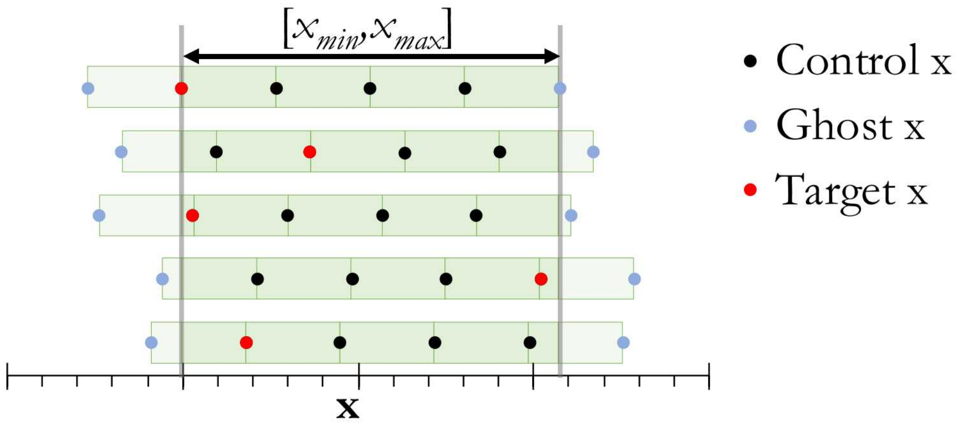
$$x_{tar} = x_{min} + r_1(x_{max} - x_{min})$$

