Center for Applied Mathematics, Computation and Statistics

Report Day

MAY 14, 2013

Simulations, Metamodeling, and Stochastic Kriging

Jian-Long Liu, Zhu Liang,
Chenchen Shen, Xueqin Yin, Wanzhen Wu

Dr. Bee Leng Lee, Dr. Bem Cayco

Sponsored by IBM—Almaden Research Center
In liaison with
Drs. Cheryl Kieliszewski, Peter Haas, and Ignacio Terrizzano

Introduction

How long do you think it takes Ford Motor Company to run one crash simulation?

About 36-160 hours*

Ford Motor Company. *1965 Mustang Fastback With Cammer Engine*. Digital image. *1965 Ford Mustang - 1st Gen 65 Mustangs for Sale & Parts*. Ford Motor Company, n.d. Web. 13 May 2013. <http://www.mustang.com/1st-gen-mustangs/1965-ford-mustang/>.

*Gu, L., "A Comparison of Polynomial Based Regression Models in Vehicle Safety Analysis," in: Diaz, A. (Ed.), ASME Design Engineering Technical Conferences - Design Automation Conference, ASME, Pittsburgh, PA, September 9-12, DAC-21063.

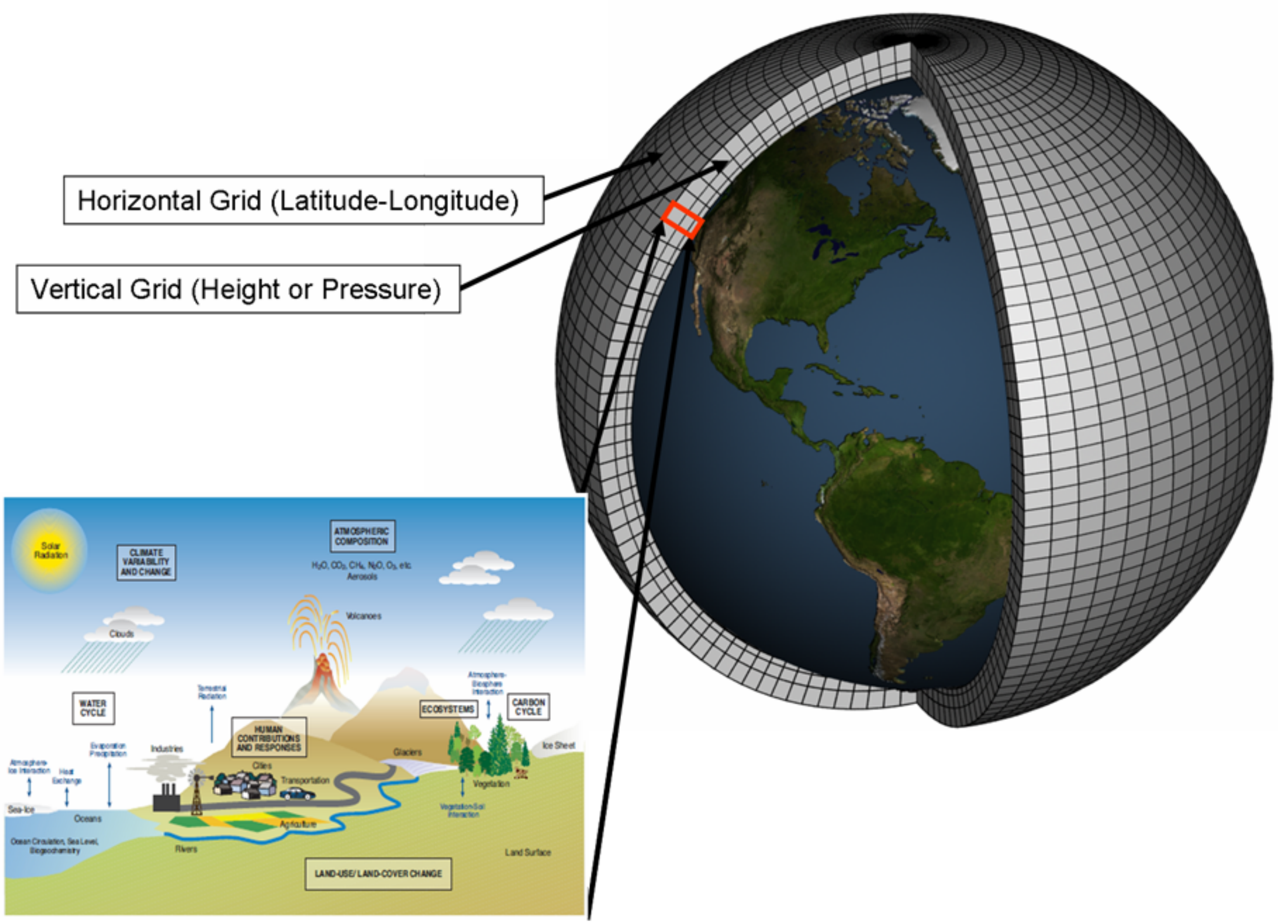
Introduction

Two-variable Optimization Problem

- Assumptions:
  - 50 iterations on average (optimization)
  - One crash simulation each iteration
- Total computation time is 3 to 11 months
- Unacceptable in practice

Brown, Adam. *Crash Testing*. Digital image. *[ESC]ape Testing Crash Test Dummies - Using Realistic Test Data*. N.p., 25 Mar. 2011. Web. 13 May 2013. <http://testing.gobanana.co.uk/?p=743>.

Introduction



NOAA. *Global Climate Model*. Digital image. *File:Global Climate Model.png - Wikipedia, the Free Encyclopedia*. Wikipedia, 18 Feb. 2012. Web. 13 May 2013. <http://en.wikipedia.org/wiki/File:Global_Climate_Model.png>.

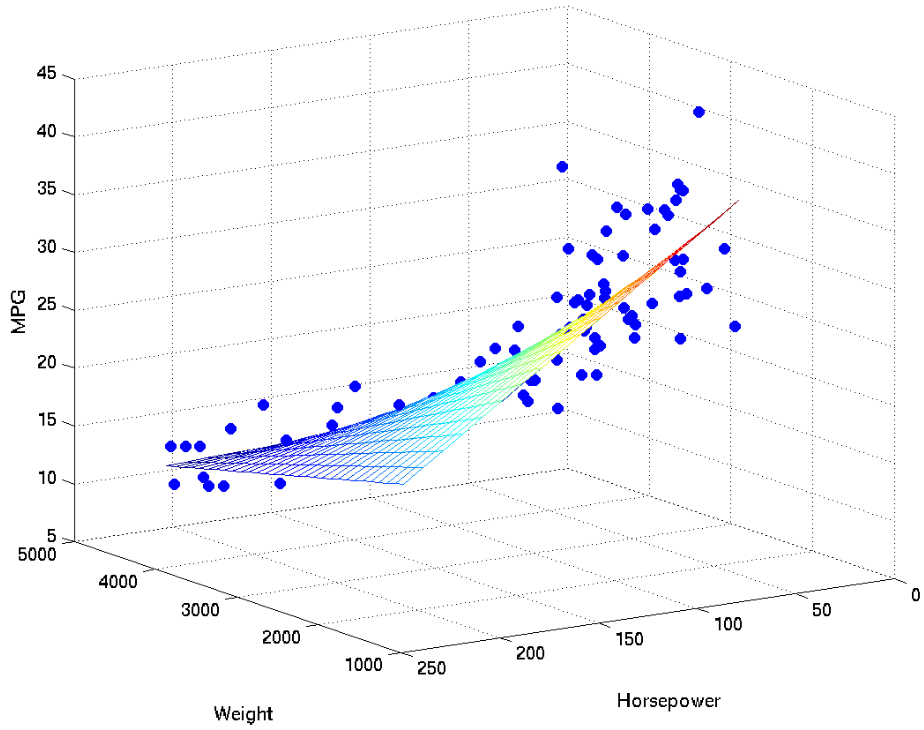
Metamodeling

- Approximation method for time-consuming, costly simulation models

- Approximates computationally intensive functions using simple analytical methods

Regression

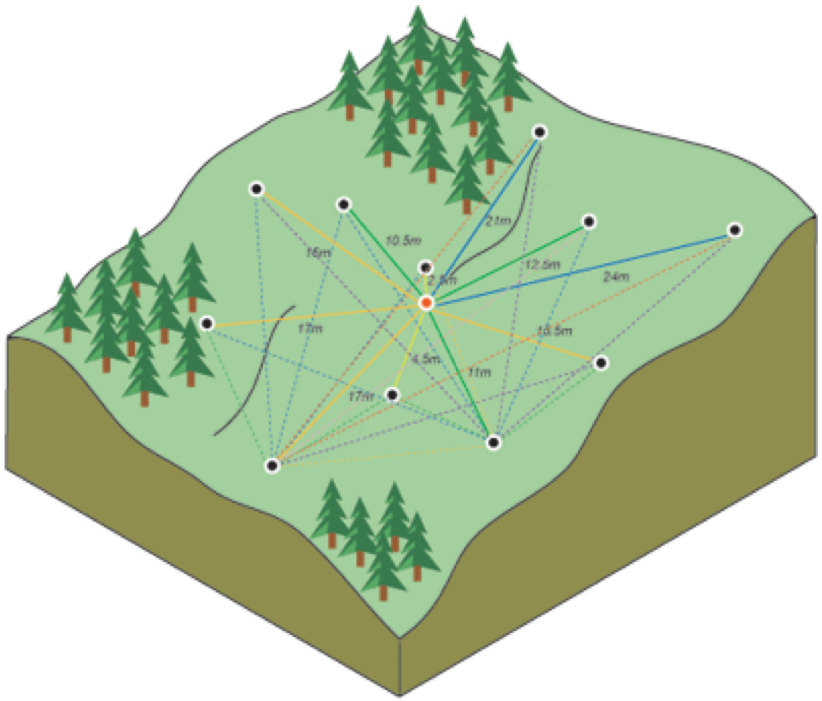
- Four standard assumptions about the random errors ε
  - Zero mean
  - Constant variance
  - Normality
  - Independence
- Accounts for the inherent variability of the data



Linear Regression Surface. Digital image. *Linear Regression and Least Squares Estimation - Statistical Machine Learning and Visualization*. N.p., n.d. Web. 13 May 2013. <http://smlv.cc.gatech.edu/2010/10/06/linear-regression-and-least-squares-estimation/>.

Standard Kriging

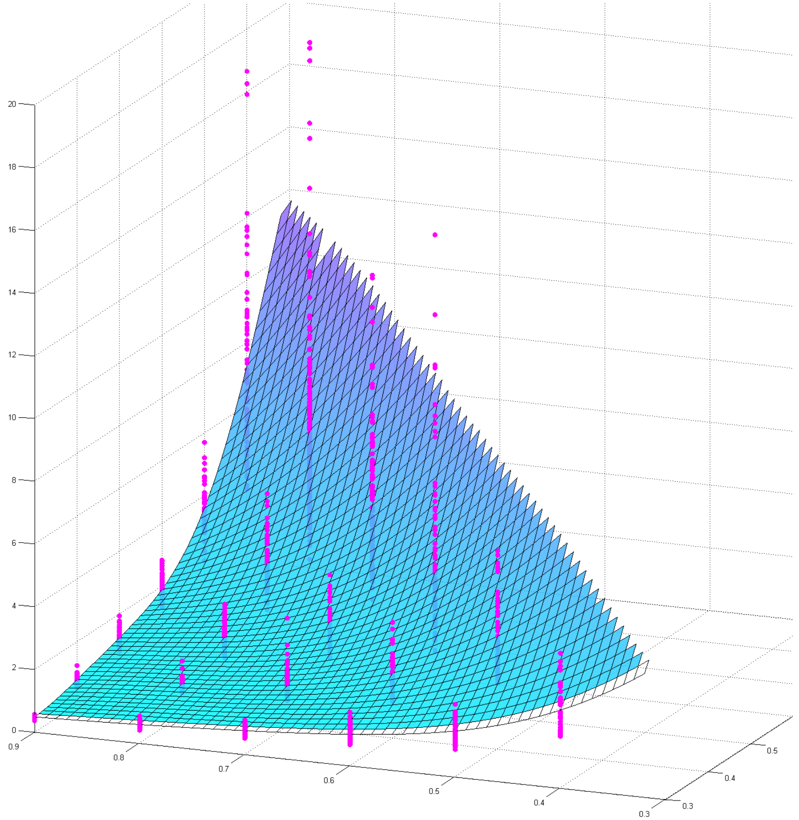
- Originated in geostatistics (i.e. spatial statistics)
- Value at an unknown point approximated by average of the known values at neighbors, weighted by distance
- Accounts for uncertainty about the response surface



ESRI. Calculating the Difference Squared between the Paired Locations. Digital image.ArcGIS Help 10.1 - How Kriging Works. N.p., 8 Nov. 2012. Web. 13 May 2013. <http://resources.arcgis.com/en/help/main/10.1/index.html>.

