Introduction to Uncertainty Quantification

Tutorial

Bruce Fryxell
University of Michigan

Turbulent Mixing and Beyond 2009
August 7, 2009- Trieste, Italy

What is Uncertainty Quantification?

- Next step beyond traditional verification and validation
- Formal framework for the quantification of errors and uncertainties in numerical simulations using statistical techniques
- Begin by identifying all possible sources of uncertainty in a numerical simulation
- Determine how these uncertainties propagate through the simulation code to create uncertainty in the output quantities of interest
- Instead of obtaining a single answer for each quantity of interest, obtain a probability distribution for each quantity with error bars
- Involves performing a very large number of numerical simulations by varying each of the input parameters

Goals

- To determine which input uncertainties produce the largest output uncertainties, and where possible, take steps to reduce them
- To use experiments to constrain the uncertainties of the input variables
- To predict the results of future experiments with error bars
- To prioritize activities that best reduce uncertainty and increase confidence in the results
- To do all of this as efficiently as possible

Steps in the uncertainty quantification process

- **Identify sources of uncertainty**
- Dimension reduction
- Design of experiments
- Screening
- Construction of statistical model
- Prediction
- Calibration

Sources of uncertainty

- X
  - Initial conditions
  - Boundary conditions
- θ
  - Material parameters
    - Equation of state
    - Opacity tables
- M
  - Mesh parameters
    - Grid size
    - Number of spatial dimensions
    - Number and structure of frequency groups
- P
  - Code tuning parameters
    - Artificial viscosity
    - Time step control
    - Constants in turbulence model

Types of uncertainty

- Aleatory
  - Uncertainty that can be characterized by a probability distribution
  - Probability distribution can be obtained using expert judgment, results of previous experiments, etc.
  - A good example is measurement error in determining initial conditions
- Epistemic
  - Uncertainty that results from lack of knowledge
  - Not mathematically correct to characterize these uncertainties using a probability distribution
  - These uncertainties should be treated in a different way
  - An example is uncertainty resulting from using physics models

Steps in the uncertainty quantification process

- Identify sources of uncertainty
- **Dimension reduction**
- Design of experiments
- Screening
- Construction of statistical model
- Prediction
- Calibration

Dimension reduction

- Cost of analysis depends on the number of input parameters to a very large power
- It is crucial to reduce the dimension of the input space as much as possible
- Cost of analysis becomes prohibitive if the number of input parameters is more than a few tens
- For X parameters
  - Use expert judgment
- For θ parameters
  - Individual numbers in equation of state and opacity tables are not independent
  - Use simple physics models to reduce the number of parameters if possible
  - Use statistical curve fitting models such as PCA, PLS, or partitioning algorithms

Dimension reduction

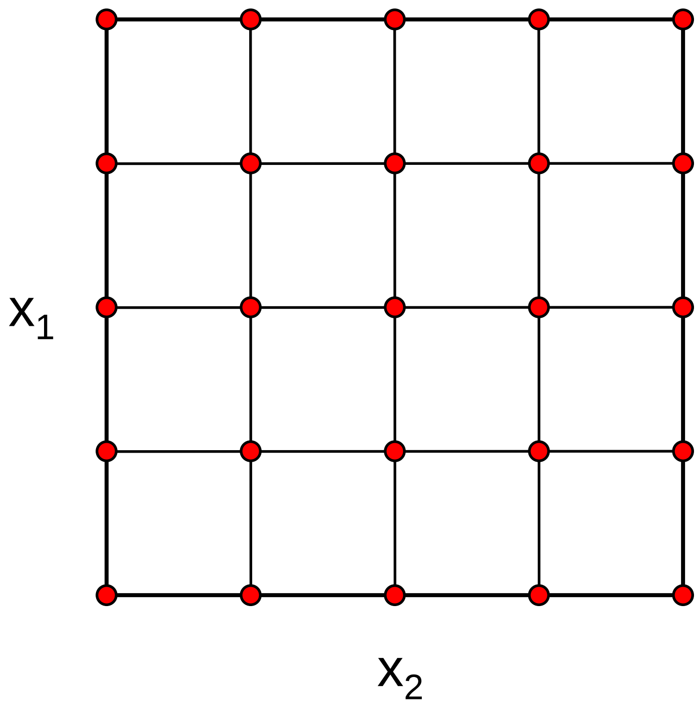
- For mesh (M) parameters
  - Ideally, perform an initial study to determine mesh parameters required to obtain a converged solution and then hold these fixed for the UQ study
  - For complex problems, convergence can not be achieved using a realistic amount of computer resources
  - In this case, the mesh parameters must be varied to determine how important they are relative to the other input parameters
- For code tuning (P) parameters
  - Many of these, such as artificial viscosity and time step controls, can be fixed at optimal values and considered as part of the discretization method
  - Others, such as parameters in a subgrid model, need to be varied to determine their impact on the results

Steps in the uncertainty quantification process

- Identify sources of uncertainty
- Dimension reduction
- **Design of experiments**
- Screening
- Construction of statistical model
- Prediction
- Calibration

Design of experiments

Simple grid



- For n parameters and m values per parameter, need $N_S = m^n$ simulations

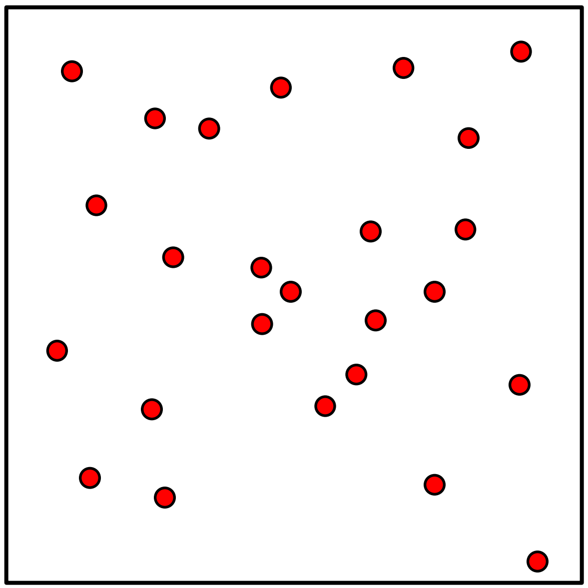
- For n = 15 and m = 5,

  $N_S \sim 3 \times 10^{10}$
- To do only corners

  $N_S = 32{,}768$

Design of experiments

Monte Carlo



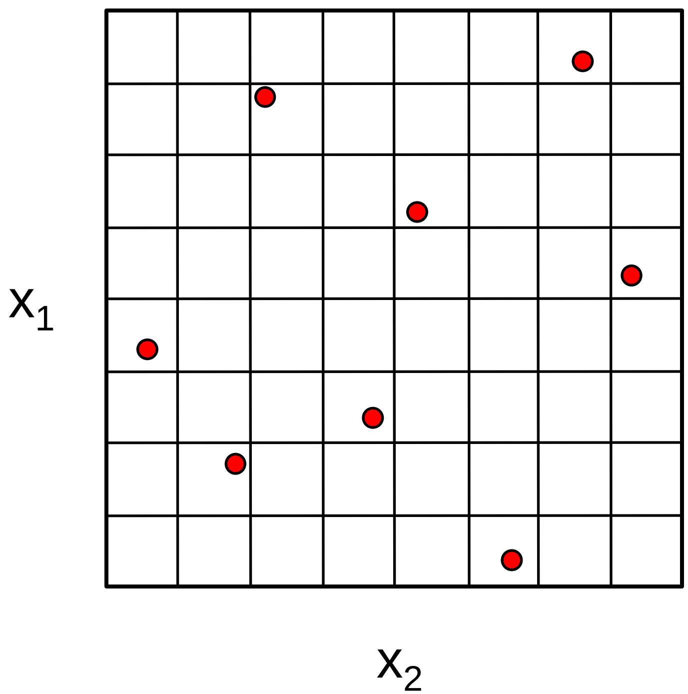
X_1

X_2

- Generate random numbers for (X_1, X_2, …, X_n)
- Choose N_S sets of parameters
- Covers space with relatively few points
- Space filling properties not particularly good

Design of experiments

Latin Hypercube



- Divide input space into
  $N_S \times N_S \times \ldots \times N_S$ grid
- Choose a cell by random
- Within that cell, choose a random location
- Use each row in each direction only once

Steps in the uncertainty quantification process

- Identify sources of uncertainty
- Dimension reduction
- Design of experiments
- **Screening**
- Construction of statistical model
- Prediction
- Calibration

Screening

- Even after using expert judgment to reduce the number of input parameters, not all the remaining parameters will be equally important in determining the output uncertainties
- It is vital to determine which parameters have the least impact on the results and remove them from the UQ study
- This can be accomplished by an initial screening process using a reduced set of simulations to determine regions of high and low sensitivity
- The use of emulators or reduced-fidelity simulations can help with this process
- Adding adjoint information into the simulation code is another approach

Steps in the uncertainty quantification process

- Identify sources of uncertainty
- Dimension reduction
- Design of experiments
- Screening
- **Construction of statistical model**
- Prediction
- Calibration

Construction of statistical model

- **Formulation of Kennedy and O'Hagan**
- **Experimental data Y can be expressed as**

$$Y(x_i) = \eta(x_i, \theta) + \delta(x_i) + \varepsilon_i$$
$$Y_S(x_i) = \eta(x_i, t)$$

- δ **accounts for discrepancy between the simulator and the real process which generates the experimental data**
- $\varepsilon \sim N(0, \lambda_\varepsilon^{-1})$ **is the experimental error**
- **Construct a statistical model for** $\delta(x_i)$ **and** $\eta(x_i, t)$
- **Now we can perform joint inference by using the posterior distribution of** $\pi(\theta, x \mid Y, Y_s)$