x_{tar} target independent variable value

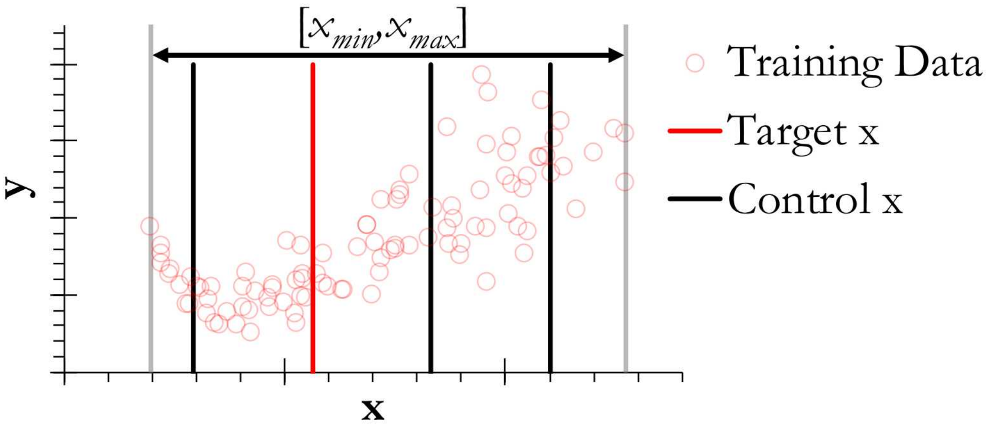
$$x_{ctrl,i} = x_{tar} + i\Delta x$$

$\boldsymbol{x}_{ctrl}$ control independent variable values

Δx cyclic independent variable value spacing



Random Independent Variable Space Division



Target and Control Input Value Selection

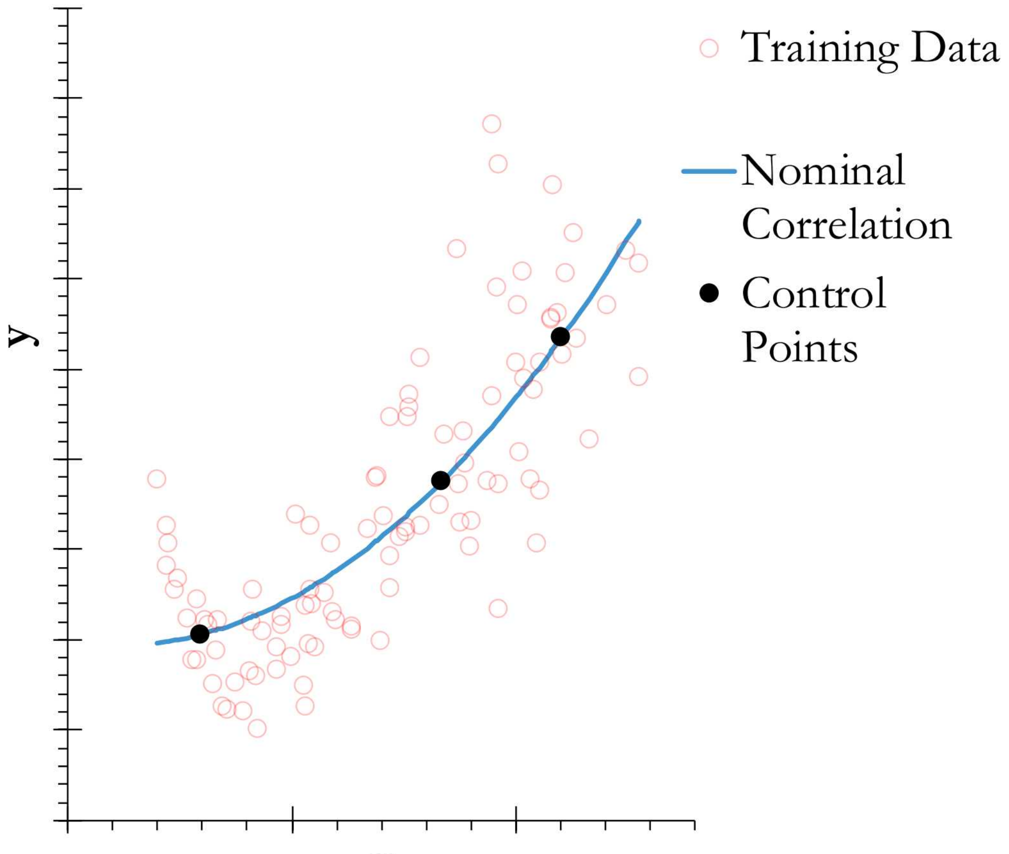
Procedure Walkthrough with Example – Step 7

Determine Control Response Values

- N_a "control" points required to capture the nominal correlation form & shape
- Evaluated from nominal correlation

$$y_{ctrl,i} = f_{nom}(x_{ctrl,i})$$

y_{ctrl} control dependent variable values



Control Response Points

Procedure Walkthrough with Example – Steps 5,8

Select Random Number

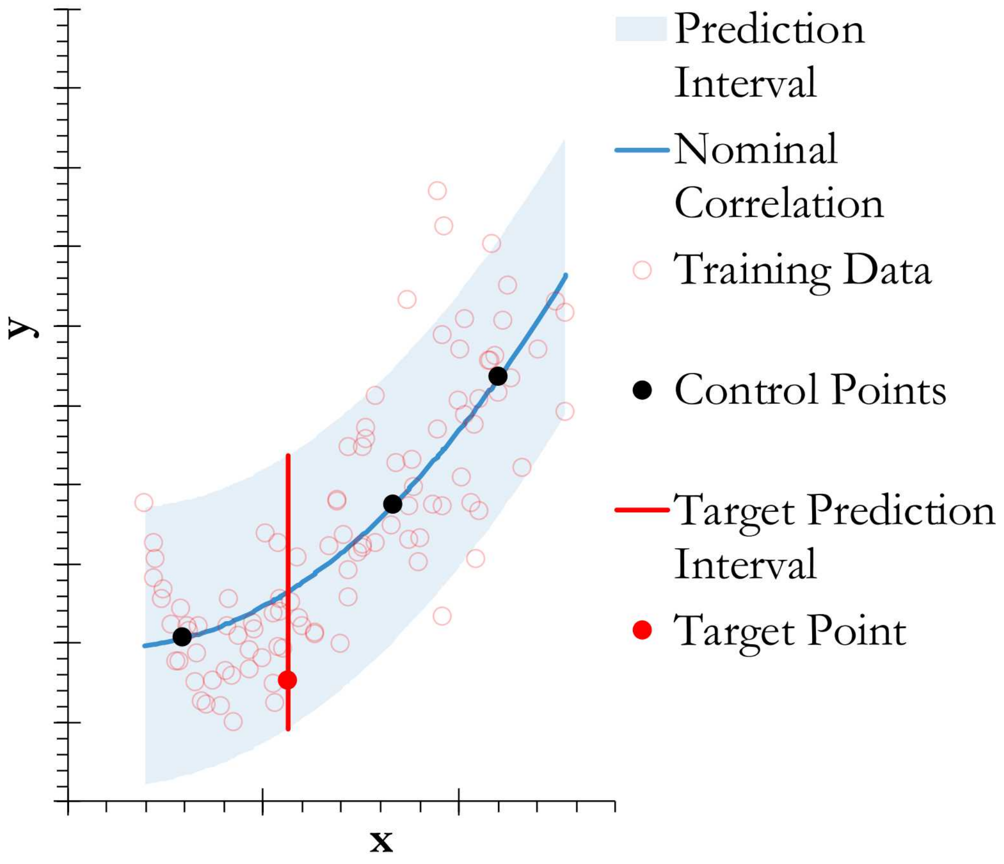
r_2 random variable from U(0,1)