Stochastic Kriging

- A metamodeling methodology developed for stochastic simulation experiments
- Distinguishes the (extrinsic) uncertainty about the response surface from the (intrinsic) uncertainty inherent in the stochastic simulation

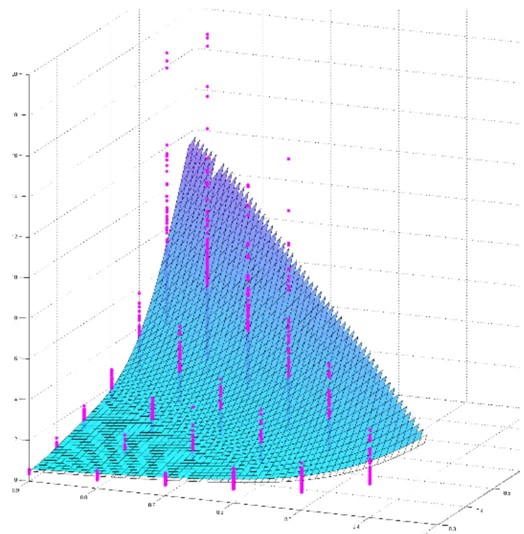


Applications

Coffee Shop

Expected Arrival Rate

Expected Service Rate

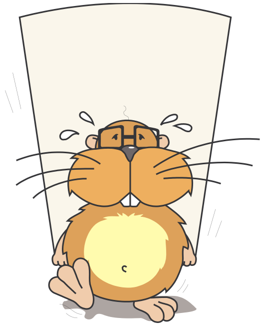


Queue Length

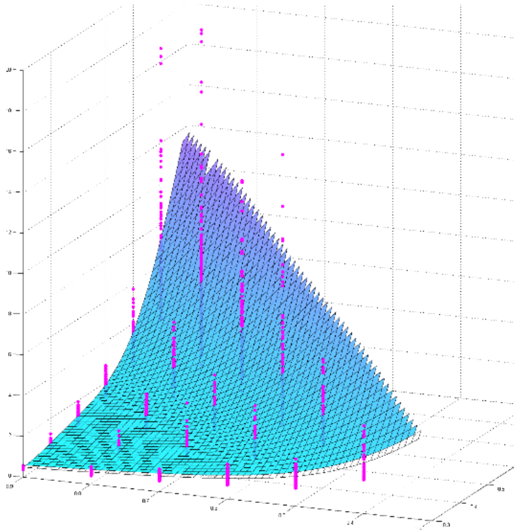
Applications

Call Centers

Staffing Levels



Load Assignments

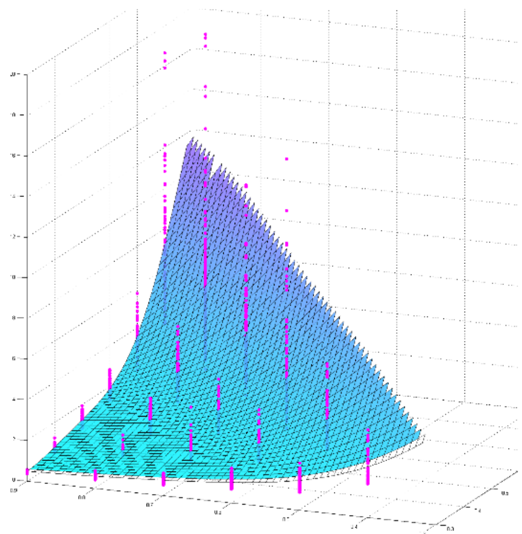


Response-time Performance

Applications

Risk Management

Portfolio Holdings



Risk Measures

General Models

Regression $\mathcal{Y}_j(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} + \boldsymbol{\varepsilon}_j(\mathbf{x})$

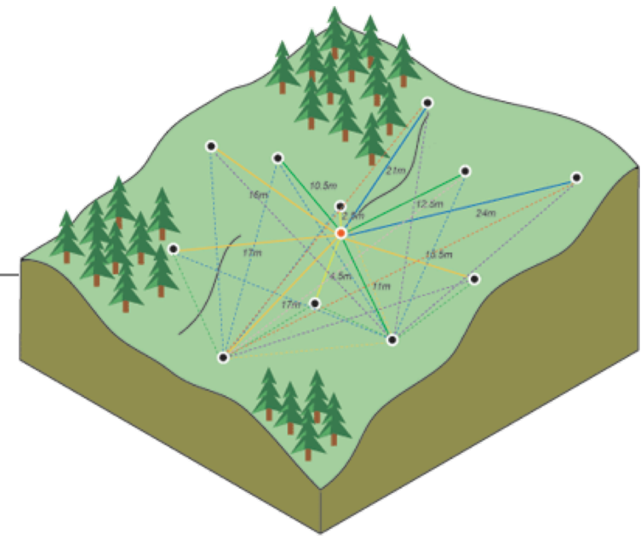
Standard kriging $\mathcal{Y}_j(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} + \mathrm{M}(\boldsymbol{x})$

Stochastic kriging $\mathcal{Y}_j(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} + \mathrm{M}(\mathbf{x}) + \boldsymbol{\varepsilon}_j(\mathbf{x})$

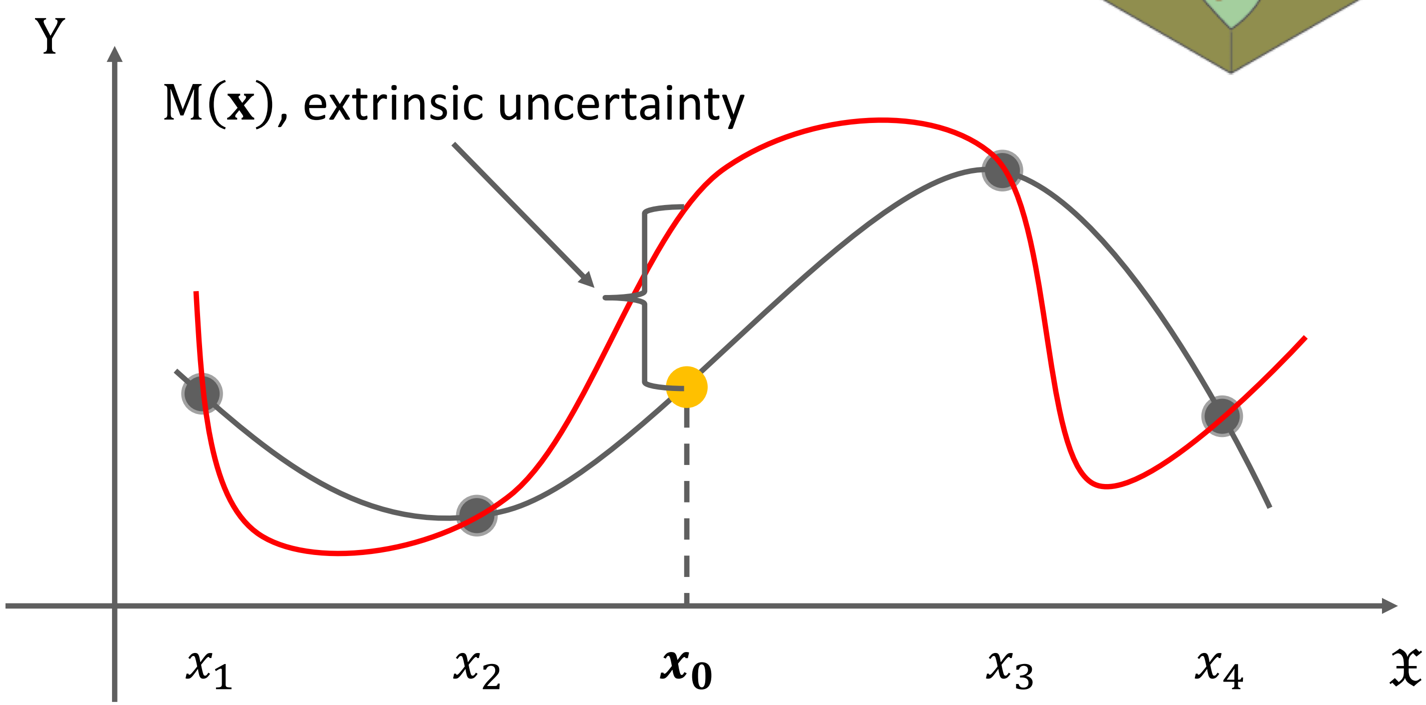
$\boldsymbol{\varepsilon}_j(\mathbf{x})$ intrinsic uncertainty

$\mathrm{M}(\mathbf{x})$ extrinsic uncertainty

Standard Kriging

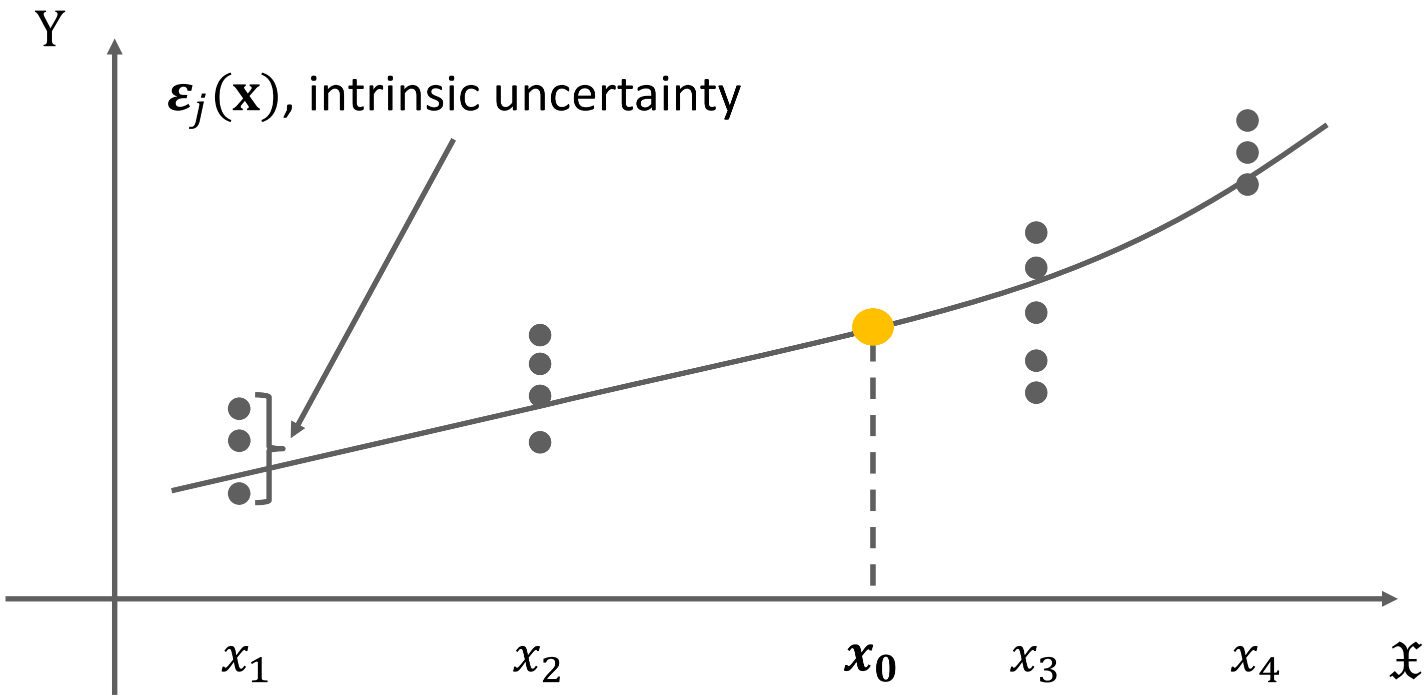


$$\mathcal{Y}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{M}(\mathbf{x})$$



Stochastic Kriging

$$\mathcal{Y}_j(\mathbf{x}) = \mathbf{f}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{M}(\mathbf{x}) + \boldsymbol{\varepsilon}_j(\mathbf{x})$$



MSE-Optimal Predictor

Suppose that all parameters are known

$$\widehat{\mathrm{Y}}(\mathbf{x}_0) = \beta_0 + \boldsymbol{\Sigma}_{\mathrm{M}}(\mathbf{x}_0,\cdot)^{\mathrm{T}}[\boldsymbol{\Sigma}_{\mathrm{M}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}]^{-1}(\bar{\mathcal{Y}} - \beta_0 \mathbf{1}_{\mathrm{k}})$$

β_0 overall response mean

$\bar{\mathcal{Y}}$ average response

$\boldsymbol{\Sigma}_{\mathrm{M}}$ extrinsic covariance matrix

$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ intrinsic covariance matrix

Assumptions

- $\mathrm{M}(\mathbf{x})$ is a stationary Gaussian random field
  - Constant mean 0
  - Constant variance τ^2
  - $\boldsymbol{\Sigma}_{\mathrm{M}} = \tau^2 \exp\left(-\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta},2}^2\right)$
- $\boldsymbol{\varepsilon}_j(\mathbf{x})$ is $N\left(0, V(\mathbf{x})\right)$

Parameter Estimation

Estimation of Predictor

- $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$, β_0, τ^2, and $\boldsymbol{\theta}$
- Variances not observable, even at design points
- Estimate with sample variances
- Covariance matrix of diagonals (i.i.d. of $\boldsymbol{\varepsilon}_j$)
- Use maximum likelihood estimator for rest

Likelihood Function

- Function of parameters
- Likelihood of observing given outputs for a set of parameters
- Complementary to probability function
- Higher likelihood is better

Gummi Bears. Digital image. *Episode 49: It Was A Gummy Bear | Mattandmondo*. N.p., 29 May 2012. Web. 13 May 2013. <http://www.mattandmondo.com/podcast/archives/438>.