Gaussian process model

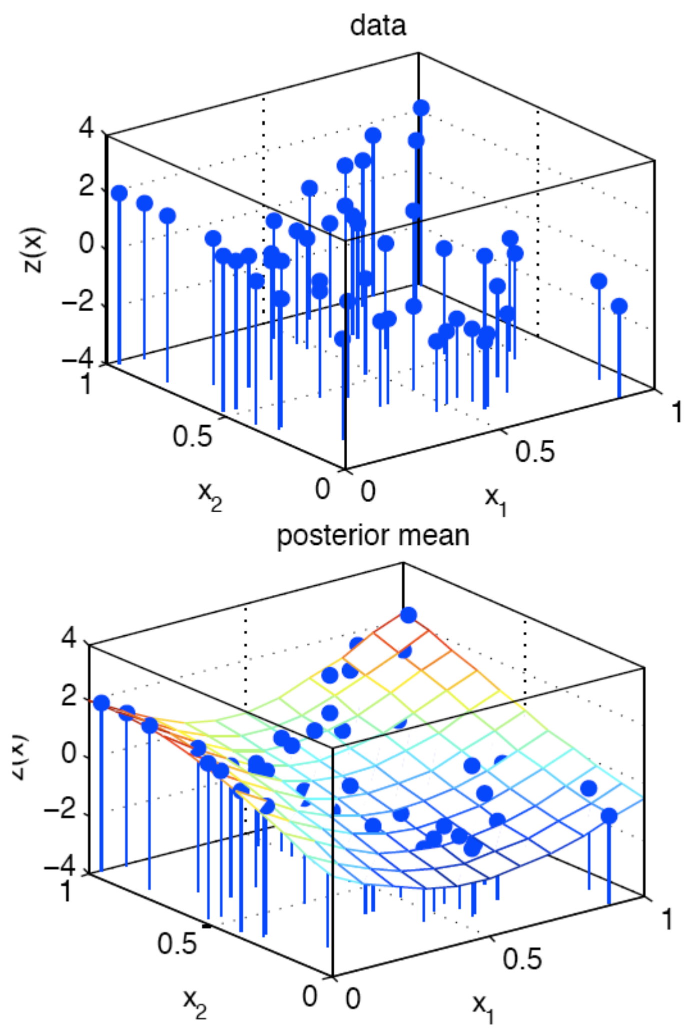
$$Y_S(x_i) = \mu + z(x_i)$$

$$z(x_i) = N(0, \Sigma_z)$$

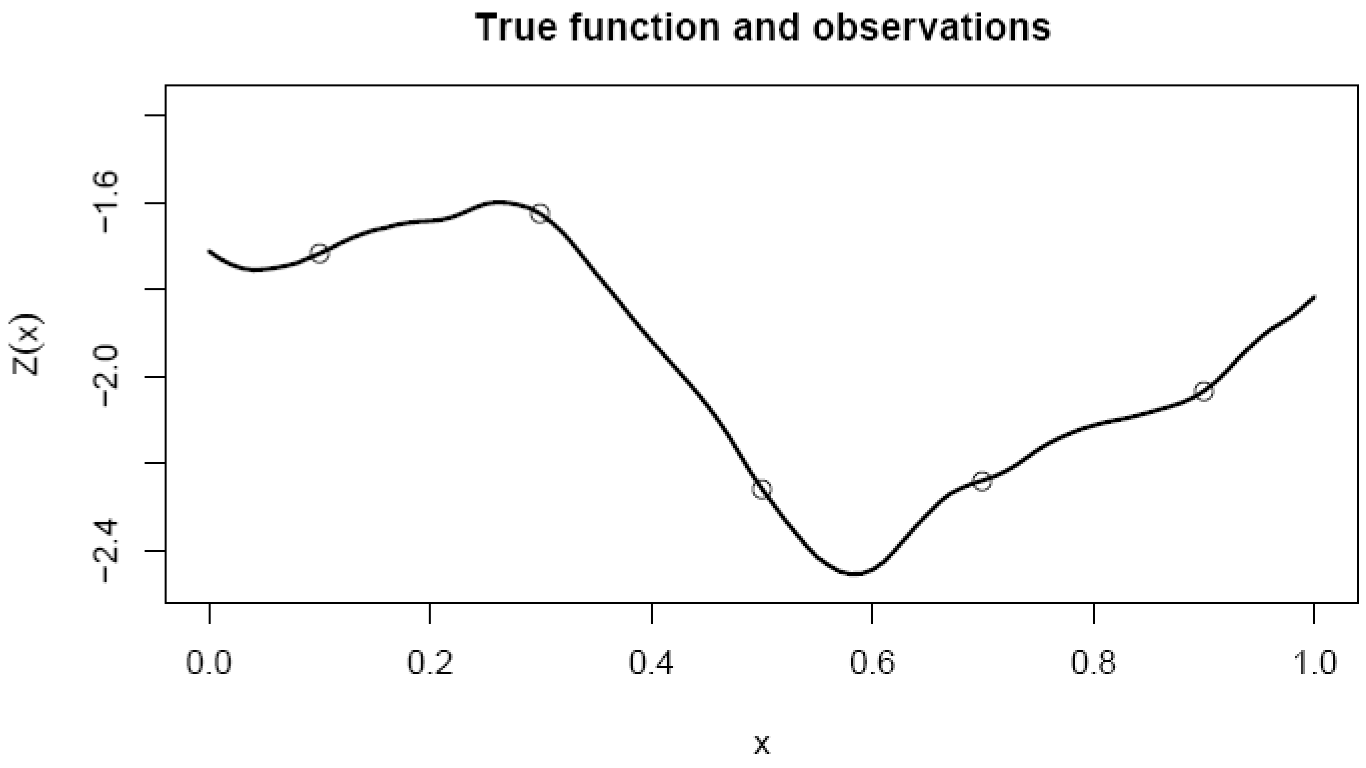
$$\Sigma_z = \mathrm{cov}\big(z(x_i), z(x_j)\big) = \frac{1}{\lambda_z}\exp\left(-\sum_{k=1}^{n}\beta_k\left(x_{ik} - x_{jk}\right)^2\right)$$

Gaussian process model

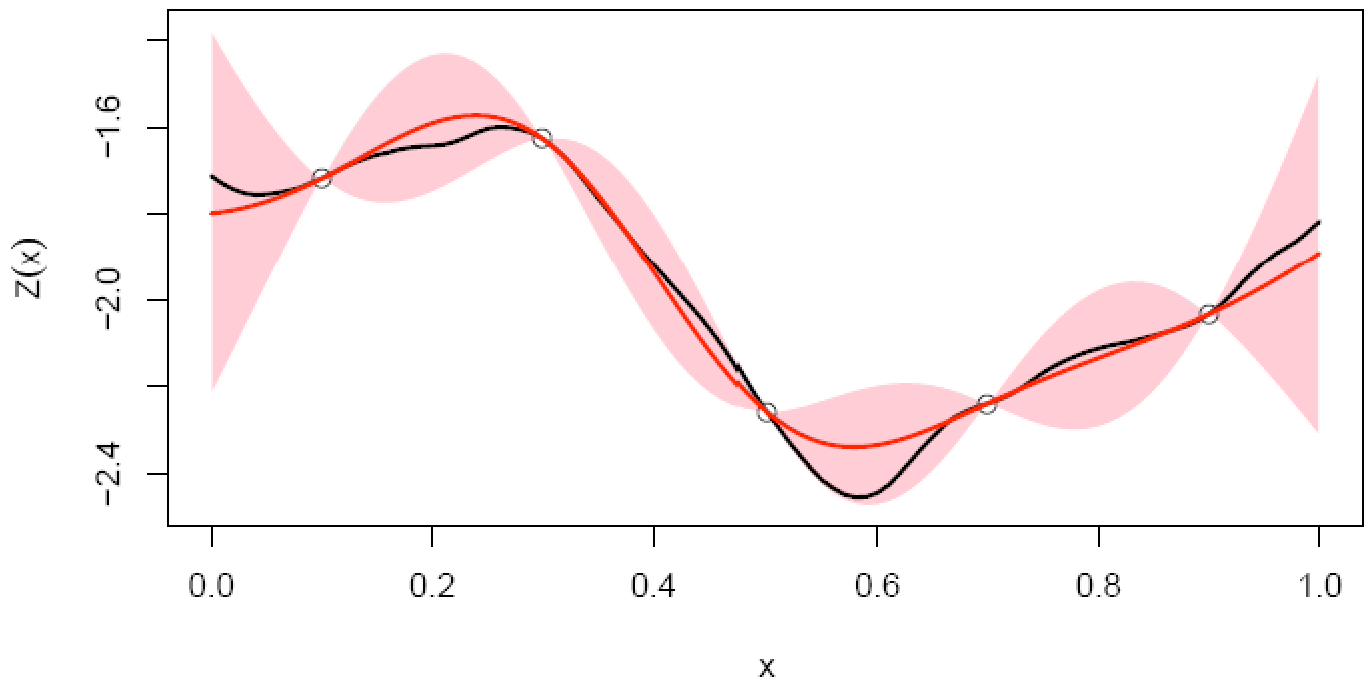
- **Essentially a regression model**
- **Response surface passes through the data points**
- **Error in fit increases with distance from the data points**