Determine Target Response

- 1 “target” point required to force the correlation to sample the uncertain space
- Response selected randomly from prediction interval at target input
  - r_2 determines the random sample from the prediction interval uniform distribution
- Constrains sample correlation response

$$y_{tar} = f_{nom}(x_{tar}) + t(r_2, N_{trn} - 2) m_{PI}(x_{tar})$$

y_{tar} target dependent variable value



Target Response Point

Procedure Walkthrough with Example – Step 9

Find Sample Coefficients

- Optimize sample correlation realization with same form as nominal correlation
- Changing sample coefficients (decision variables)
- Constrain sample correlation to pass through target point
- Minimize difference between sample & nominal correlations at control points

$$f_{smp}(x) = f_{corr}(\boldsymbol{a}_{smp}, x)$$

$\boldsymbol{a}_{smp}$ set of sample correlation coefficients

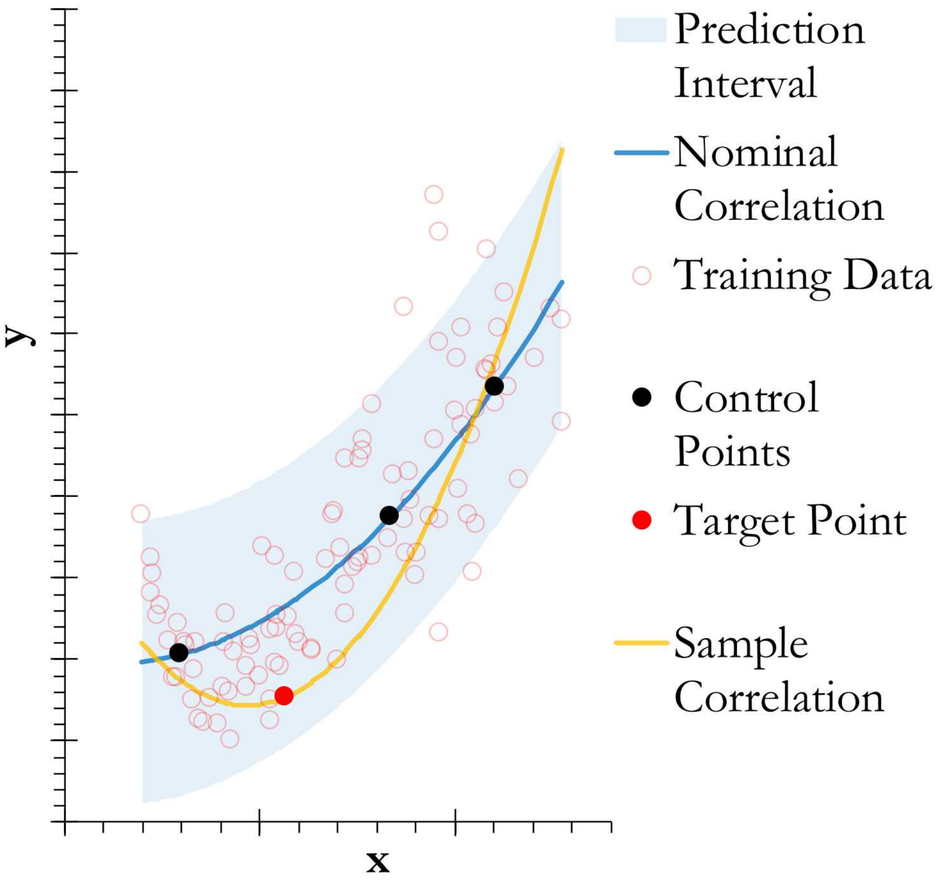
f_{smp} sample correlation

Using least squares objective function

$$\min_{\boldsymbol{a}_{smp}} \Sigma_{i=1}^{N_a}\left(f_{smp}\left(x_{ctrl,i}\right) - y_{ctrl,i}\right)^2$$