Likelihood Function

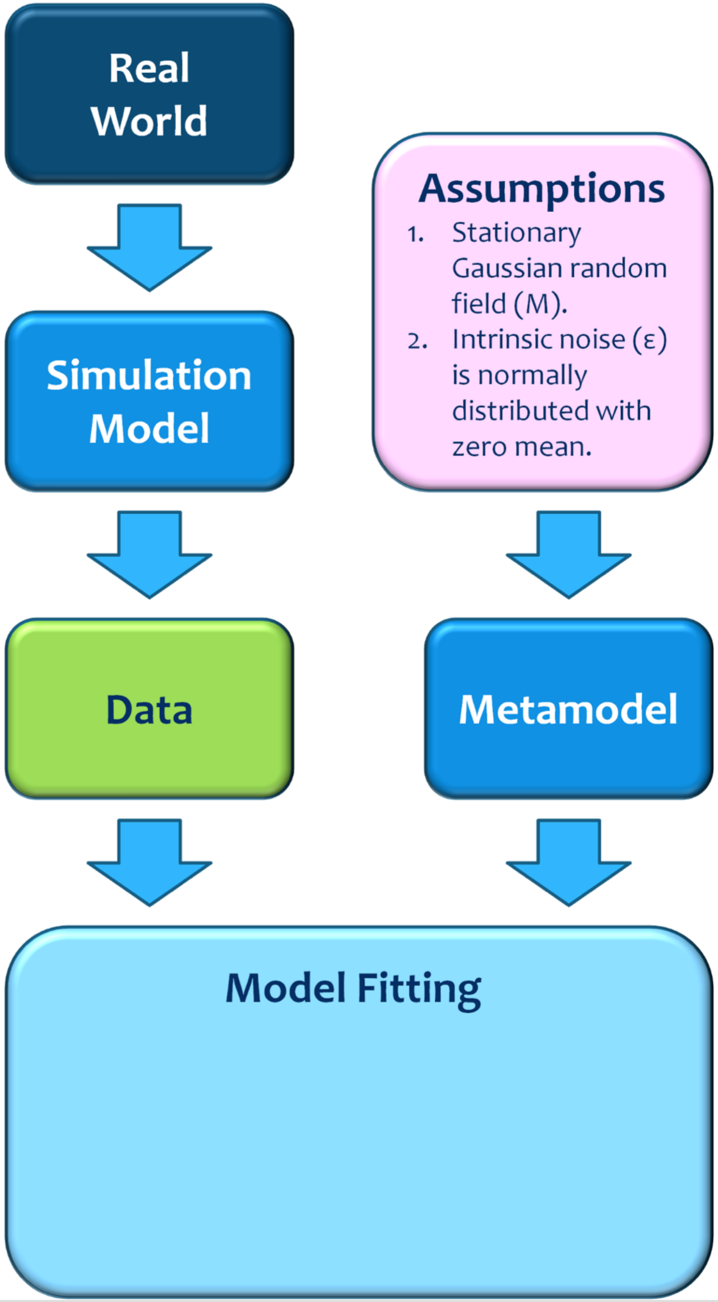
- Function of parameters
- Likelihood of observing given outputs for a set of parameters
- Complementary to probability function
- Higher likelihood is better

Nonlinear Optimization

- Find the combination of parameters to maximize the likelihood function for our predictor

- R package MLEGP
  - "Maximum Likelihood Estimates of Gaussian Processes"

- Plug resulting parameters into predictor


Real World
Simulation Model
Data
Assumptions
1. Stationary Gaussian random field (M).
2. Intrinsic noise (ε) is normally distributed with zero mean.
Metamodel
Model Fitting

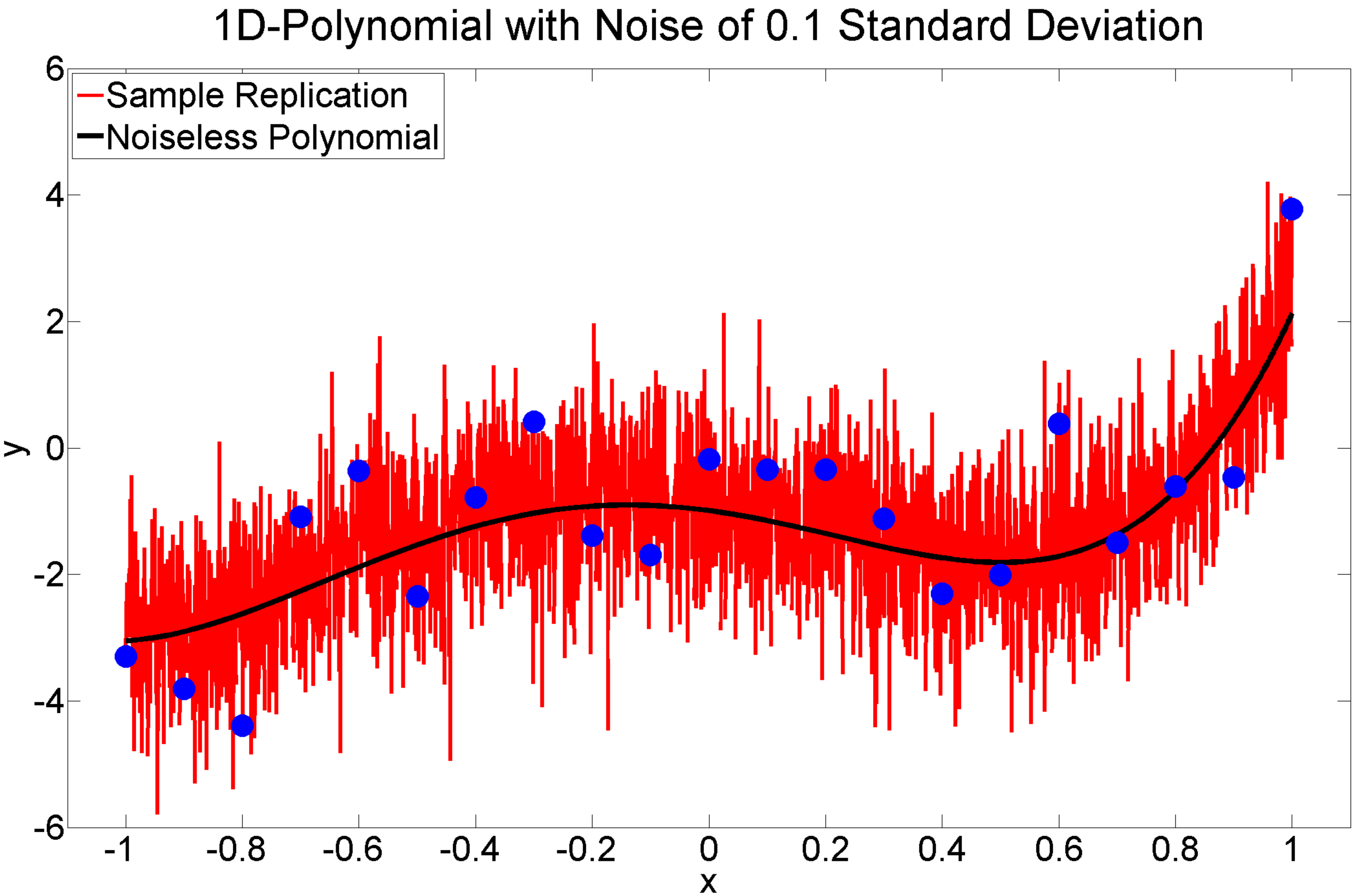
BREAK

Preliminary Results

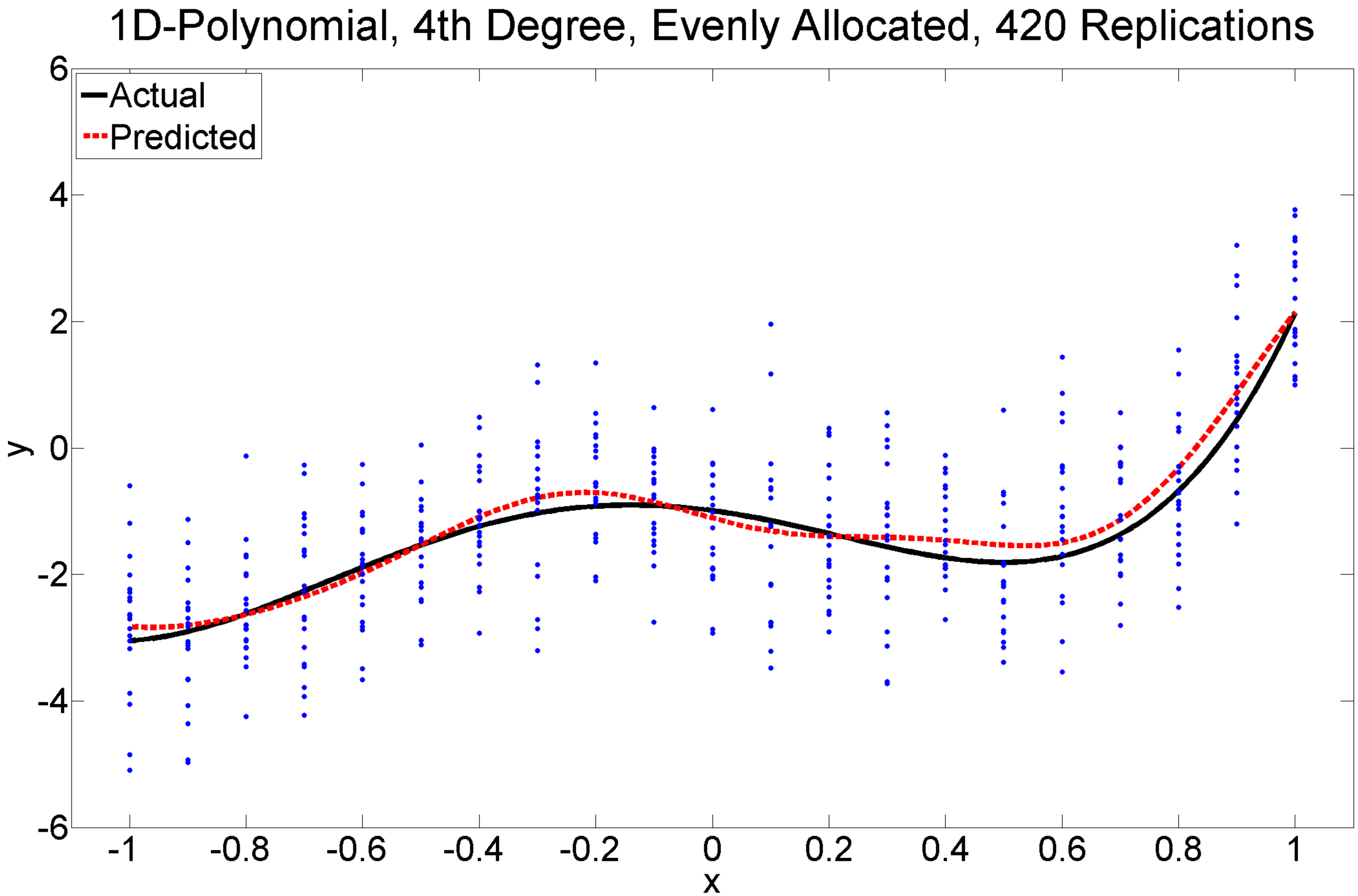
Polynomials

- Simple to use as test case
- Can test as high-dimensional as we want
- Evenly distributed noise