Gaussian process model



Gaussian process model

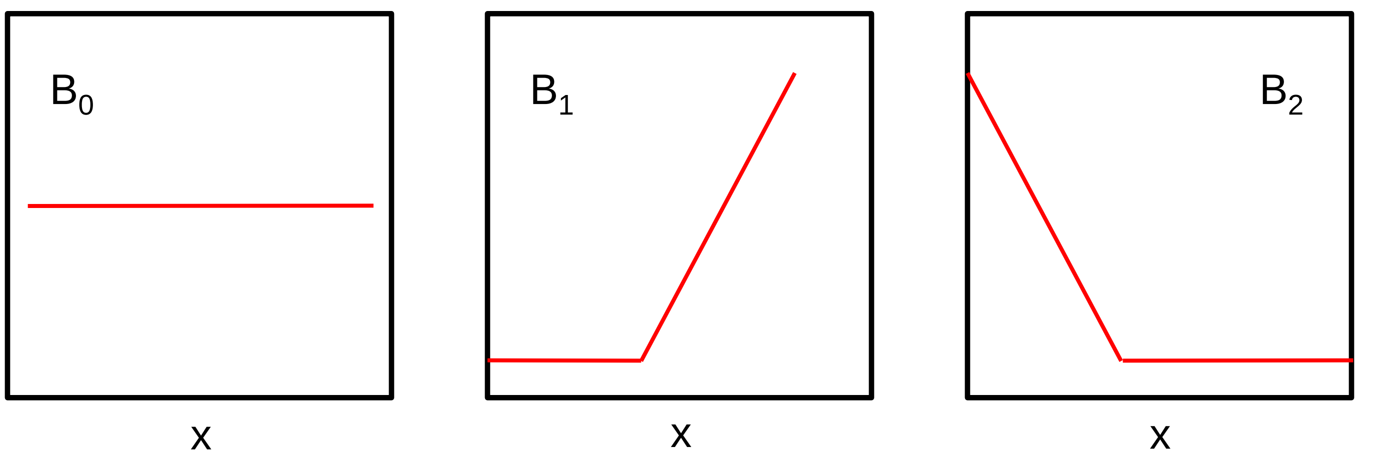


MARS (Multivariate Adaptive Regression Splines)

- **Expand using splines as basis functions**

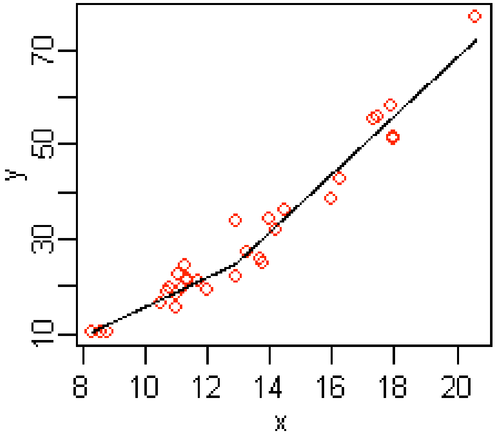
$$\widehat{Y}_S\left(x_i\right)=\sum_{j=1}^{k} c_j B_k\left(x_i\right)$$

- **Basis functions are a constant, hinge functions, and interactions among hinge functions**
- **Response surface does not, in general, pass through the data points**



MARS

- Better for high-dimensional case with sparsity
- More efficient for high-dimensional case – instead of including all possible effects, unimportant effects can be excluded
- Computationally faster than Gaussian process models – can handle larger data sets
- Each main effect and interaction effects can be easily estimated



Steps in the uncertainty quantification process

- Identify sources of uncertainty
- Dimension reduction
- Design of experiments
- Screening
- Construction of statistical model
- **Prediction**
- Calibration

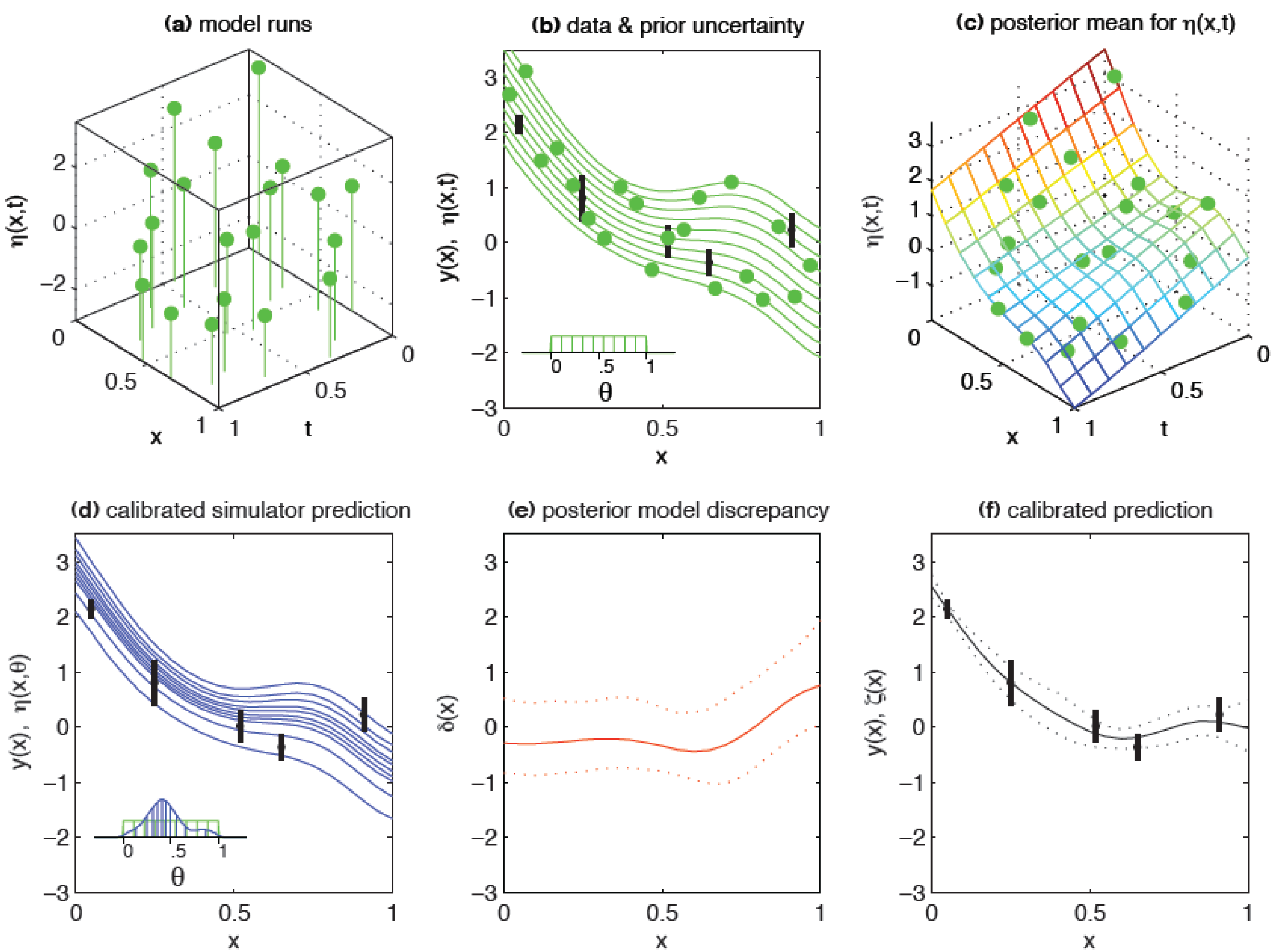
Prediction

- Use statistical model to generate response surface
- Response surface can be used to predict output quantities of interest using any new combinations of input parameters
- Goodness of fit of statistical model can be evaluated in a number of ways
  - Divide input data into two sets – training data and test data
    - Use the training data to construct the statistical model
    - Use statistical model to predict test data
  - Leave one out cross validation
    - Leave out data for one simulation, construct model using the remaining data, then try to predict the omitted simulation
    - Repeat for each simulation

Steps in the uncertainty quantification process

- Identify sources of uncertainty
- Dimension reduction
- Design of experiments
- Screening
- Construction of statistical model
- Prediction
- **Calibration**

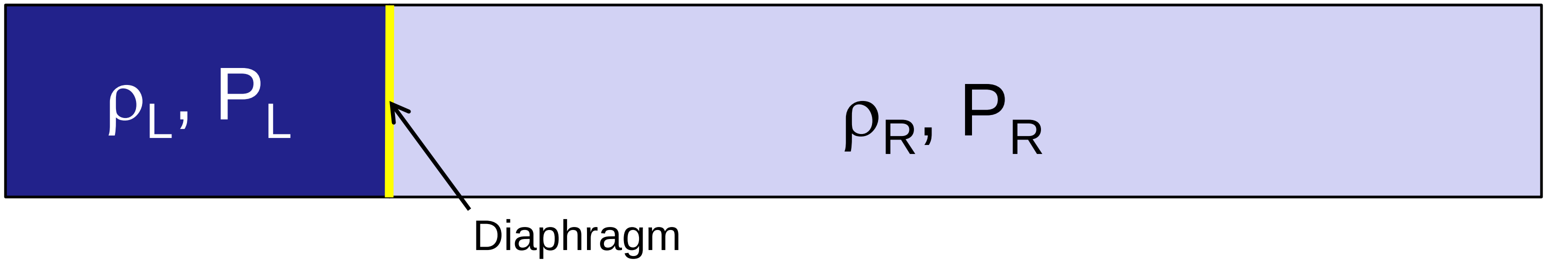
Calibration (inverse problem)

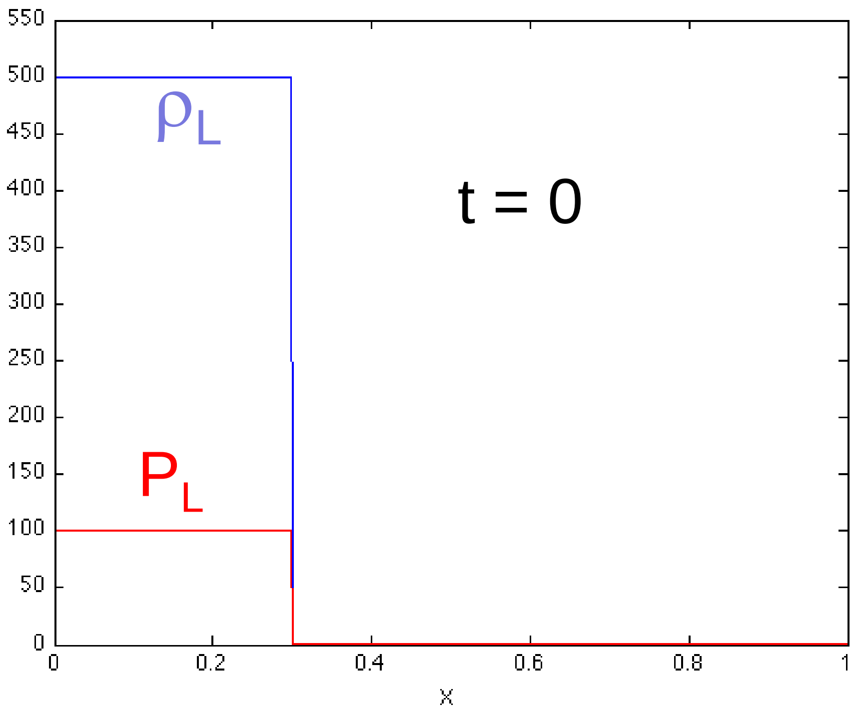


Sample UQ problem

- One-dimensional shock tube
  - Limited set of input and calibration parameters
  - Fast to simulate
  - Analytic solution
    - Can compare UQ analysis of analytic solution to analysis of simulation
    - Can use analytic solution as a substitute for experimental results