Subject to constraint

$$f_{smp}(x_{tar}) - y_{tar} = 0$$

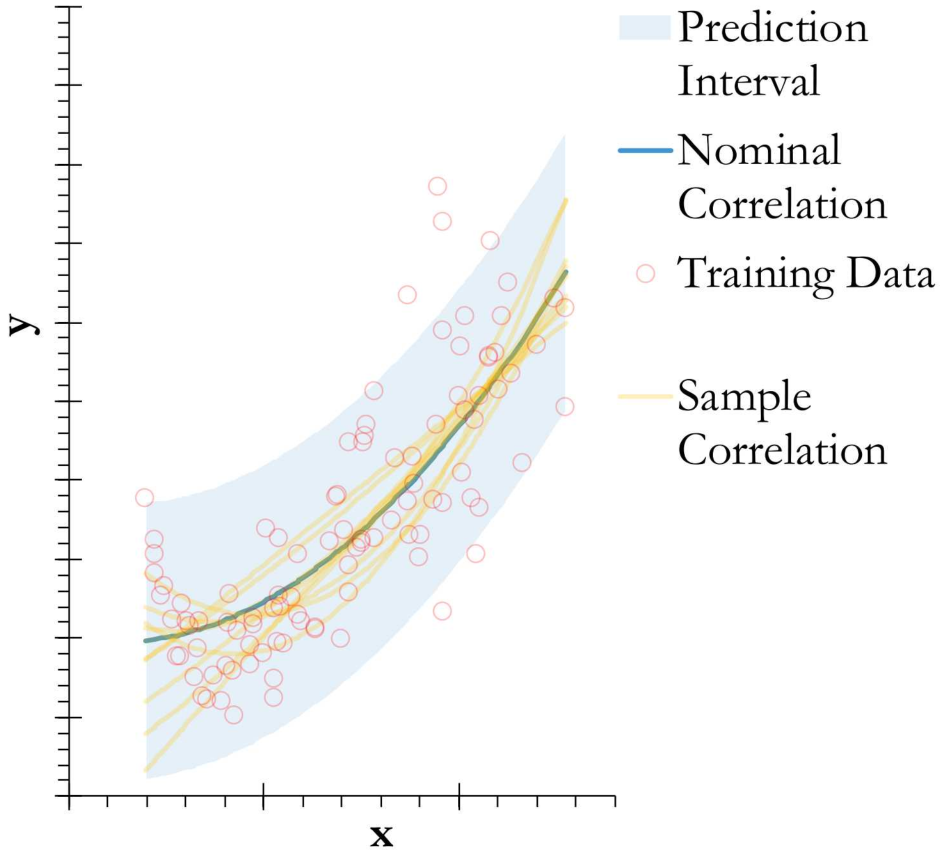


Constrained Optimization of Sample Correlation

Procedure Walkthrough with Example – Step 10

Iterate Sampling

- Repeat random sampling
- Generates discrete sets of sample correlation realizations



Ensemble of Sample Correlations

Procedure Walkthrough with Example – Step 11

Perform F-test

- Comparing the equivalence of two variances
- Assumes variations are normal
- Assumes underlying population variances of nominal correlation and sample ensemble are equivalent
- Not robust when assumptions are inaccurate
- Assume prediction interval deviation rate is nominal correlation standard deviation