Sample Replication

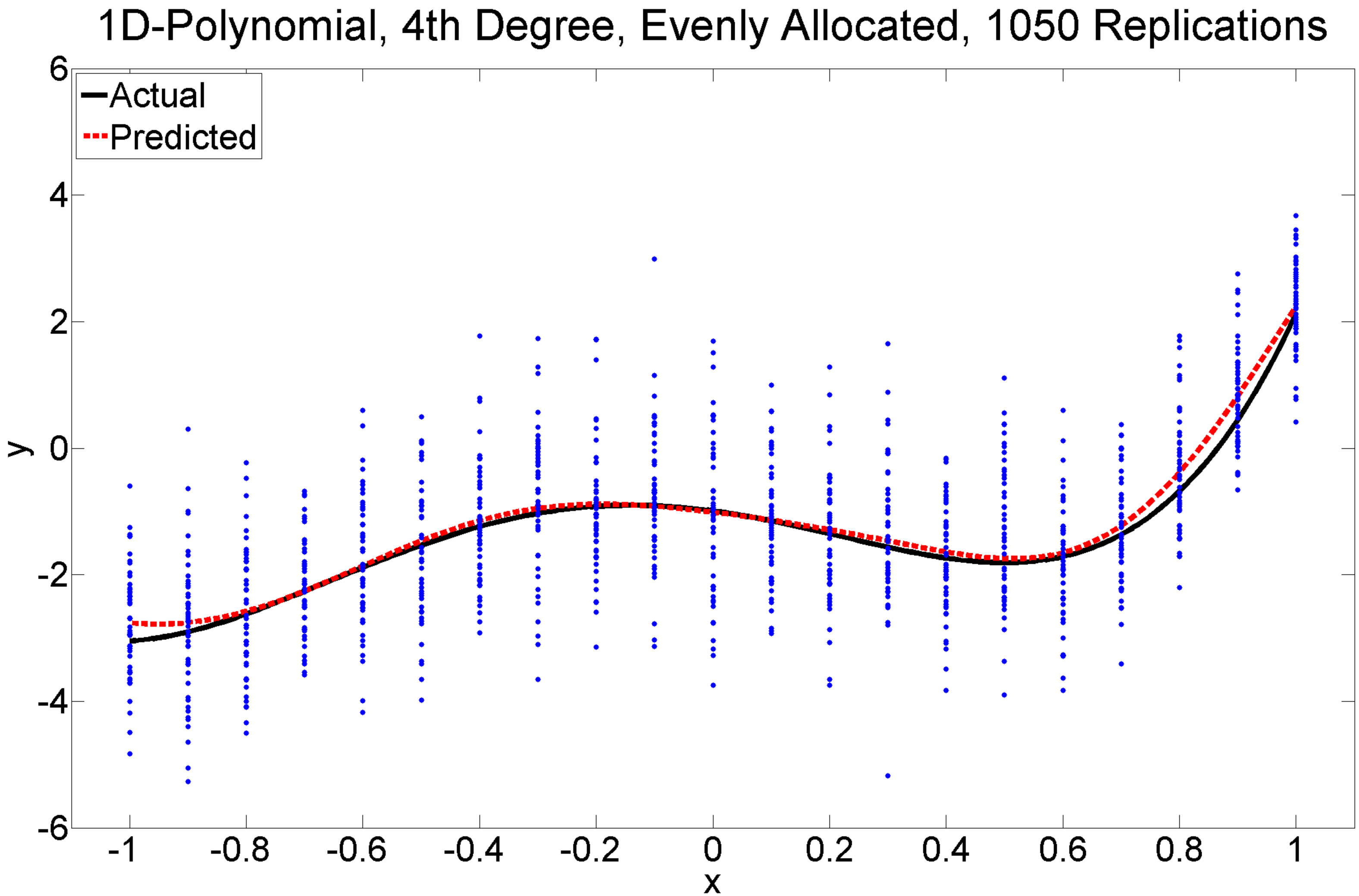


Results



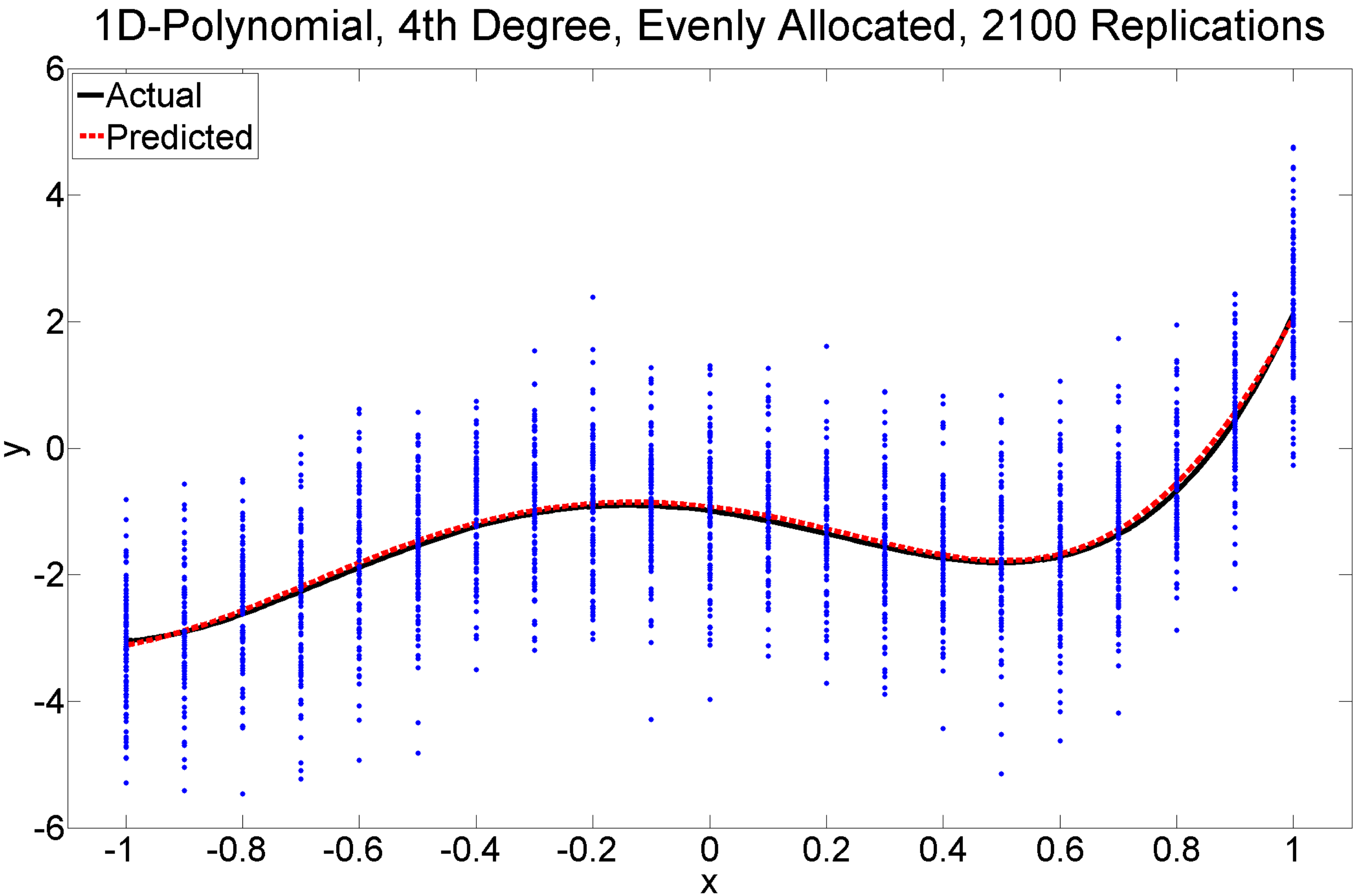
1D-Polynomial, 4th Degree, Evenly Allocated, 420 Replications
Actual
Predicted
y
x
6
4
2
0
-2
-4
-6
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

Results



1D-Polynomial, 4th Degree, Evenly Allocated, 1050 Replications
Actual
Predicted
6
4
2
0
-2
-4
-6
y
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
x

Results

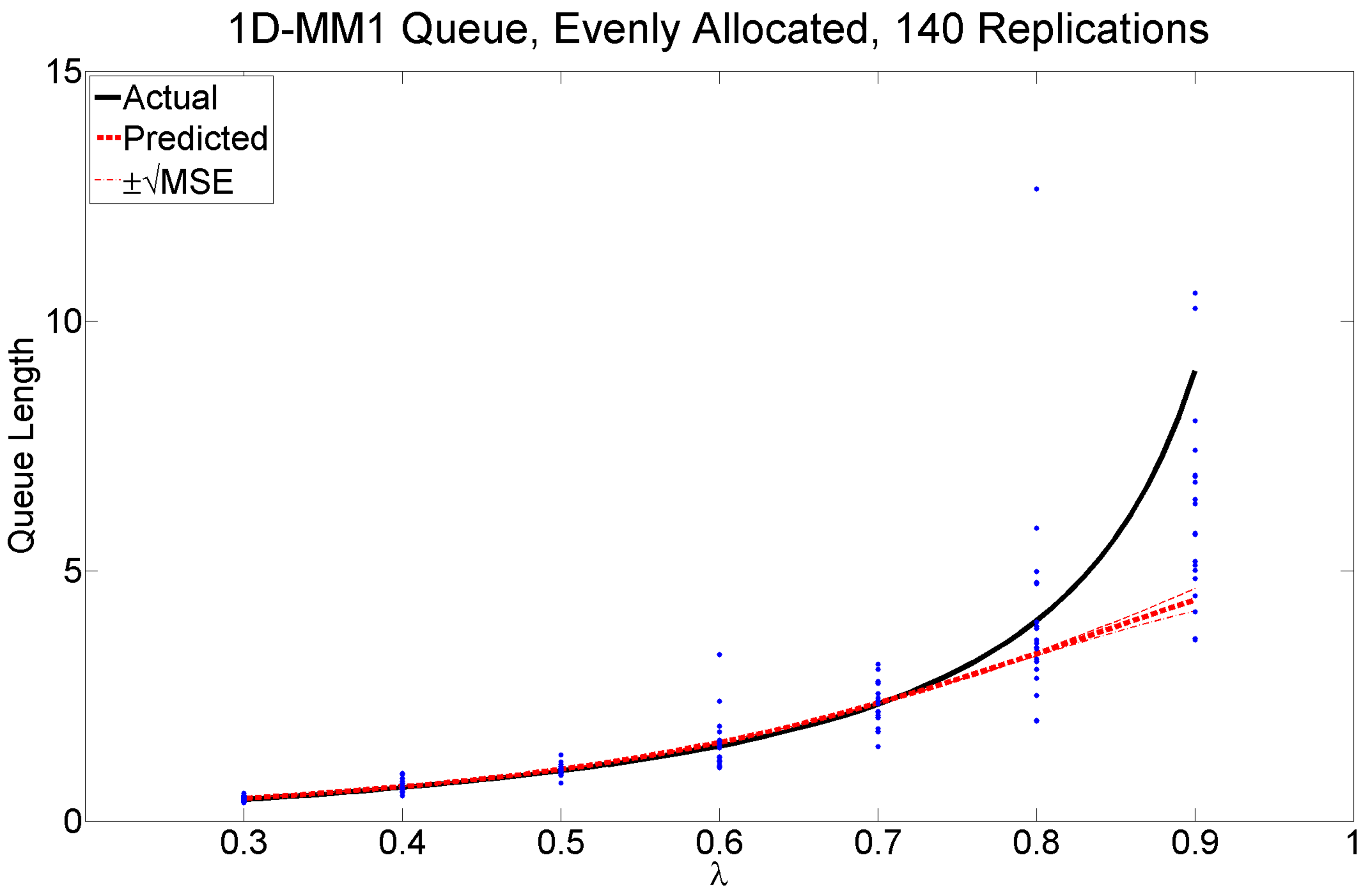

1D-Polynomial, 4th Degree, Evenly Allocated, 2100 Replications
Actual
Predicted
6
4
2
0
-2
-4
-6
y
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
x

Queuing Model

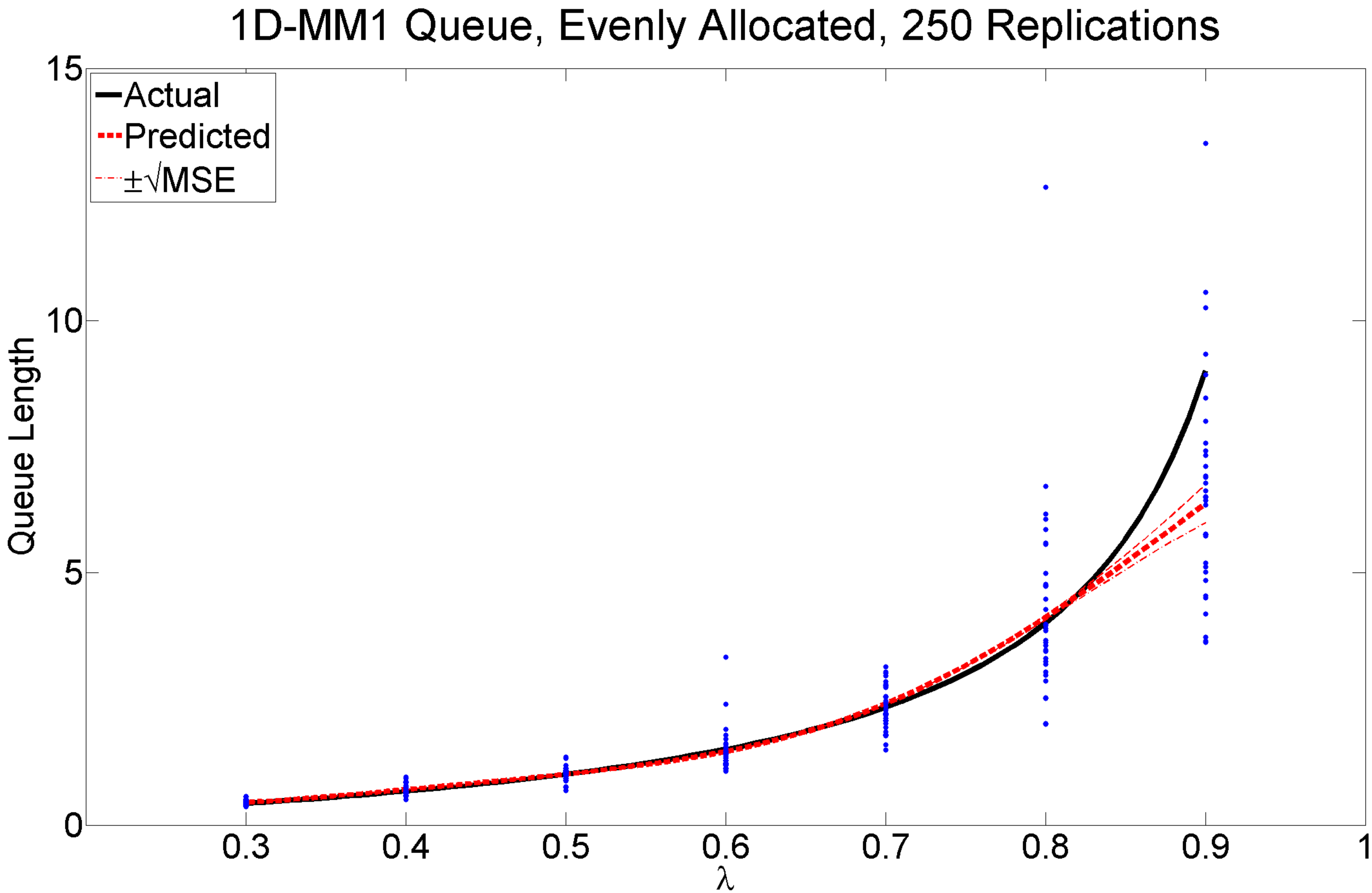
MM1 Queue

- Single server, single queue
- λ is expected arrival rate, μ is expected service rate
- Expected queue length is $\frac{\lambda}{\mu - \lambda}$
- Assume no trend
- Assume $0 < \lambda < \mu$
- Average queue length