Initial Conditions





Fixed initial conditions

$\rho_R = P_R = 1$

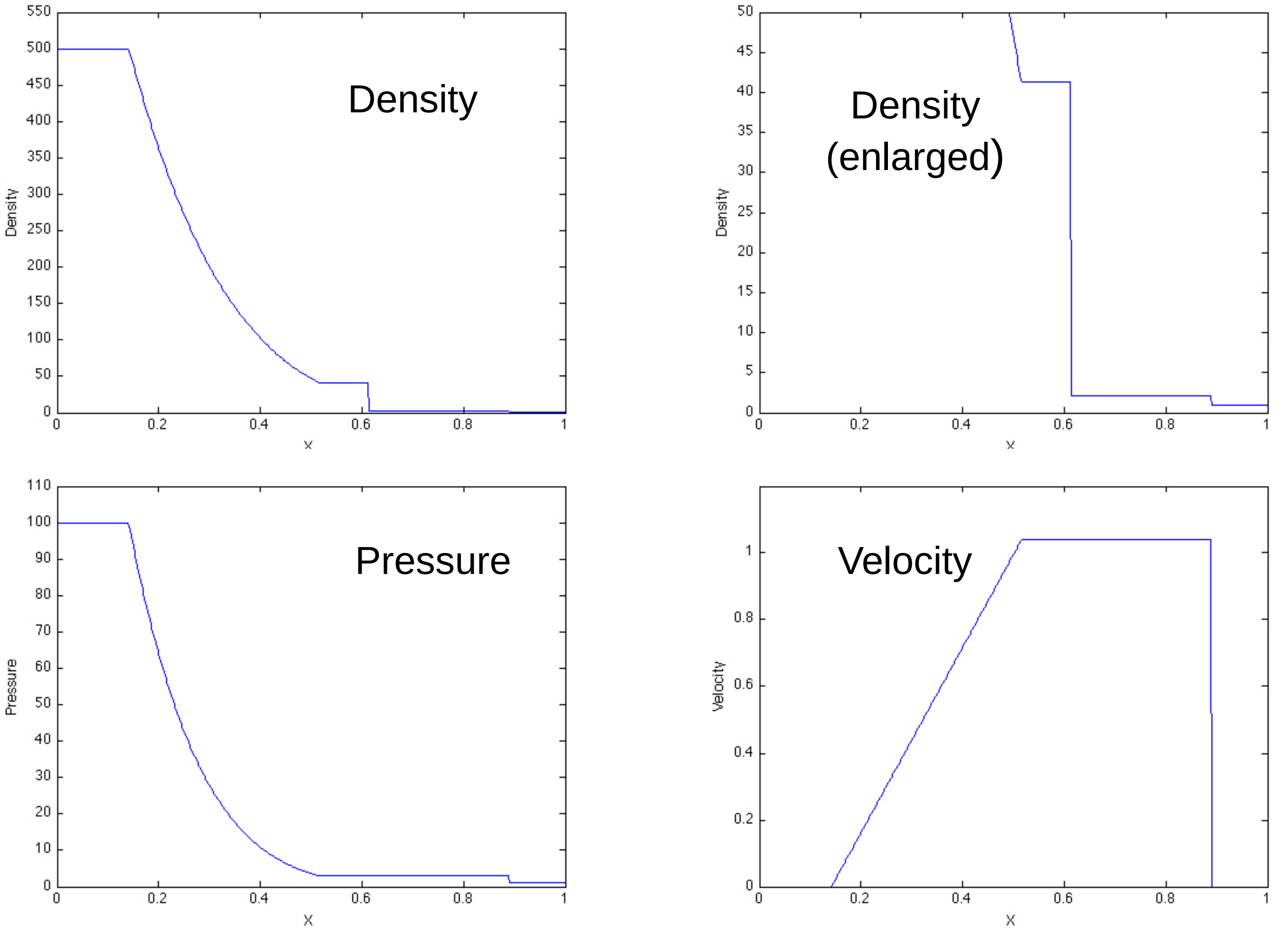
$u_L = u_R = 0$

Varied input parameters:

ρ_L, P_L, γ

Also added five inert input parameters that had no effect on the output

Analytic solution

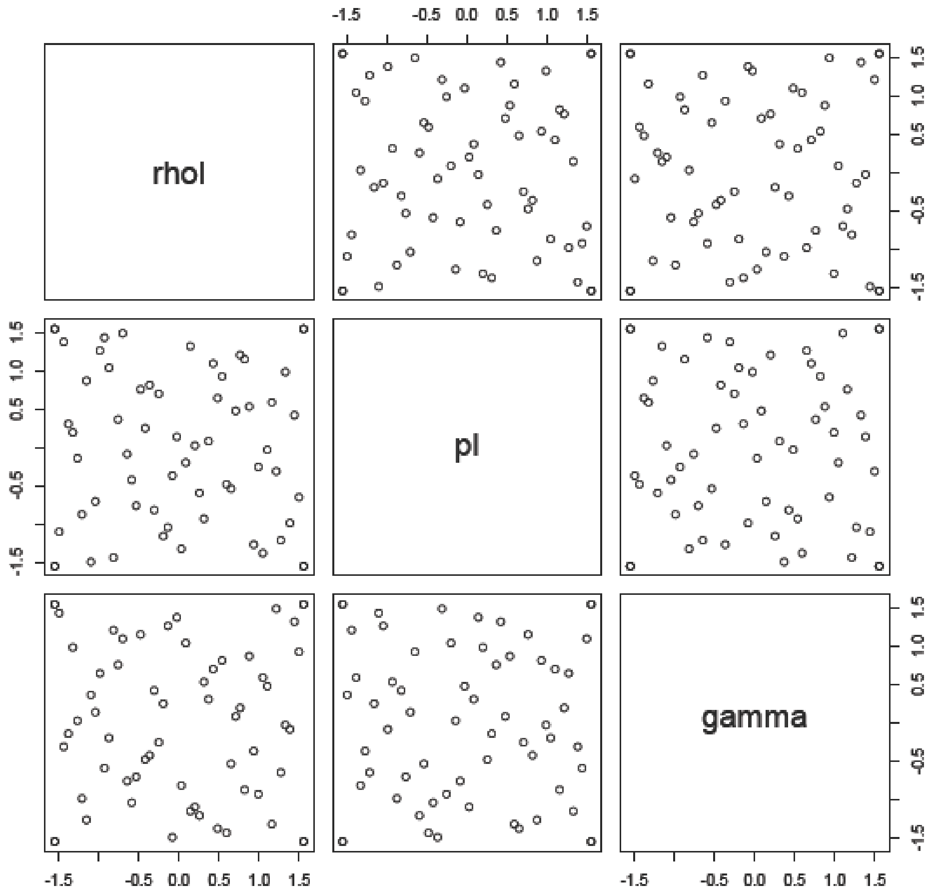


Eight output values

- **Four positions**
  - x_{shk} location of shock
  - x_{cd} location of contact discontinuity
  - x_{tail} location of tail of rarefaction
  - x_{head} location of head of rarefaction
- **Four values of state variables**
  - ρ_{shk} density at left of shock
  - ρ_{cd} density at left of contact discontinuity
  - P_{shk} pressure at left of shock
  - u_{shk} velocity at left of shock

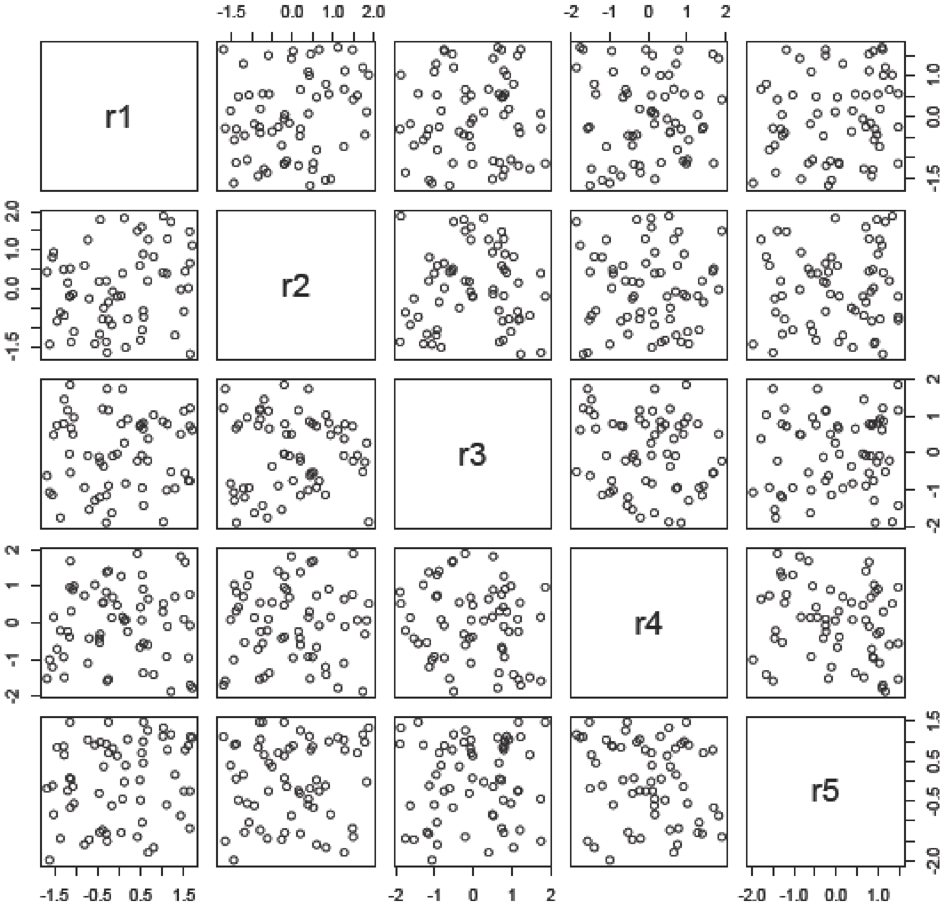
Distribution of input parameters

- **62 simulations**
  - **350 < ρ_L < 650**
  - **70 < P_L < 130**
  - **4/3 < γ < 5/3**
- **Orthogonal Array Latin Hypercube design**