$$F_{cmp} = \frac{\sigma_1^2}{\sigma_2^2}$$

F_{cmp} computed F-statistic

σ_1 largest standard deviation between PI & samples

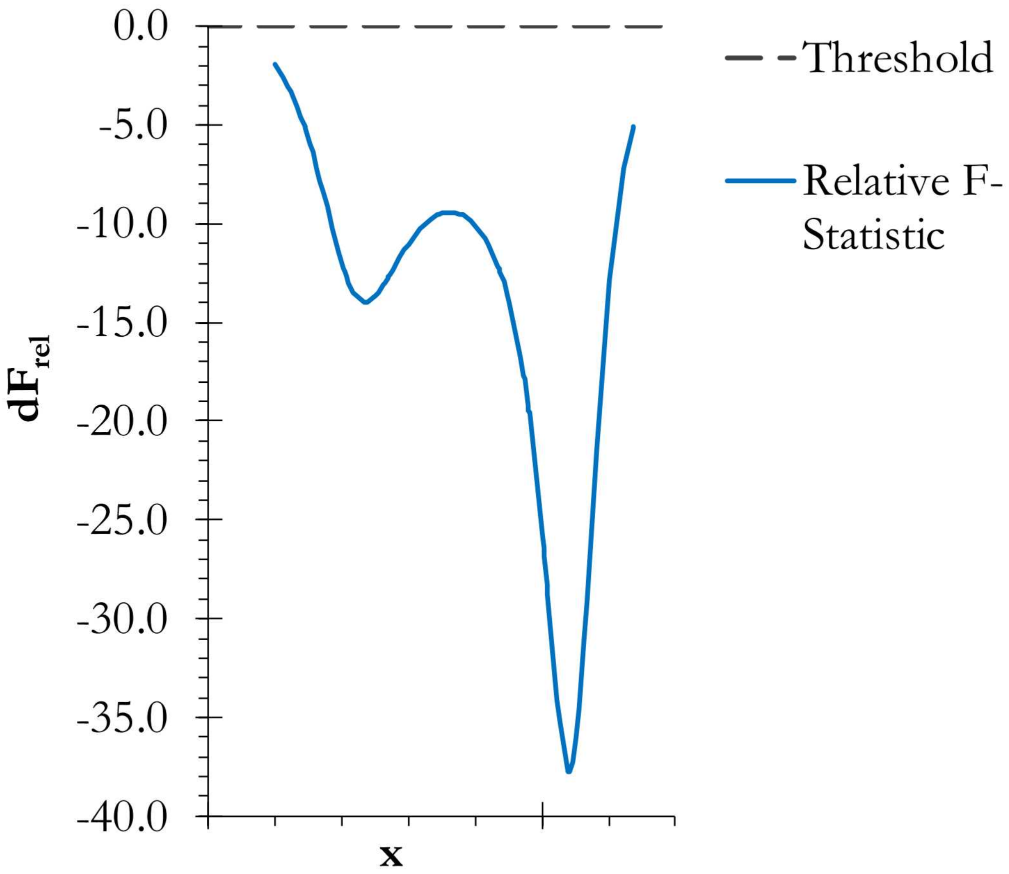
σ_1 smallest standard deviation between PI & samples

$$dF_{rel} = \frac{F_{crt} - F_{cmp}}{F_{crt}}$$

dF_{rel} relative F-statistic

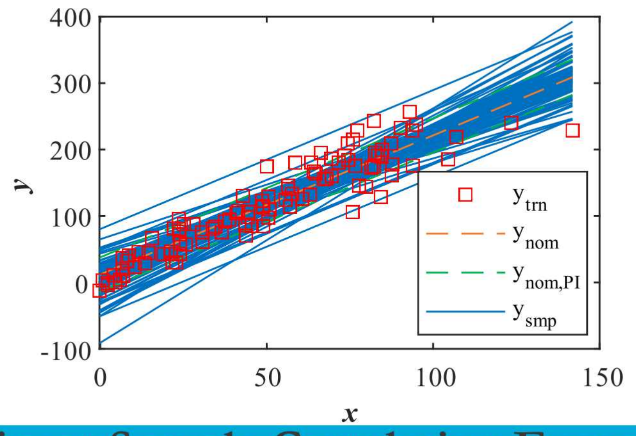
F_{crt} critical F-statistic from tables