Results

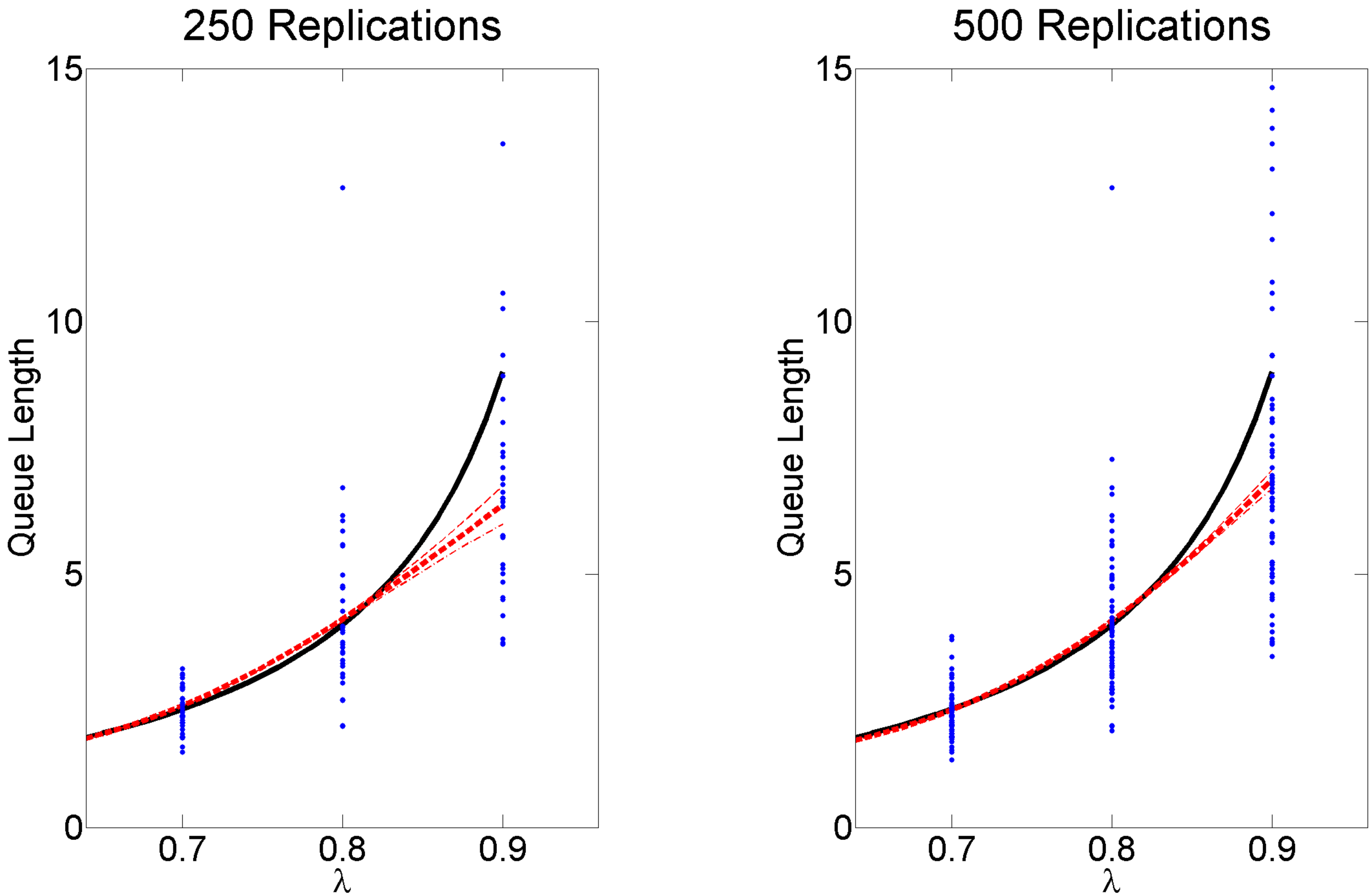

1D-MM1 Queue, Evenly Allocated, 140 Replications
Actual
Predicted
±√MSE
Queue Length
λ
15
10
5
0
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Results



1D-MM1 Queue, Evenly Allocated, 250 Replications
Actual
Predicted
±√MSE
Queue Length
λ
0
5
10
15
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Results—Comparison



Experimental Design

Experimental Design

- Goal is to minimize $IMSE(\mathbf{n}) = \int_{\mathbf{x}_0 \in \mathfrak{X}} MSE(\mathbf{x}_0; \mathbf{n})\, d\mathbf{x}_0$
- $\mathfrak{X} \subseteq \mathbb{R}^d$ is the experimental design space
- k is the number of fixed design points
- $\mathbf{n}^T = (n_1, n_2, \dots, n_k)$
- $n_i^* = n_i^*\big(N, V(\mathbf{x}_1), \dots, V(\mathbf{x}_k), \boldsymbol{\Sigma}_\mathrm{M}, \mathbf{r}(\mathbf{x}_0)\big)$

Two-Stage Design

Stage 1

- Select m predetermined design points $\mathbf{x}_1, \dots, \mathbf{x}_m$ and allocate n_0 replications to each $\mathbf{x}_i$

- Estimate V and $\mathbf{\Sigma}_\mathrm{M}$
  - V can be estimated by standard kriging method $V(\mathbf{x}) = \boldsymbol{\sigma}^2 + Z(\mathbf{x})$
  - $\mathbf{\Sigma}_\mathrm{M} = \tau^2 \exp\left(-\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta},2}^2\right)$

Two-Stage Design

Stage 2

- Jointly select $k - m$ additional design points
- Optimally allocate $N - mn_0$ additional replications among all design points
- $n_i^* = n_i^*\left(N, V(\mathbf{x}_1), \ldots, V(\mathbf{x}_k), \mathbf{\Sigma}_\mathrm{M}, \mathbf{r}(\mathbf{x}_0)\right)$

Algorithm

Algorithm

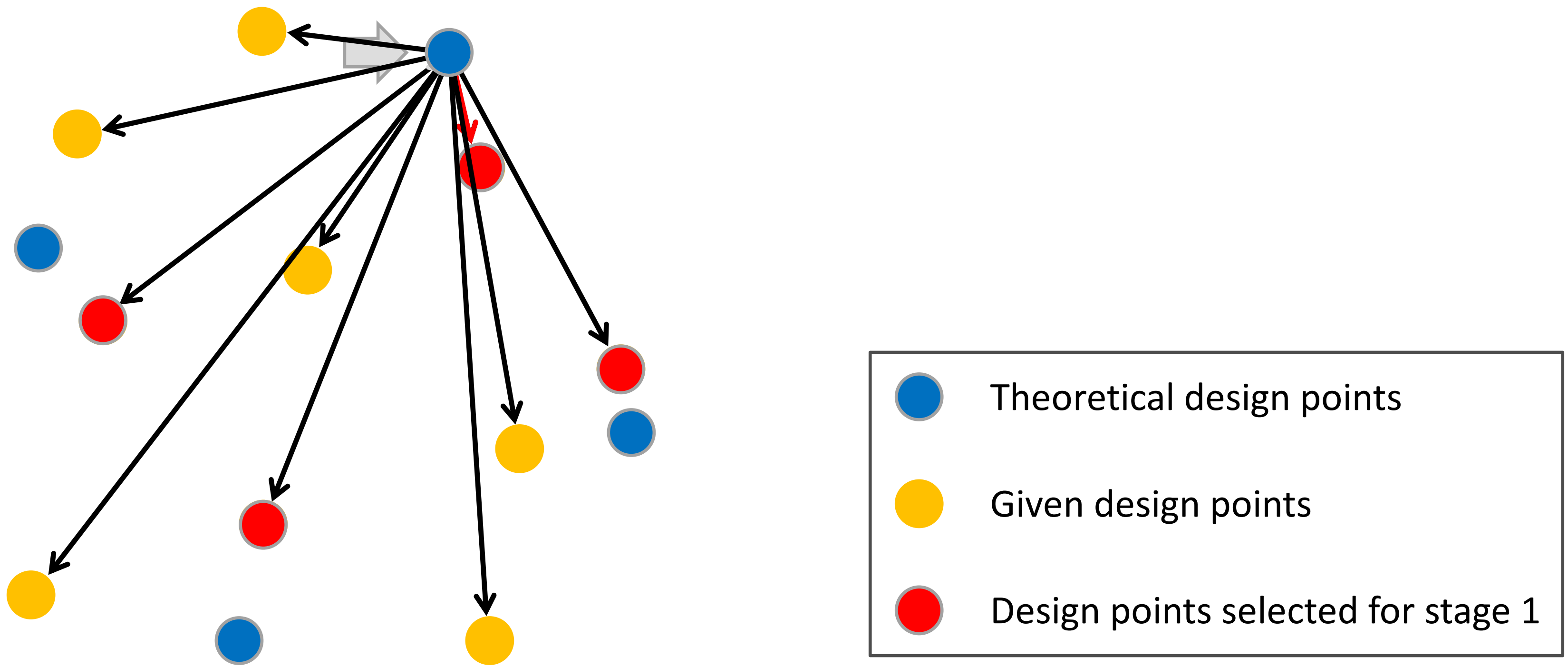
Stage 1

- Generate m points from Latin Hypercube Sampling

- Calculate distances of given design points and theoretical points

- Choose the design points closest to the theoretical points

Algorithm


Theoretical design points
Given design points
Design points selected for stage 1

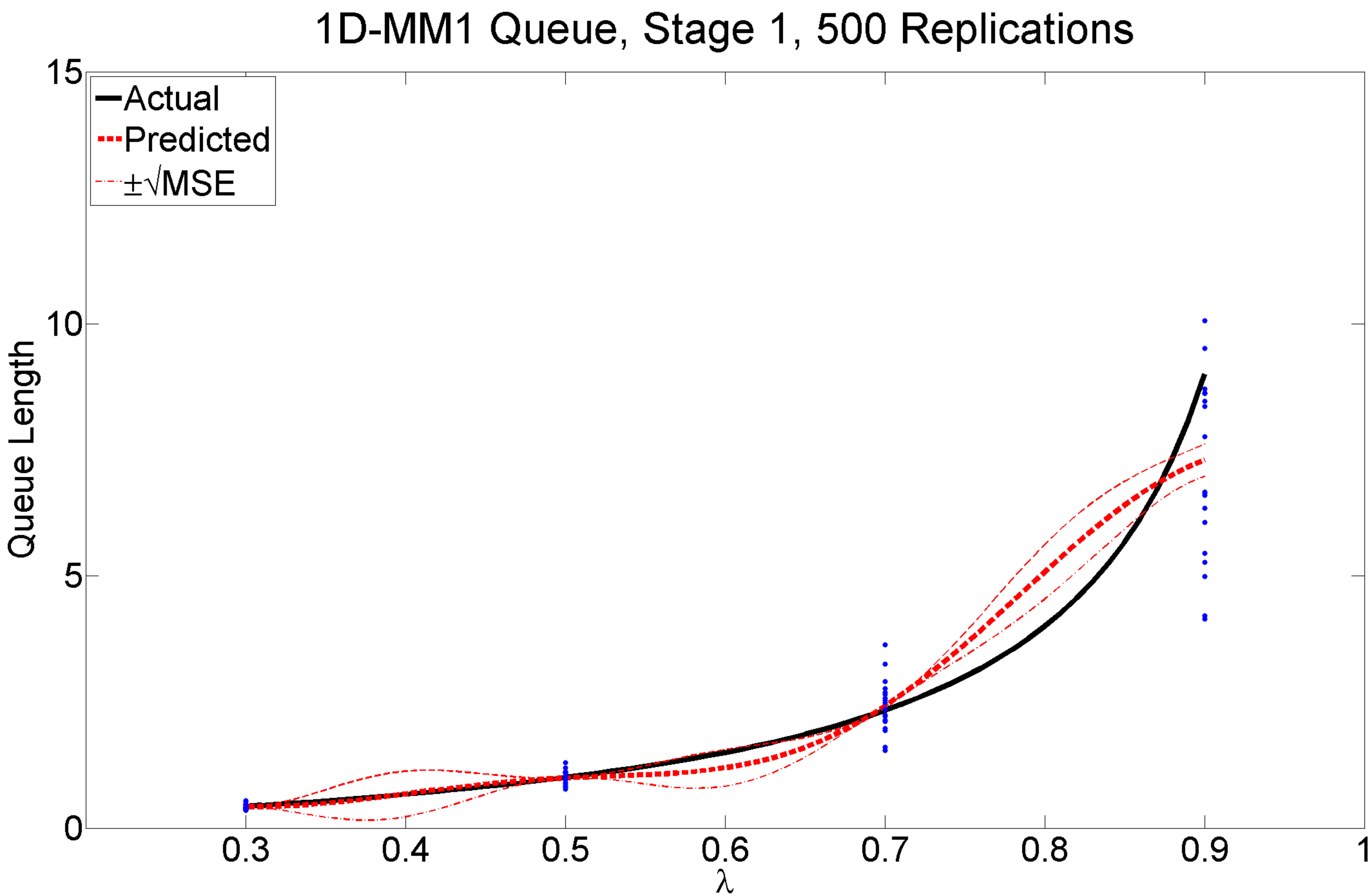
Algorithm

Stage 2

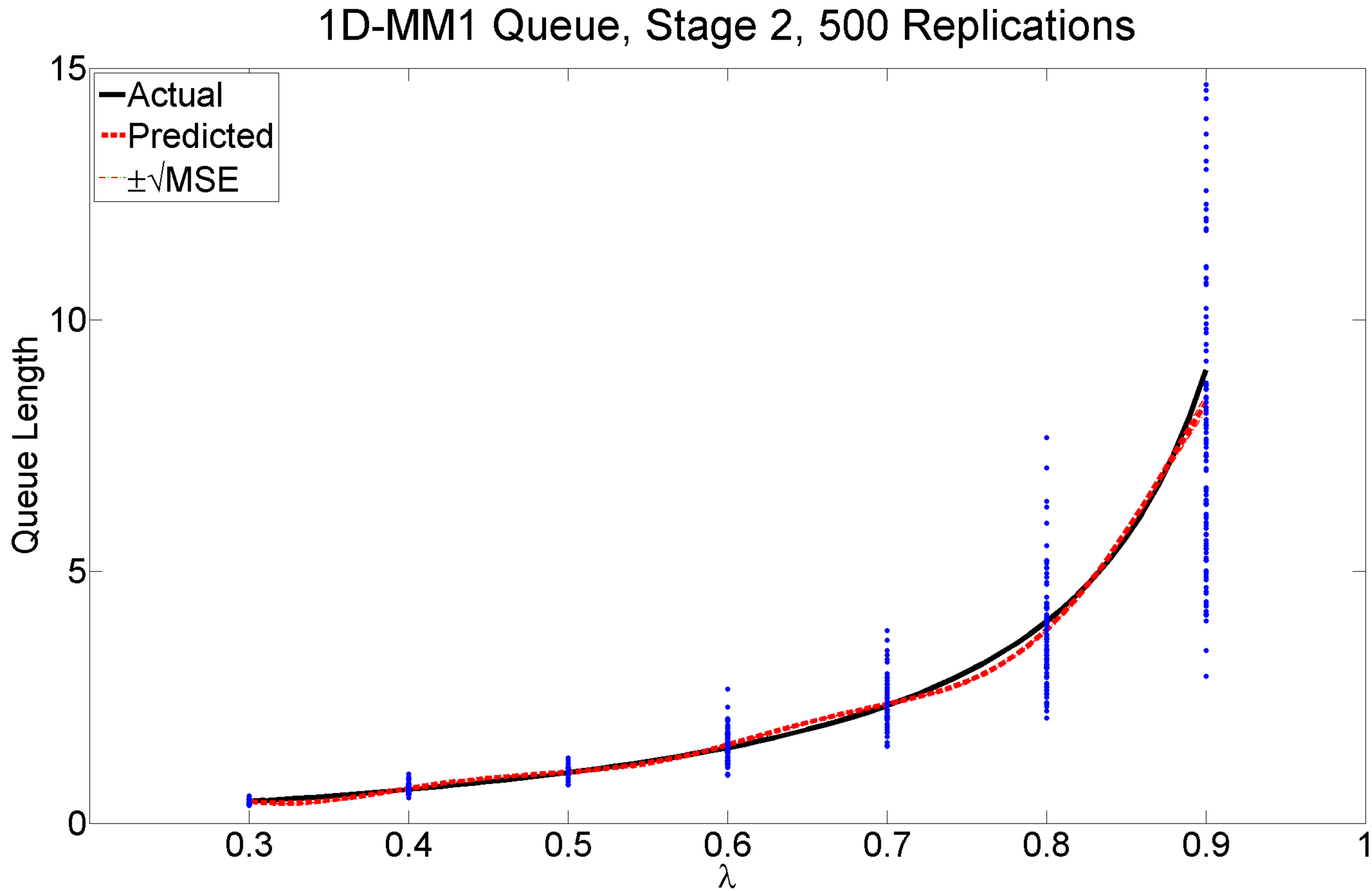
- Simulate stage 1 data with m design points