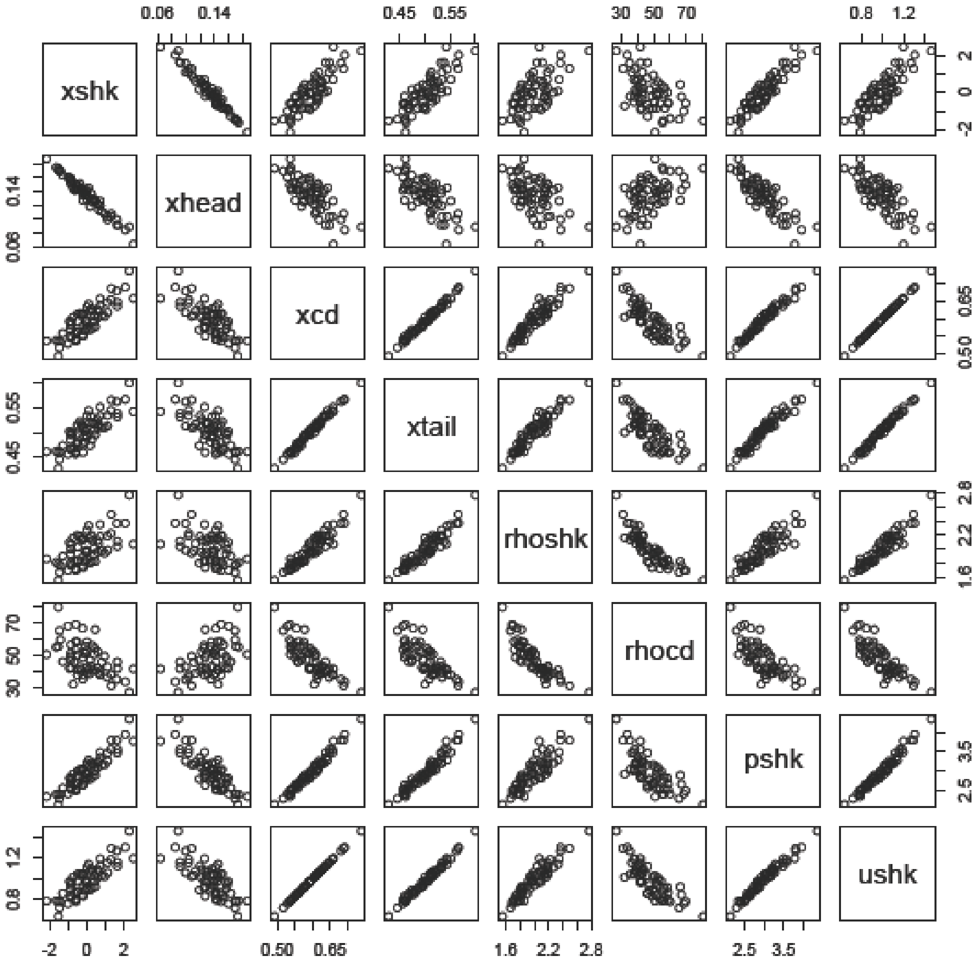


Inert input variables

Inert variables given a uniform random distribution between 0 and 1

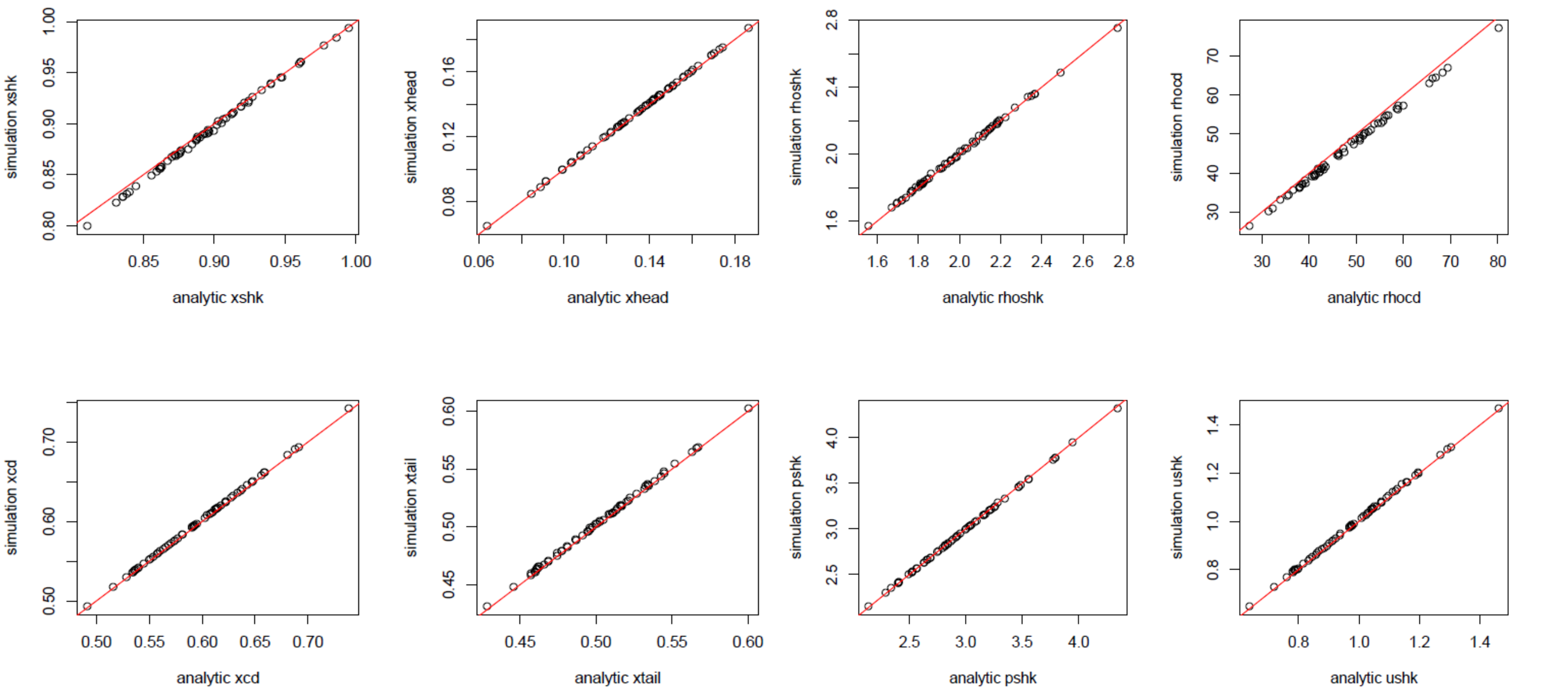


Correlations between output values - analytical



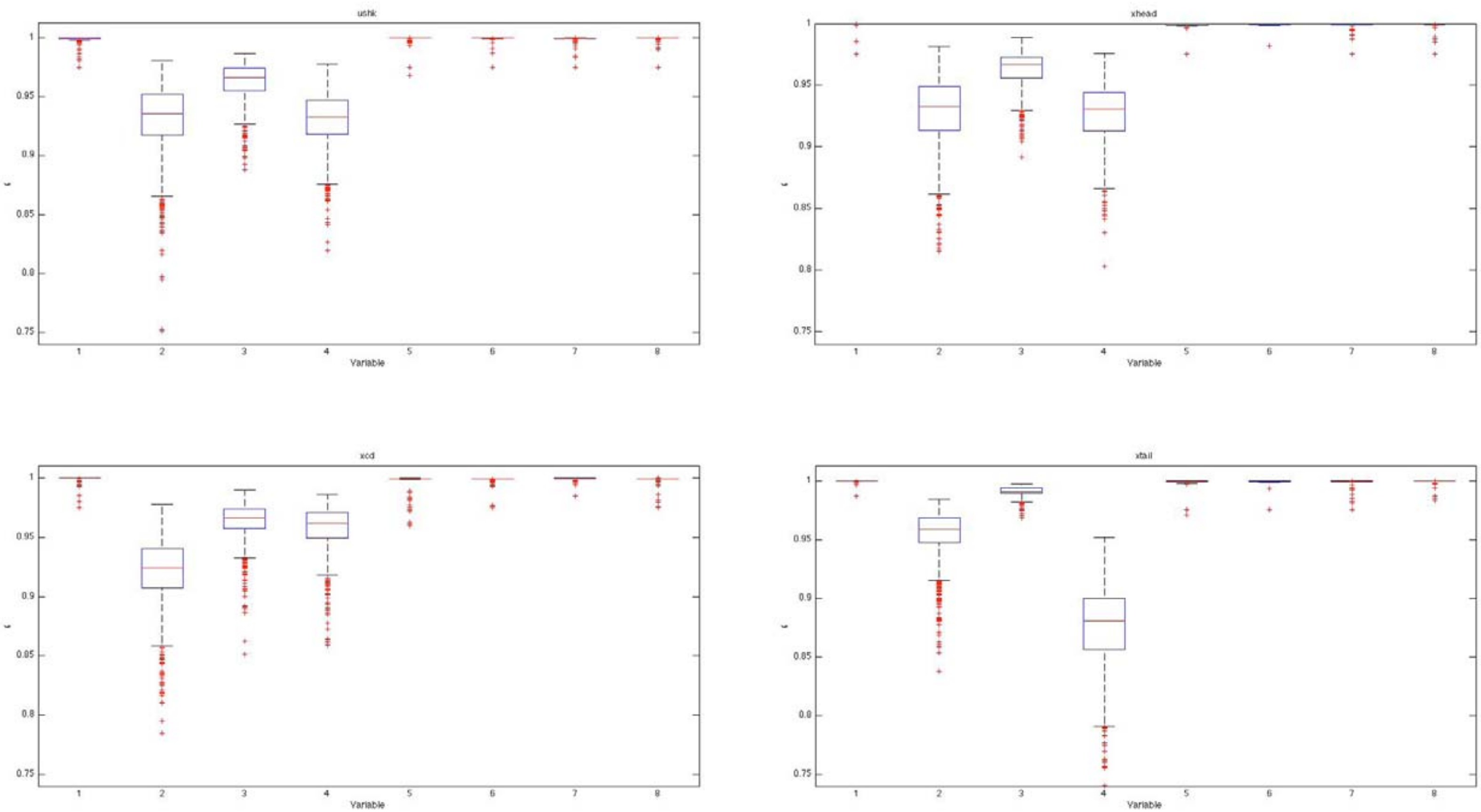
Correlations between output values - simulation