- If dF_{rel}>0 cannot reject hypothesis that variances are equal

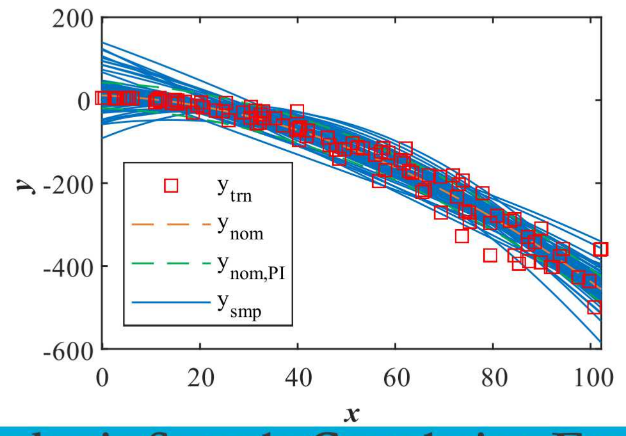


Variation Comparison F-Test

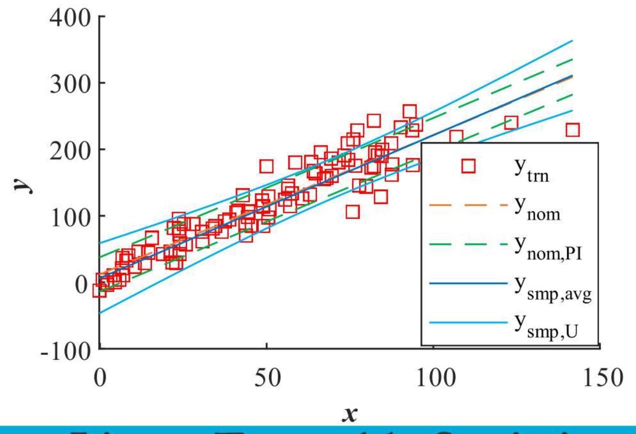
Example Results [1]



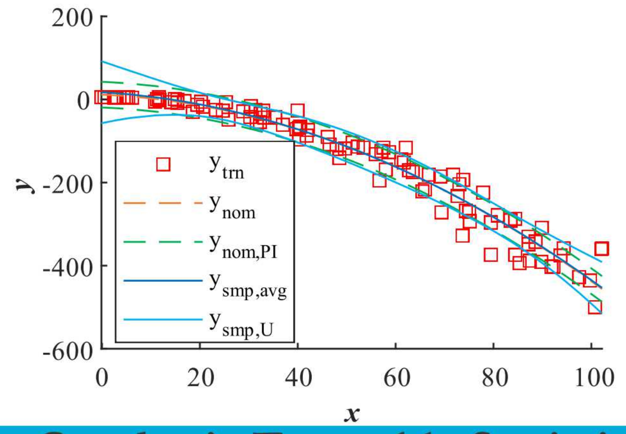
Linear Sample Correlation Ensemble



Quadratic Sample Correlation Ensemble

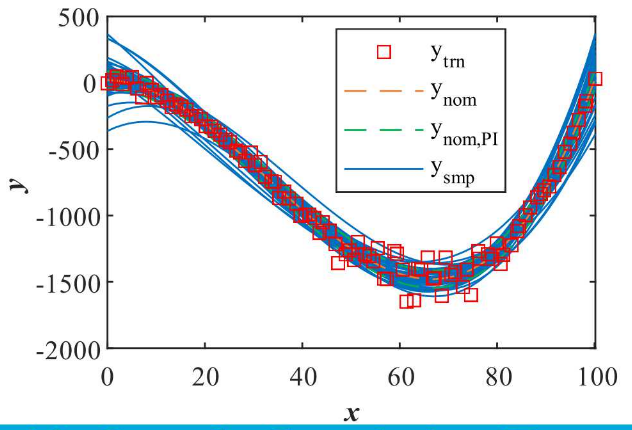


Linear Ensemble Statistics

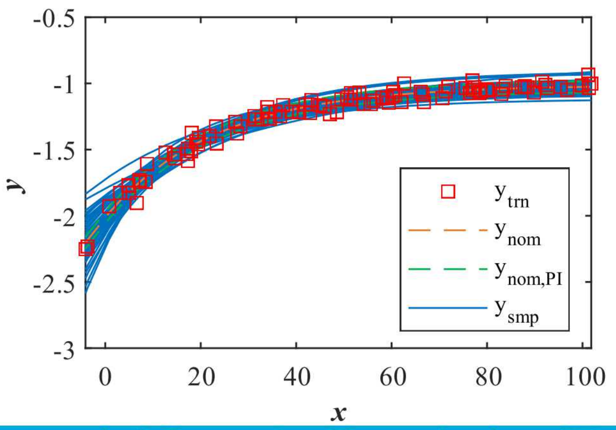


Quadratic Ensemble Statistics

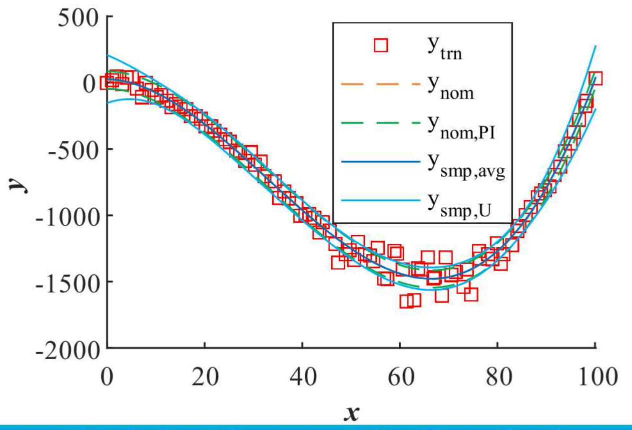
Example Results [2]