- Allocate optimal replications to all design points

Results

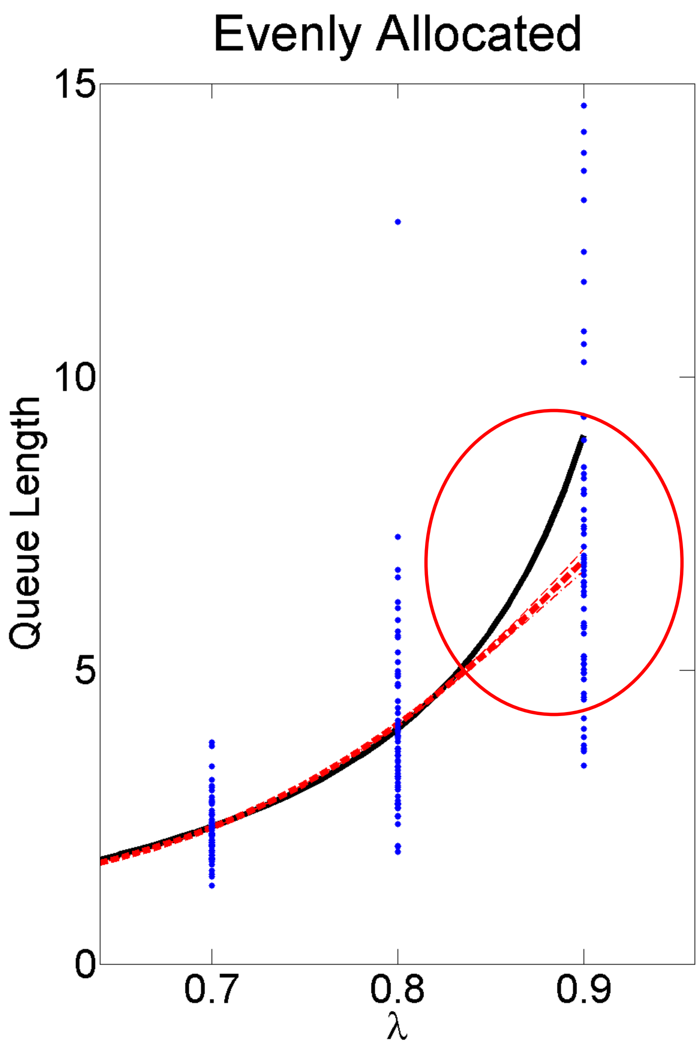

1D-MM1 Queue, Stage 1, 500 Replications
Actual
Predicted
±√MSE
Queue Length
λ
15
10
5
0
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

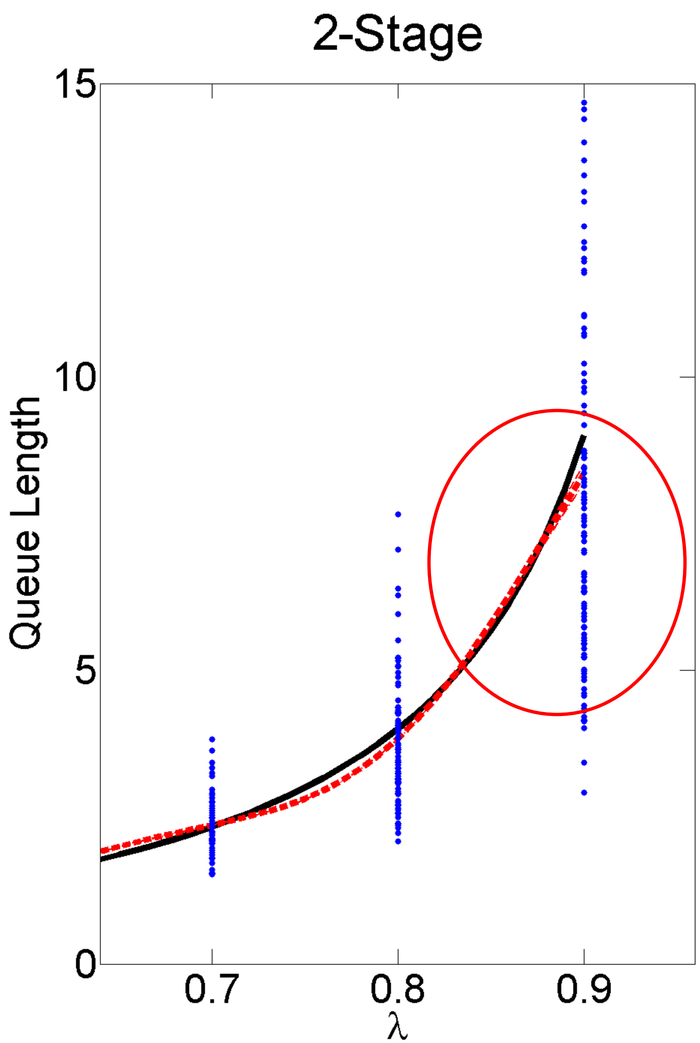
Results



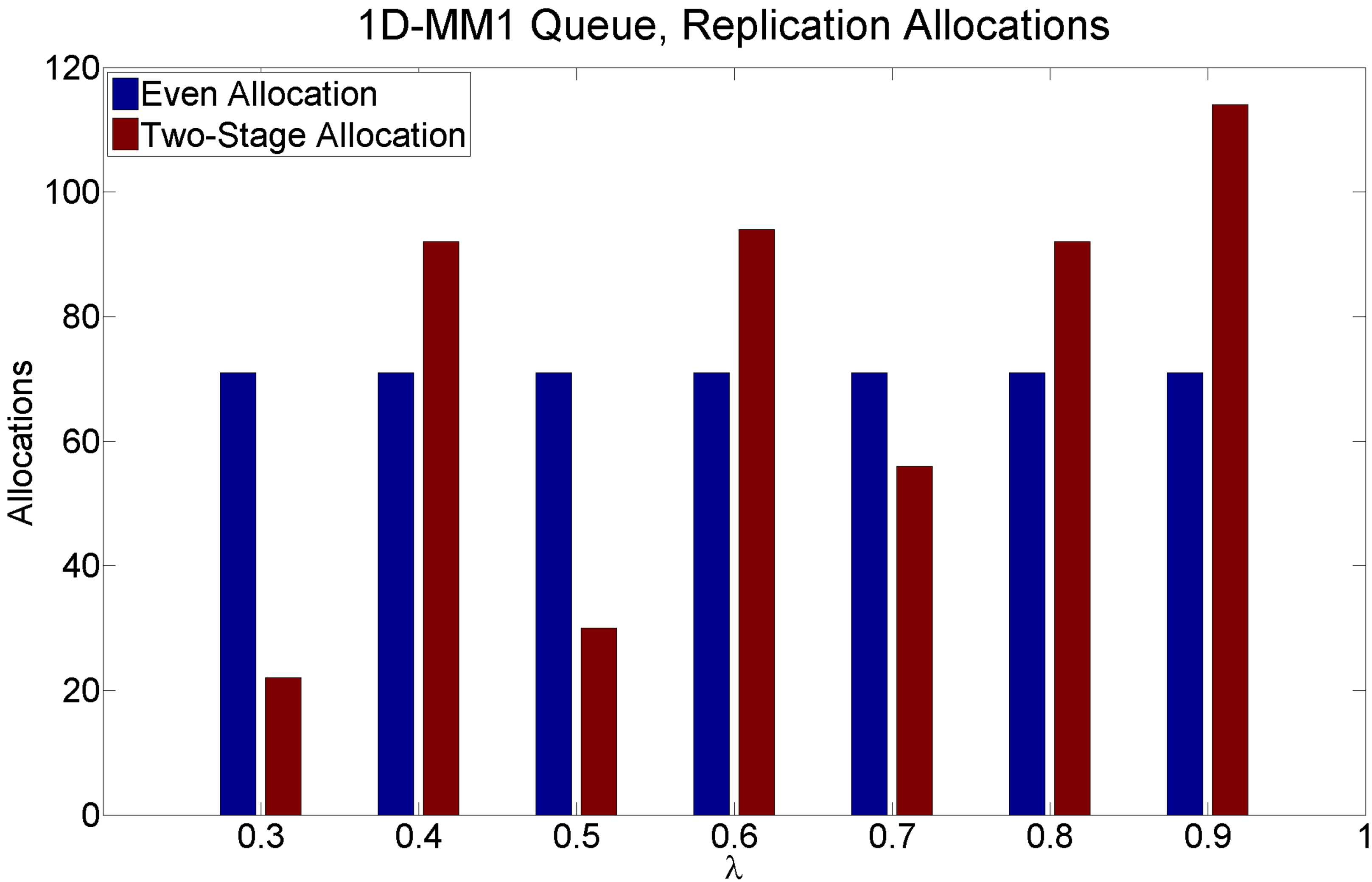
1D-MM1 Queue, Stage 2, 500 Replications
Actual
Predicted
±√MSE
Queue Length
λ
0
5
10
15
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Results—Comparison





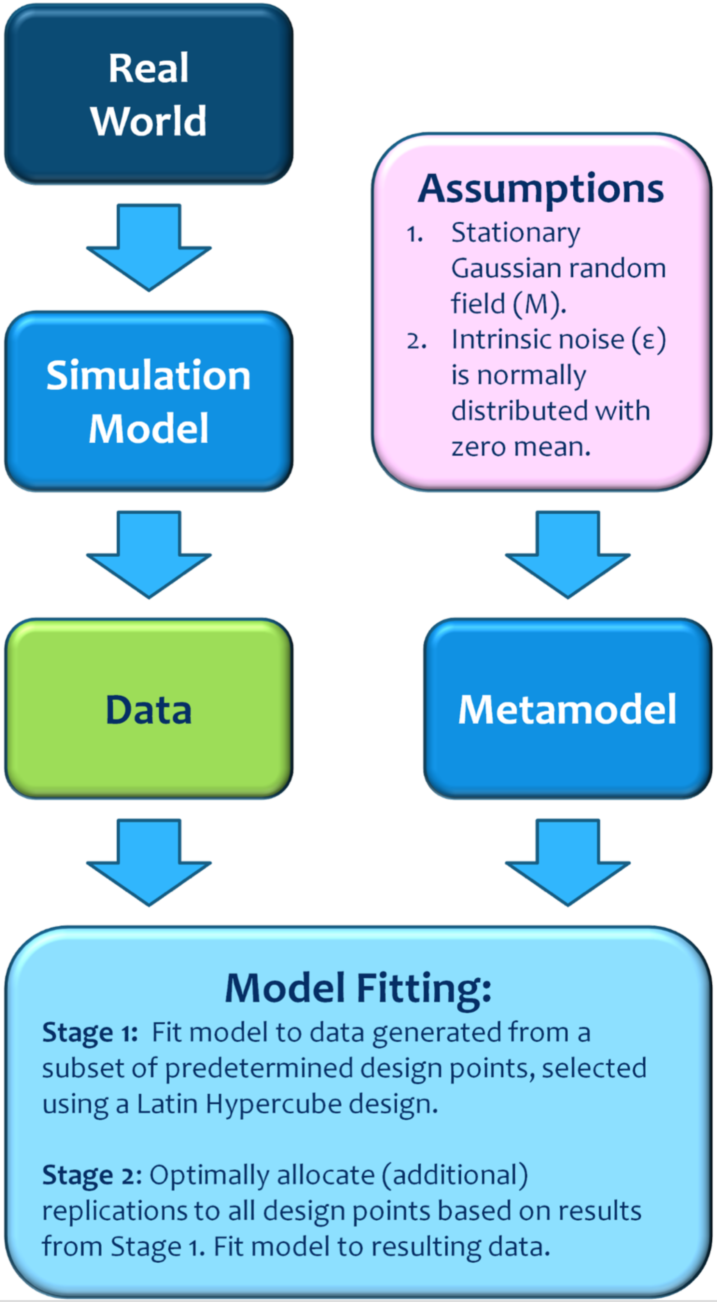
Results—Allocations



Simulation Effort

- How many replications to allocate in stage 1?
  - Too few implies inadequate estimation in τ^2, β_0, and V
  - Too many implies reduced advantage of 2-stage procedure