Simulation results vs. analytic solution

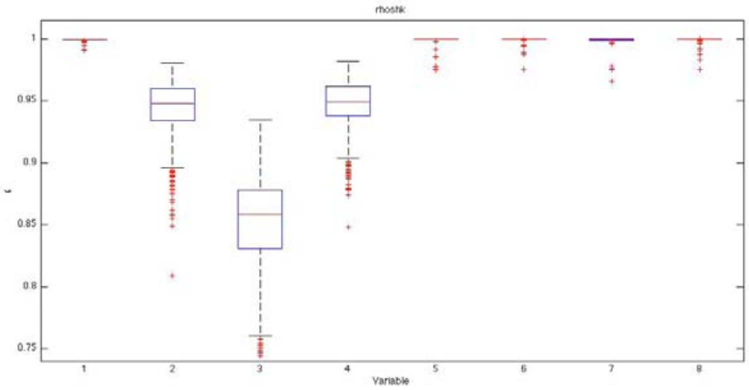


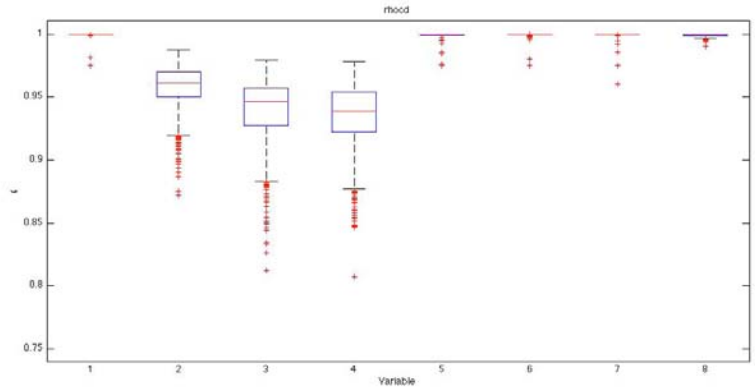
Some output variables show a small bias

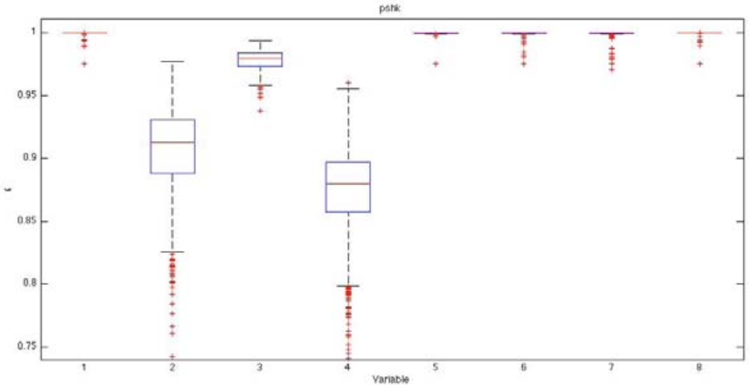
Gaussian process results – input significance

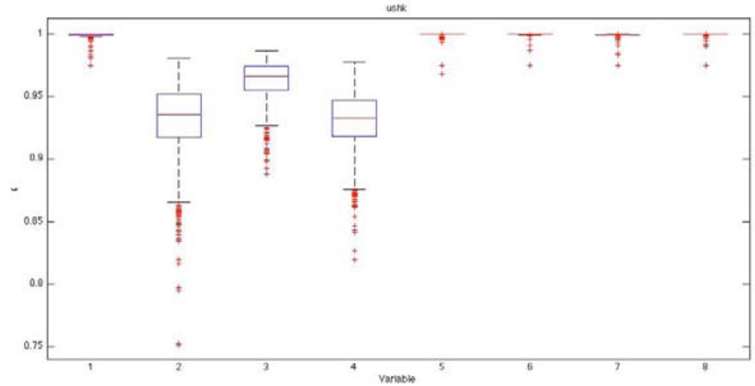


Gaussian process results – input significance









Bayesian MARS results

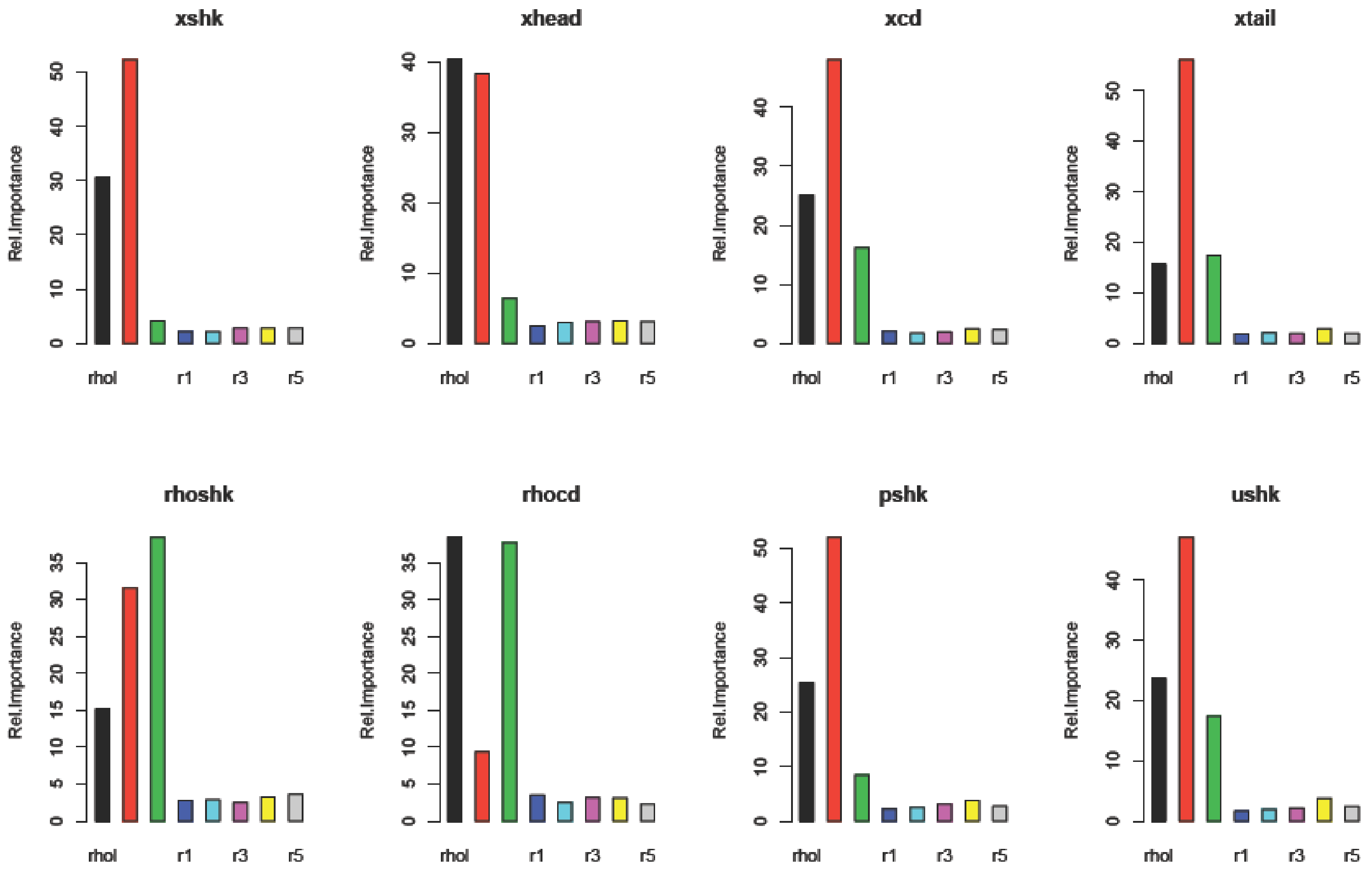
Significancy of effects for each outputs

x_{shk}		x_{cd}		x_{tail}		x_{head}	
Effect	Prob.	Effect	Prob.	Effect	Prob.	Effect	Prob.
ρ	1.00	ρ	1.00	ρ	1.00	ρ	1.00
P	1.00	P	1.00	P	1.00	P	1.00
γ	1.00	γ	1.00	γ	1.00	γ	1.00
(ρ, P)	0.64	(ρ, P)	0.02	(ρ, P, γ)	0.03	(ρ, P)	0.04
(ρ, P, r_4)	0.05	r_5	0.02	(P, γ)	0.02	r_5	0.02
(ρ, P, γ)	0.02	r_4	0.02	(ρ, P)	0.02	r_3	0.02
(ρ, γ)	0.02	(ρ, P, γ)	0.01	r_4	0.01	(ρ, r_4)	0.01
(ρ, P, r_3)	0.01	(ρ, r_4)	0.01	r_5	0.01	r_2	0.01
(P, γ)	0.01	r_2	0.01	(γ, r_1, r_2)	0.01	(ρ, P, r_1)	0.01
(P, γ, r_1)	0.01	(P, r_1)	0.01	(ρ, P, r_2)	0.01	r_4	0.01