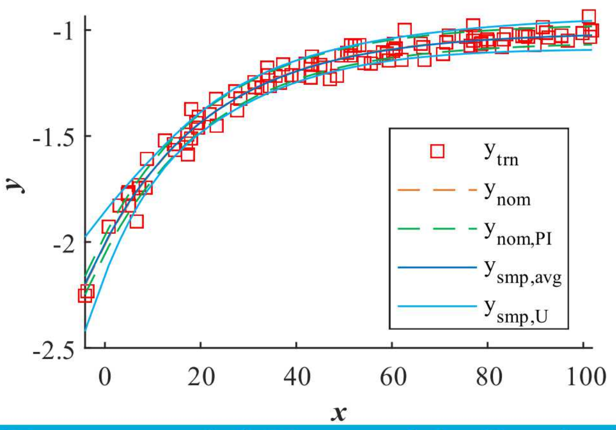
Cubic Sample Correlation Ensemble



Exponential Sample Correlation Ensemble

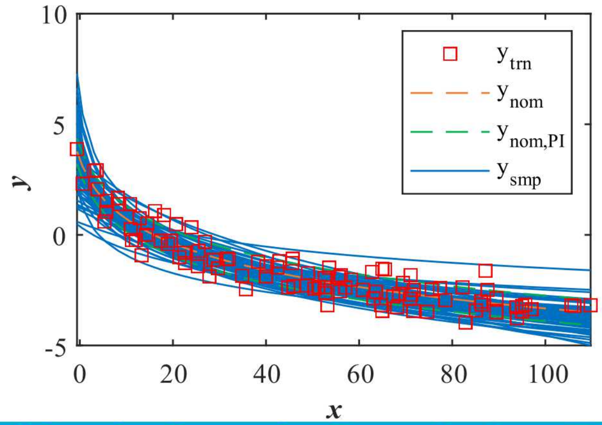


Cubic Ensemble Statistics

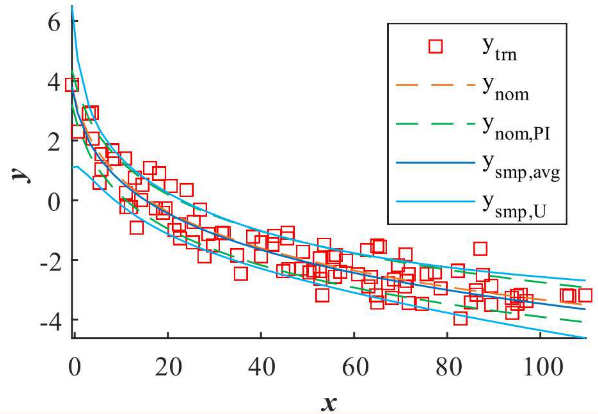


Exponential Ensemble Statistics

Example Results [3]



Logarithmic Sample Correlation Ensemble



Logarithmic Ensemble Statistics

Caveats, Conclusions & Future Work

Caveats
- Relatively wide uncertainties (variations) generated near upper and lower bounds of range
  - Primarily due to unconstrained nature of optimization process
- Additional constraints can be added to optimization process where needed
- Produces outlier realizations
  - Looking at function evaluations at x_{min} and x_{max} can provide insight
  - Can filter based on the least squares objective function value and/or the target point constraint

Conclusions
- "Abstract" function sampling methodology
- Method respects statistical behavior of entire function
- Method respects shape of underlying correlation
- Requires generation of two random variables
- Non-biased sampling

Future Work
- Other metrics of evaluation (as opposed to F-test)?
- How to apply to sensitivity analysis? (I.e., how to consider overall contribution of correlation uncertainty?)
- Working on implementation with PDEs

Acknowledgements

Acknowledgements

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. *SAND2020-2831 C*