- How many design points to pick?
  - Depends on structure of simulation model



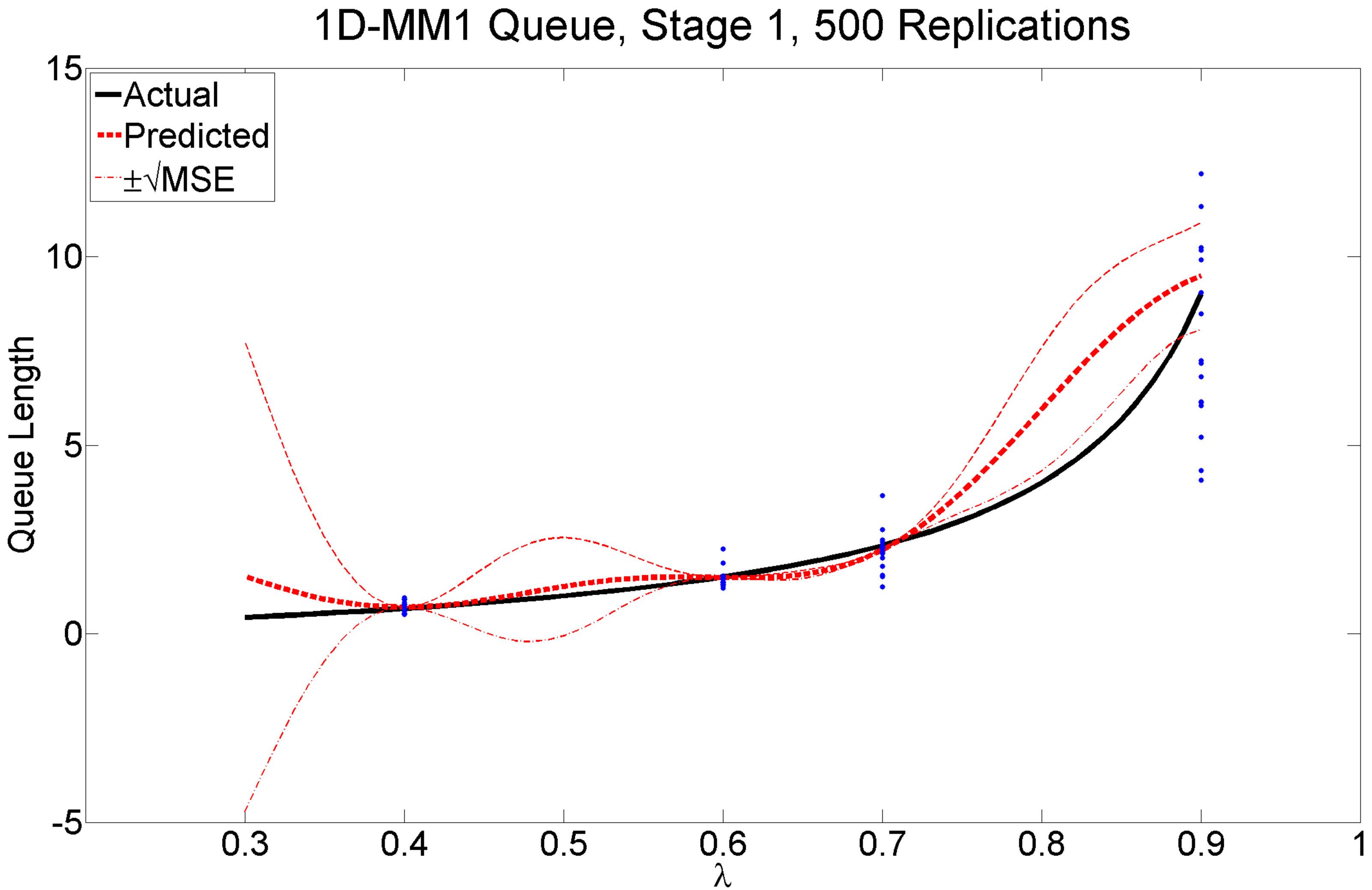
Real World
Simulation Model
Data
Assumptions
1. Stationary Gaussian random field (M).
2. Intrinsic noise (ε) is normally distributed with zero mean.
Metamodel
Model Fitting:
Stage 1: Fit model to data generated from a subset of predetermined design points, selected using a Latin Hypercube design.
Stage 2: Optimally allocate (additional) replications to all design points based on results from Stage 1. Fit model to resulting data.

Problems and Future Work

Problems

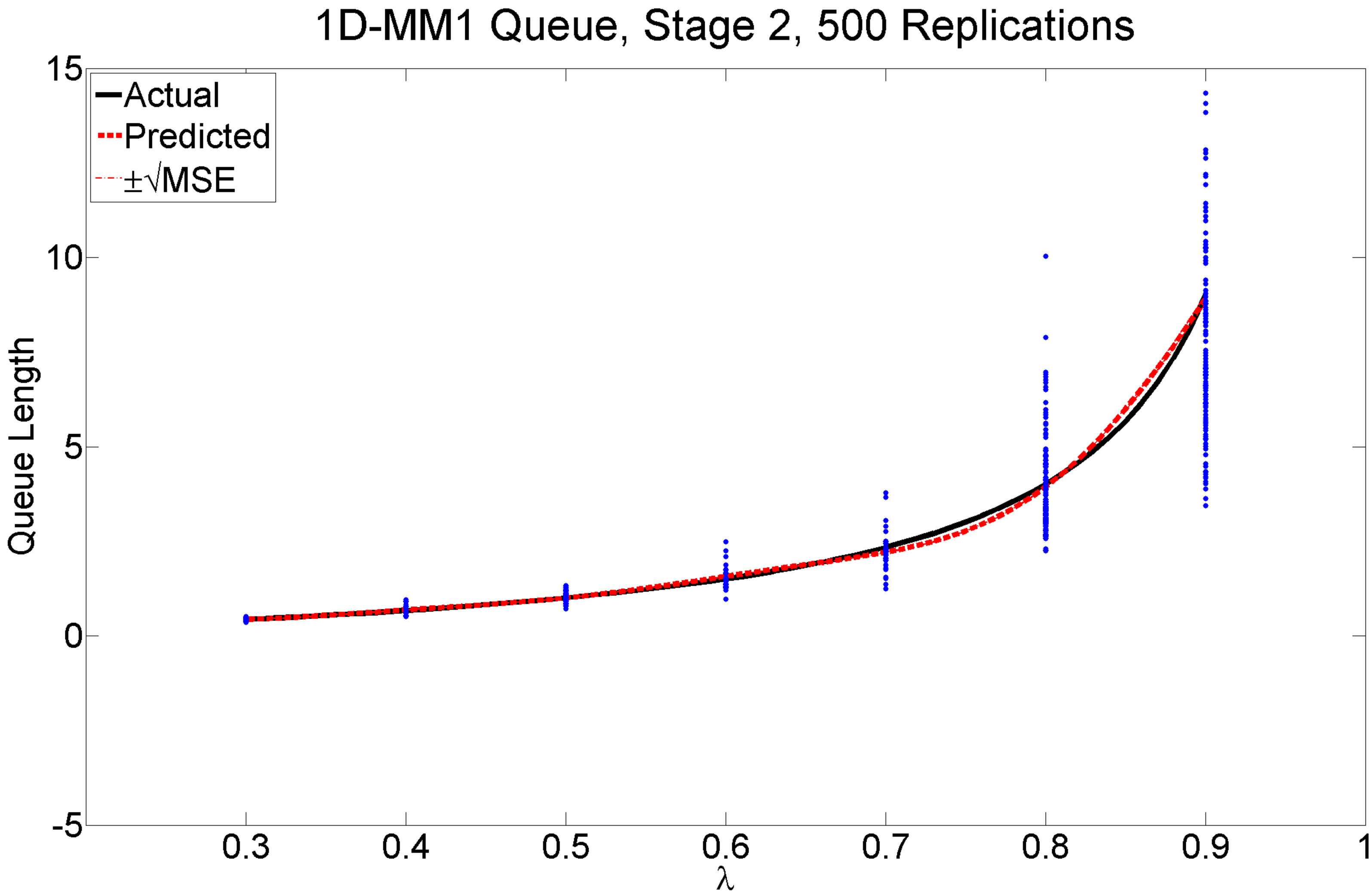
- Nonidentical simulation output at each design point
- Estimated variance may end up nonpositive
- Overestimated MSE and variance in stage 1
- Bumps

Problems



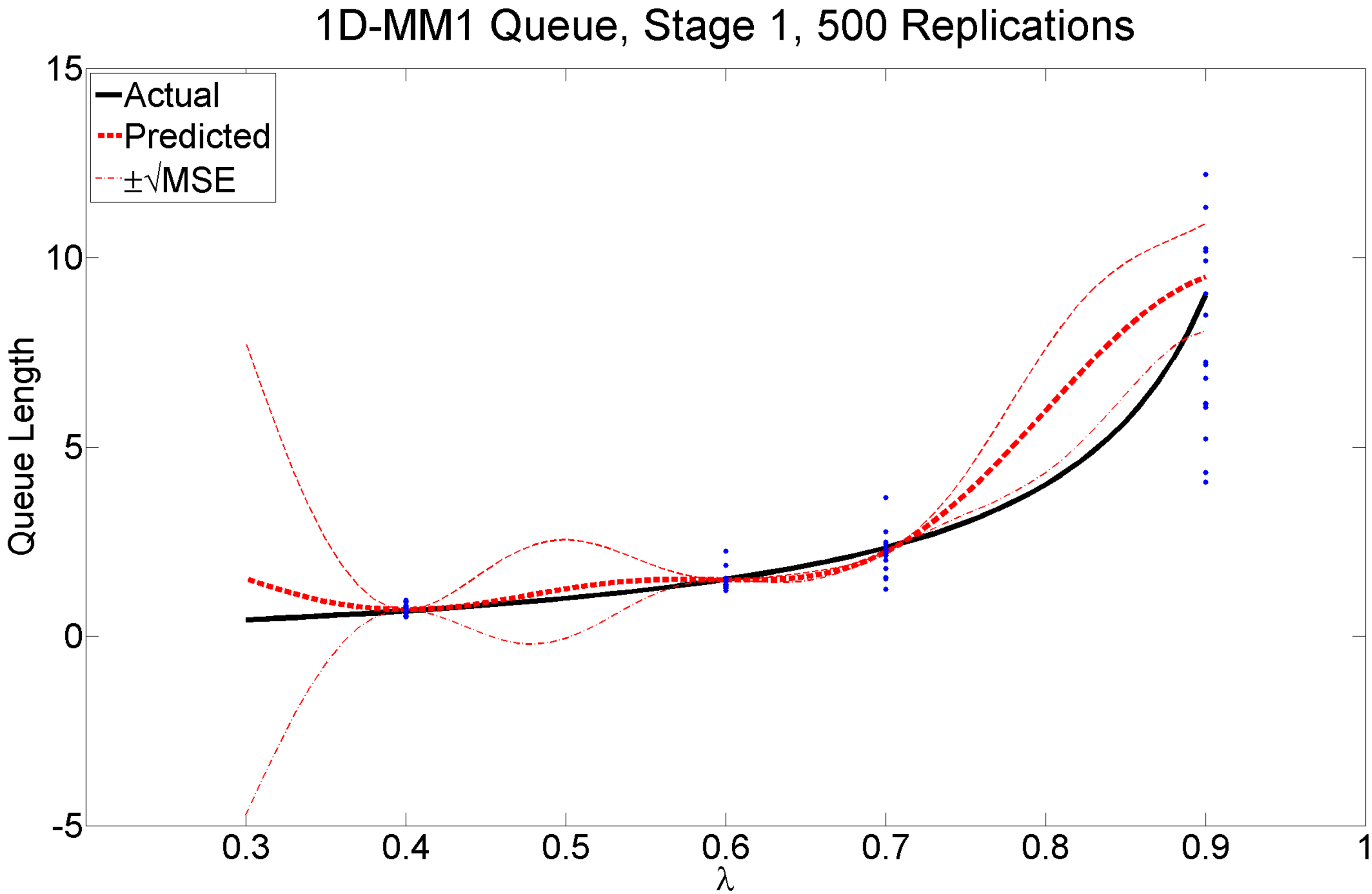
1D-MM1 Queue, Stage 1, 500 Replications
Actual
Predicted
±√MSE
Queue Length
λ
15
10
5
0
-5
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Problems



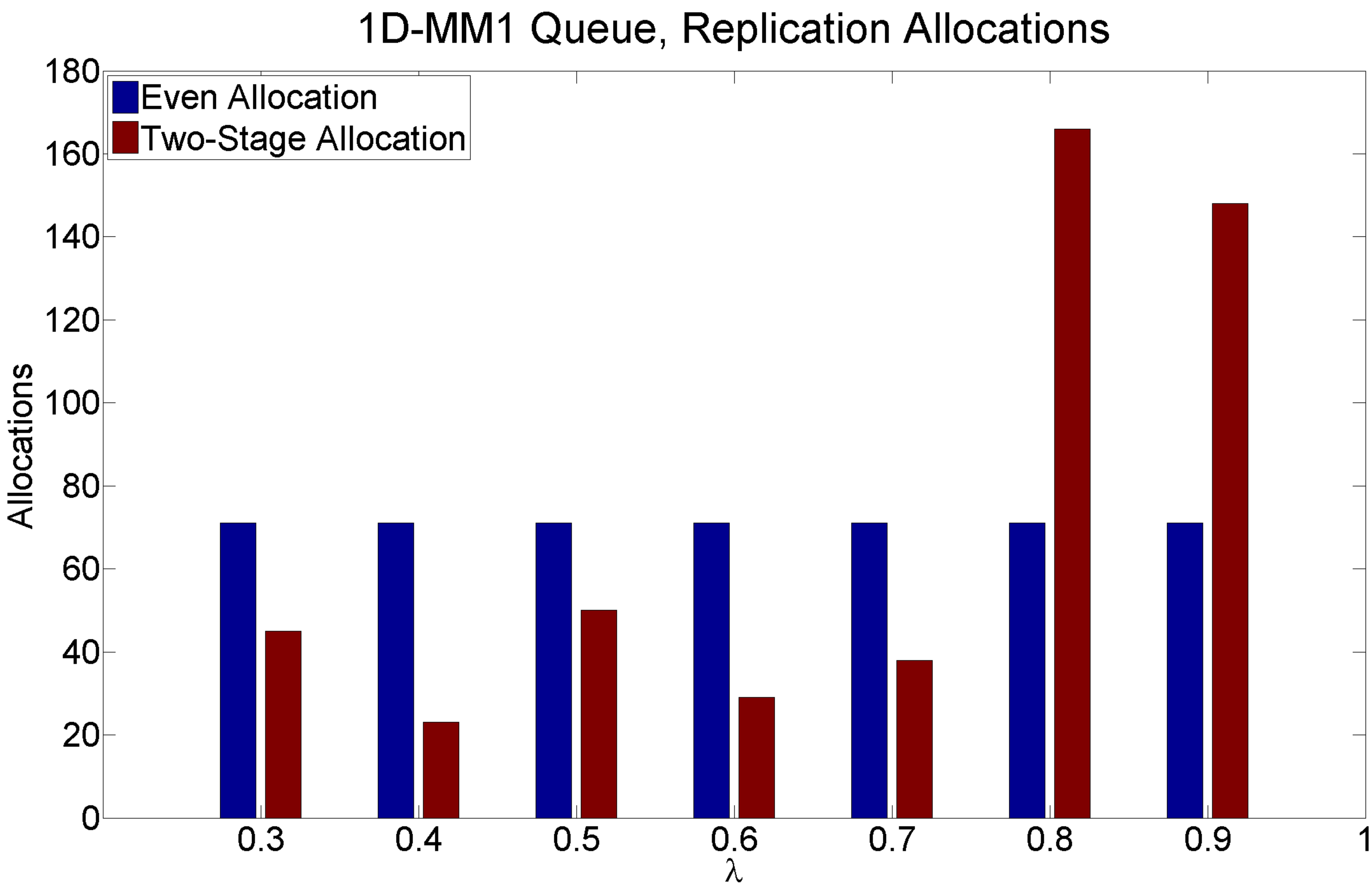
1D-MM1 Queue, Stage 2, 500 Replications
Actual
Predicted
±√MSE
Queue Length
λ
15
10
5
0
-5
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Problems