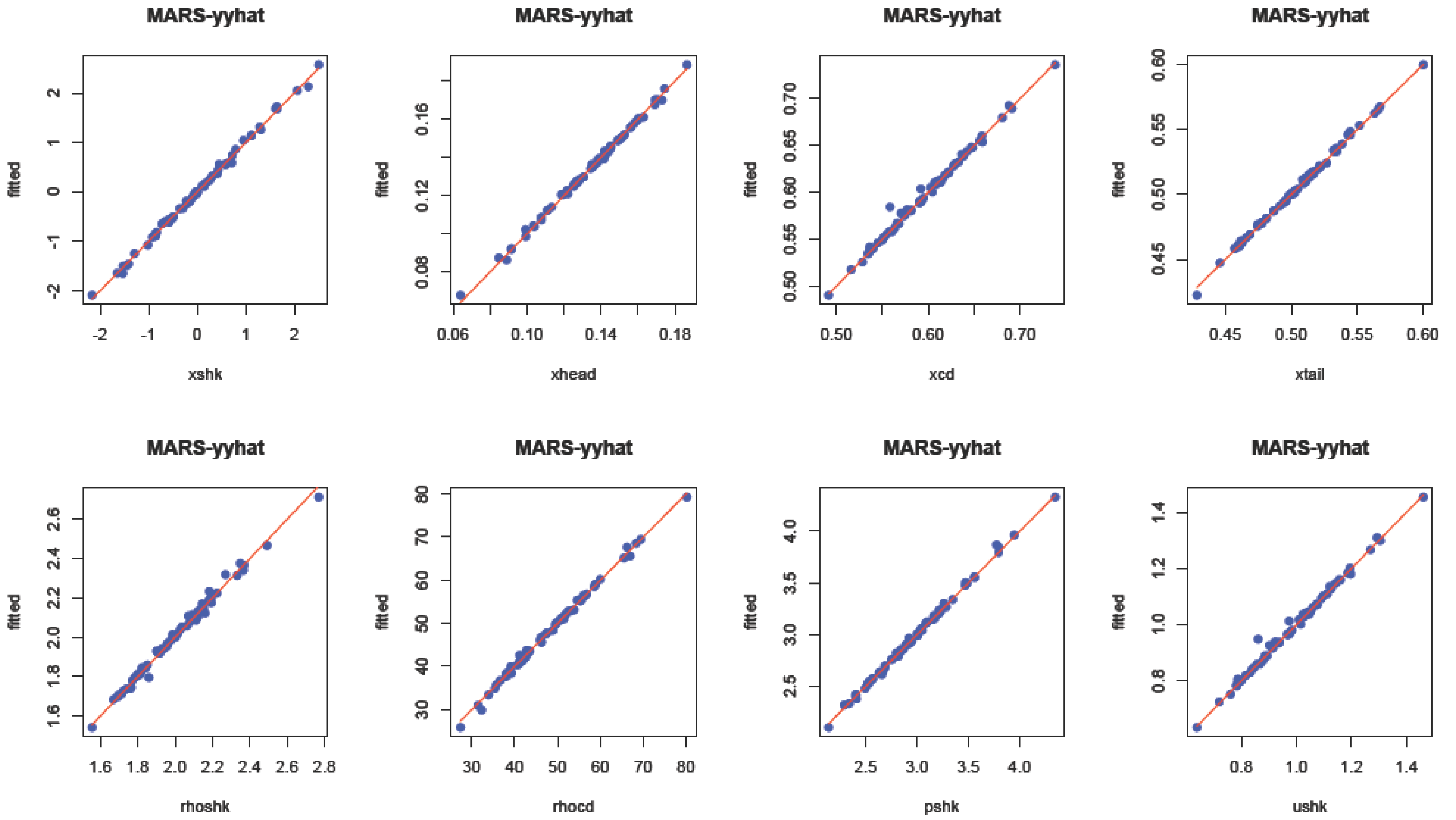
Significancy of effects for each outputs

ρ_{shk}		ρ_{cd}		P_{shk}		u_{shk}	
Effect	Prob.	Effect	Prob.	Effect	Prob.	Effect	Prob.
ρ	1.00	ρ	1.00	ρ	1.00	ρ	1.00
P	1.00	P	1.00	P	1.00	P	1.00
γ	1.00	γ	1.00	γ	1.00	γ	1.00
(P, γ)	0.59	(ρ, P)	1.00	(ρ, P)	1.00	r_4	0.01
(ρ, γ, r_5)	0.07	(ρ, γ)	1.00	(P, γ)	0.52	(ρ, P)	0.01
(P, γ, r_3)	0.06	(P, γ)	1.00	(ρ, γ)	0.04	r_1	0.01
(ρ, P)	0.04	(P, γ, r_4)	0.04	(P, γ, r_3)	0.02	r_5	0.01
(P, γ, r_5)	0.03	(ρ, r_1)	0.03	r_2	0.02	(ρ, r_4)	0.01
r_2	0.02	(ρ, r_3, r_4)	0.02	(P, γ, r_5)	0.02	(ρ, γ)	0.01
(ρ, γ)	0.02	(γ, r_4, r_5)	0.02	(ρ, r_3, r_5)	0.01	r_2	0.01

MART results



Leave one out cross validation

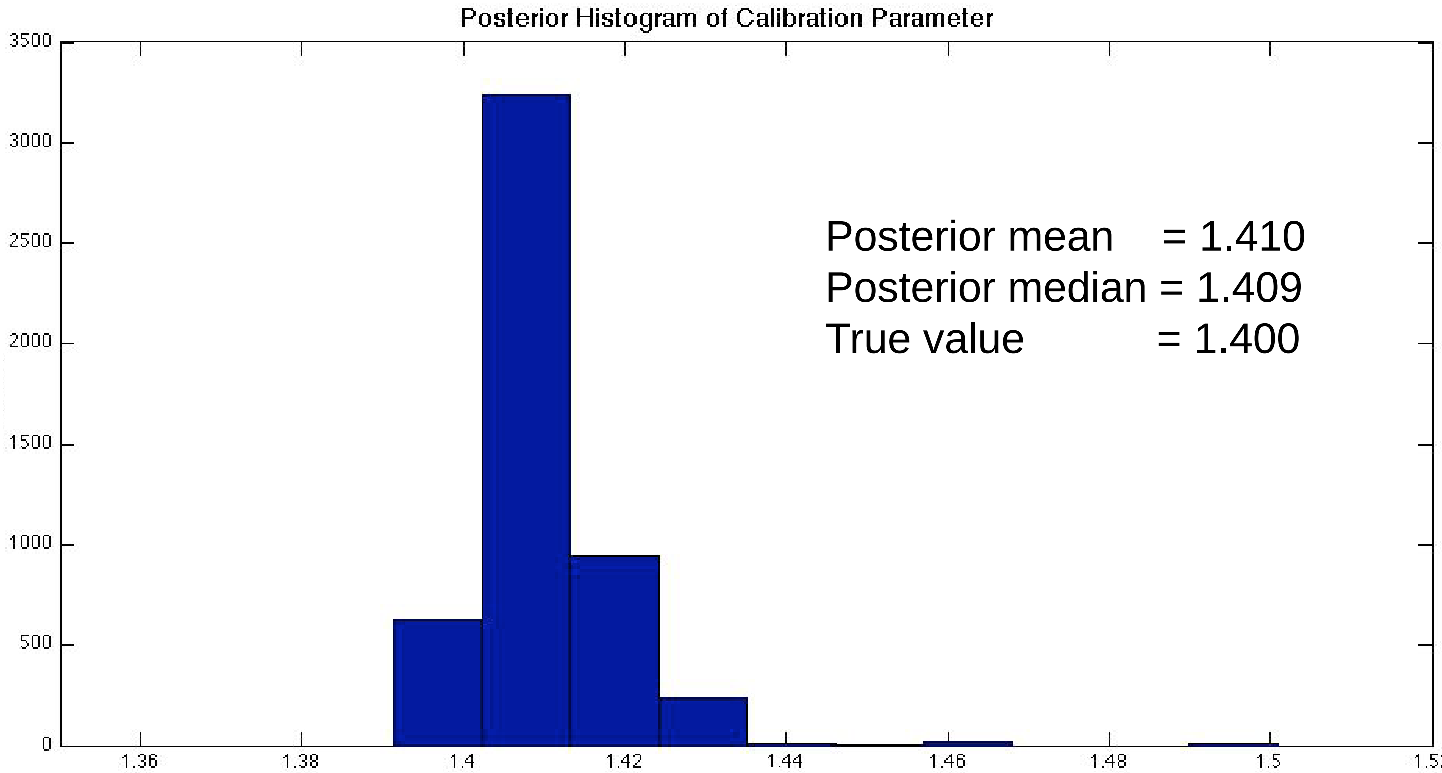


Calibration and prediction

- Used the 62 runs from the simulation code and 10 data points from the analytic solution (substitute for experiment)
- Analytic solution was run computed with a fixed (but unknown) value of γ
- Calibration (the inverse problem) was performed using all eight output variables
- Fit the above model, estimated γ, and predicted the remaining 52 runs

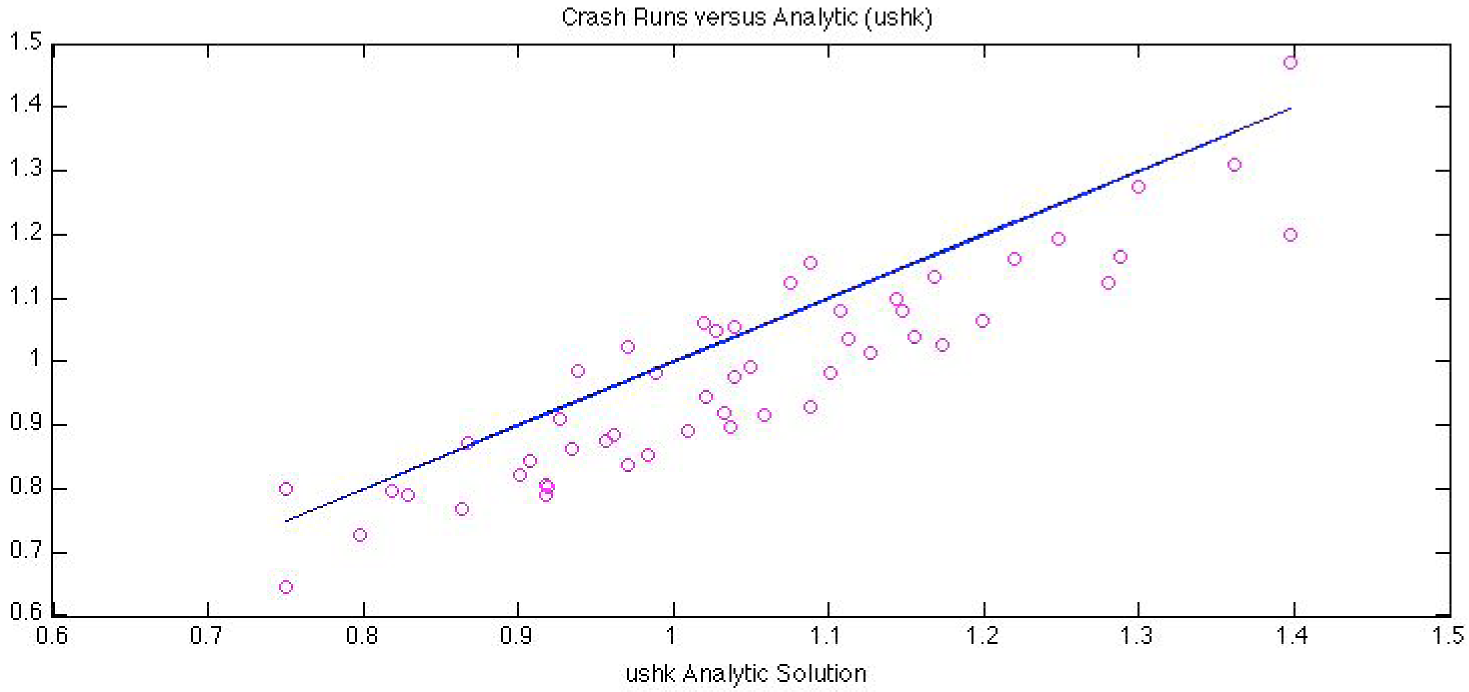
Calibration

Prediction of γ



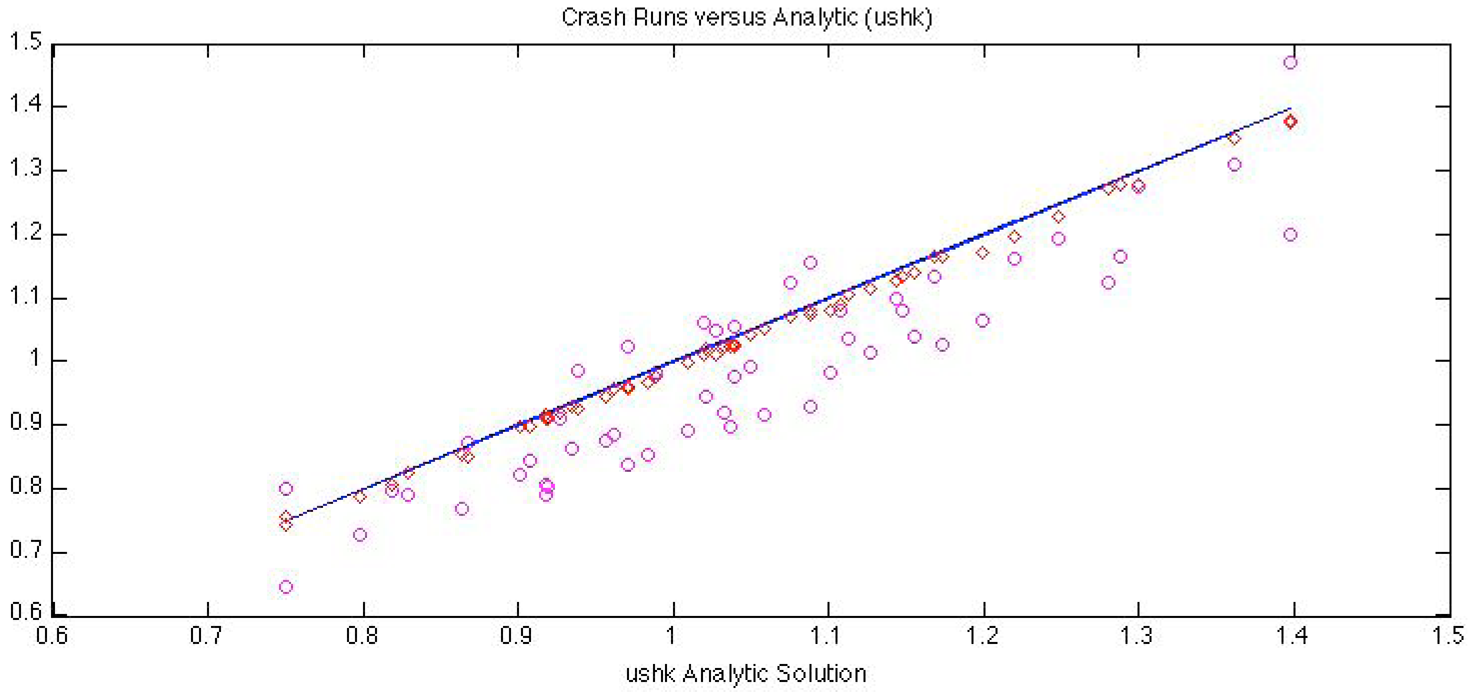
γ

Prediction



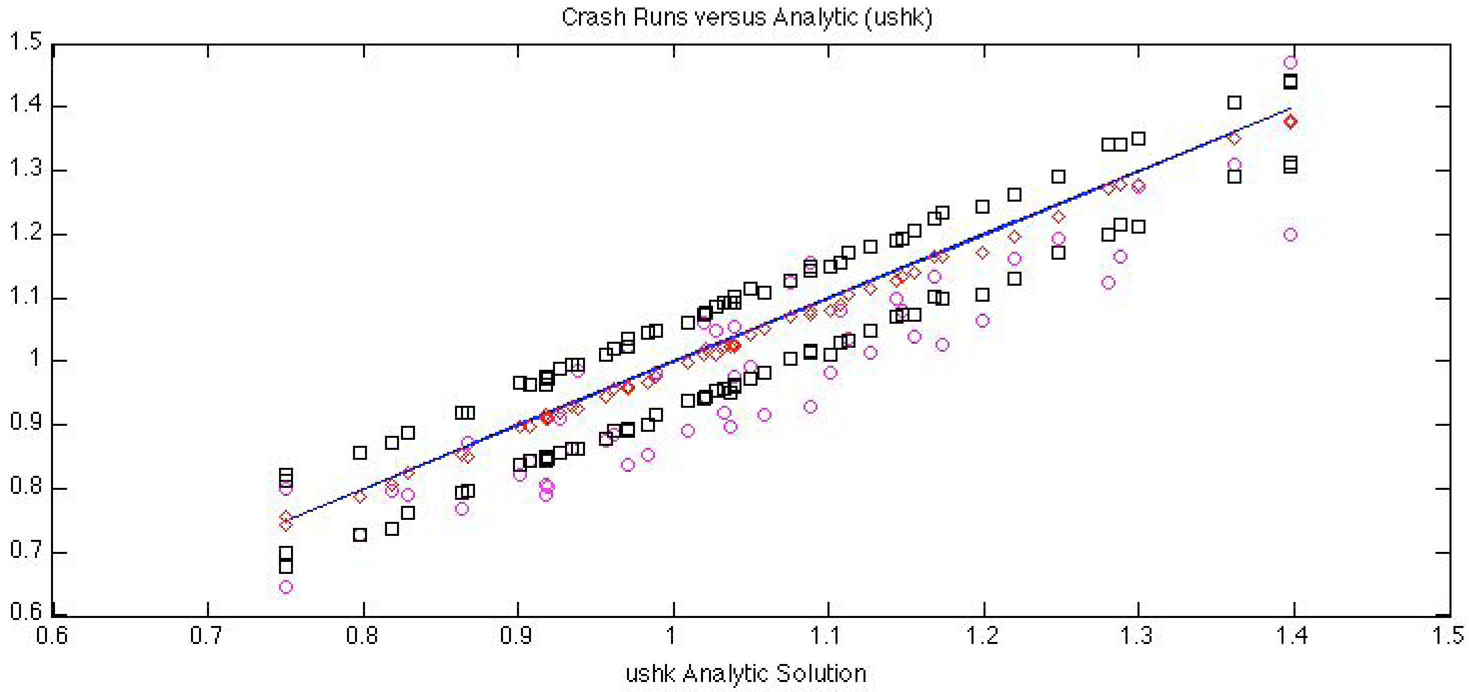
Crash Runs versus Analytic (ushk)
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
ushk Analytic Solution

Prediction



Crash Runs versus Analytic (ushk)
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
ushk Analytic Solution

Prediction



Crash Runs versus Analytic (ushk)
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
ushk Analytic Solution

Acknowledgement

This work was supported by the US DOE NNSA under the Predictive Science Academic Alliance Program by grant DE-FC52-08NA28616