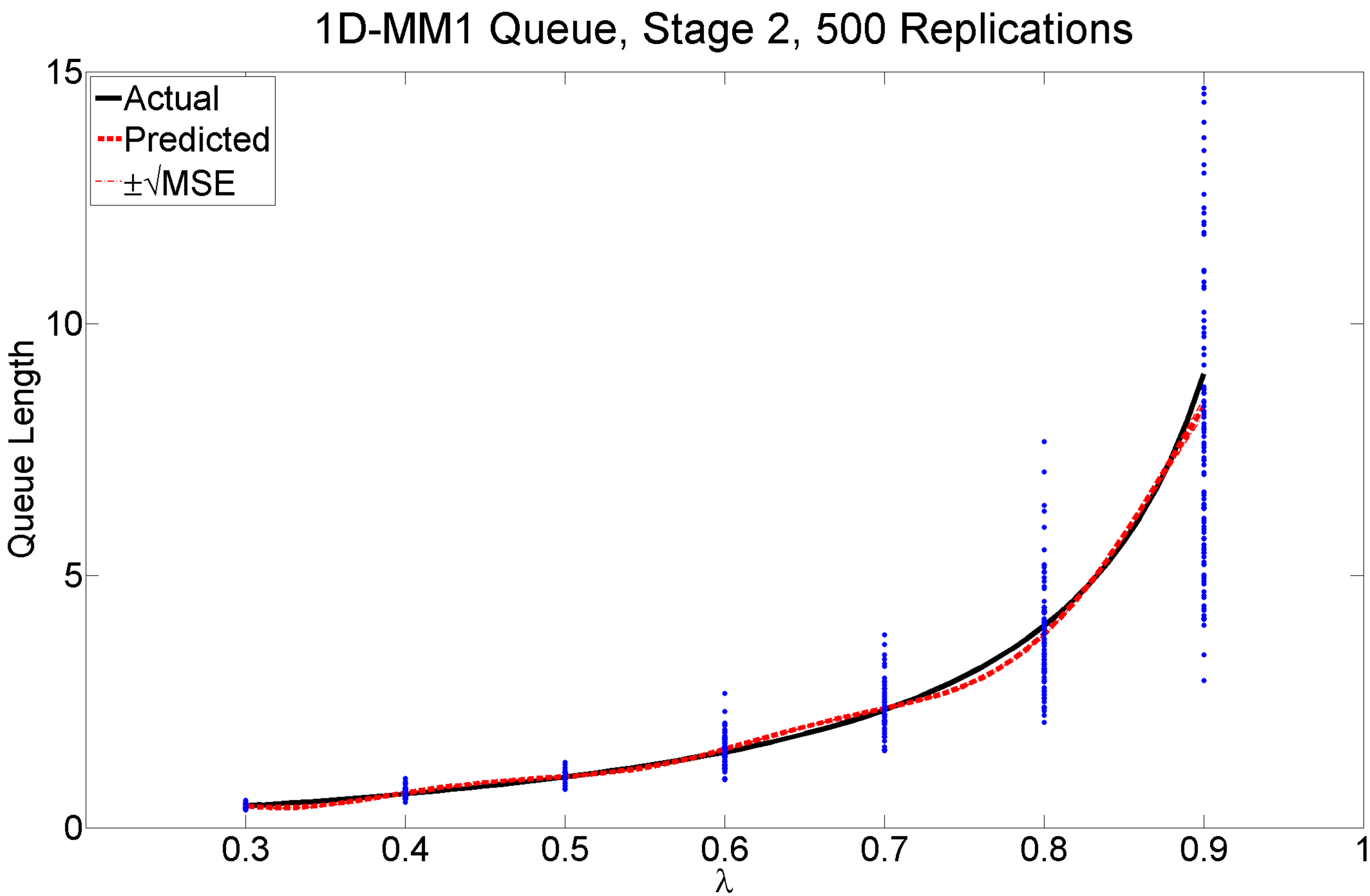


1D-MM1 Queue, Stage 1, 500 Replications
Actual
Predicted
±√MSE
Queue Length
λ
15
10
5
0
-5
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Problems

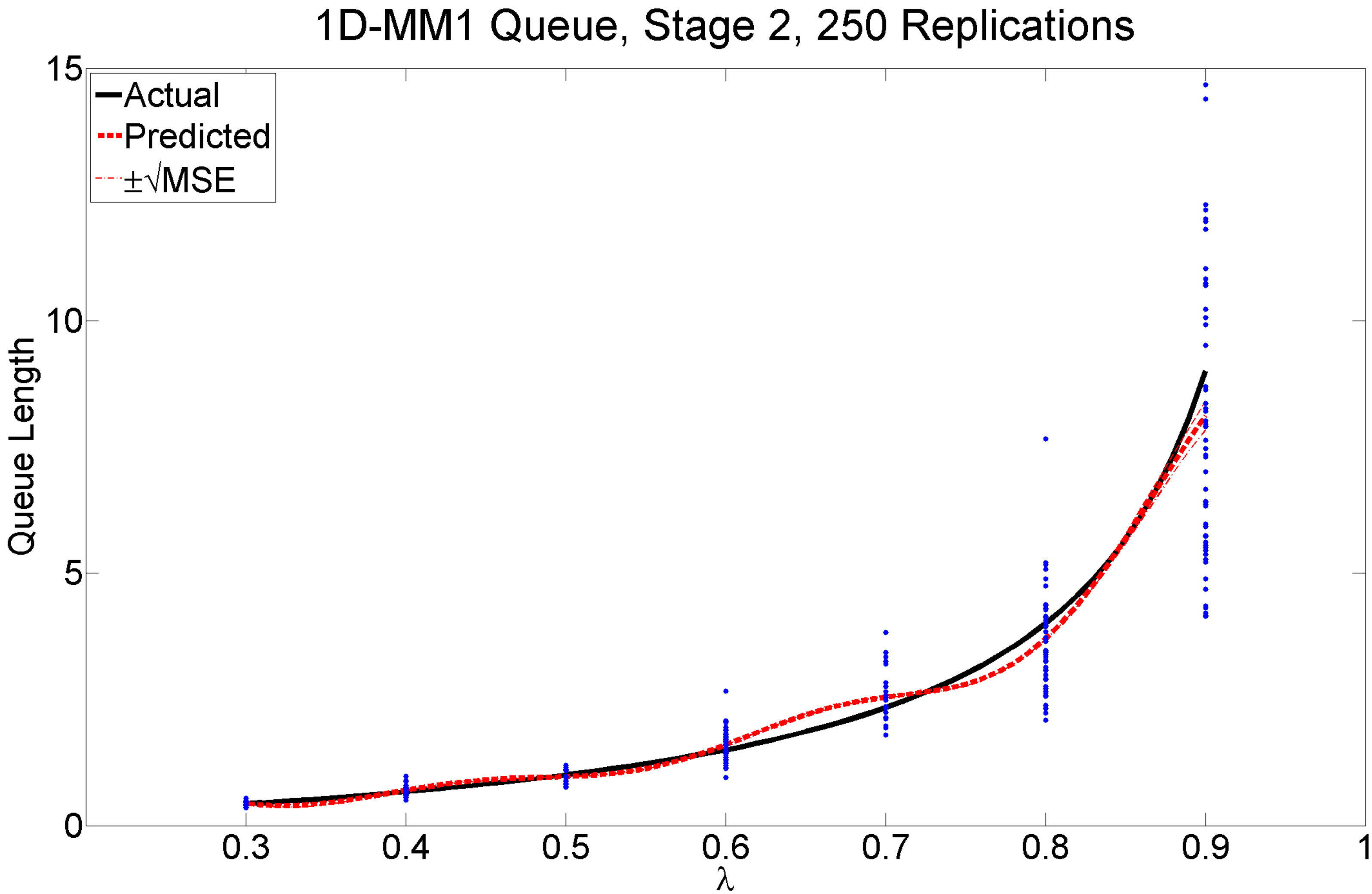

1D-MM1 Queue, Replication Allocations
Even Allocation
Two-Stage Allocation
Allocations
180
160
140
120
100
80
60
40
20
0
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
λ

Problems

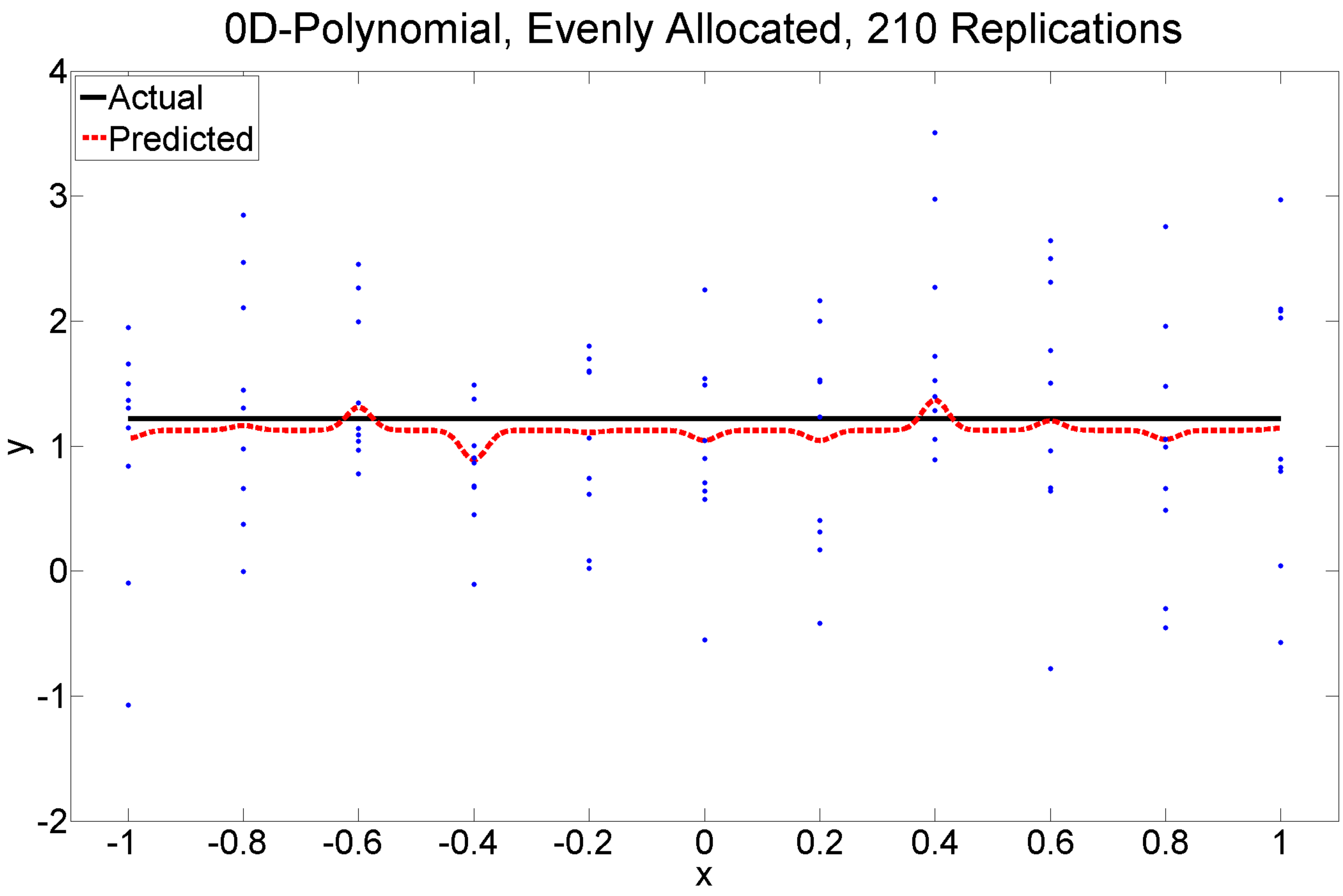


1D-MM1 Queue, Stage 2, 500 Replications
Actual
Predicted
±√MSE
Queue Length
λ
15
10
5
0
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Problems


1D-MM1 Queue, Stage 2, 250 Replications
Actual
Predicted
±√MSE
Queue Length
λ
15
10
5
0
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Problems



0D-Polynomial, Evenly Allocated, 210 Replications
Actual
Predicted
4
3
2
1
0
-1
-2
y
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
x

Future Work

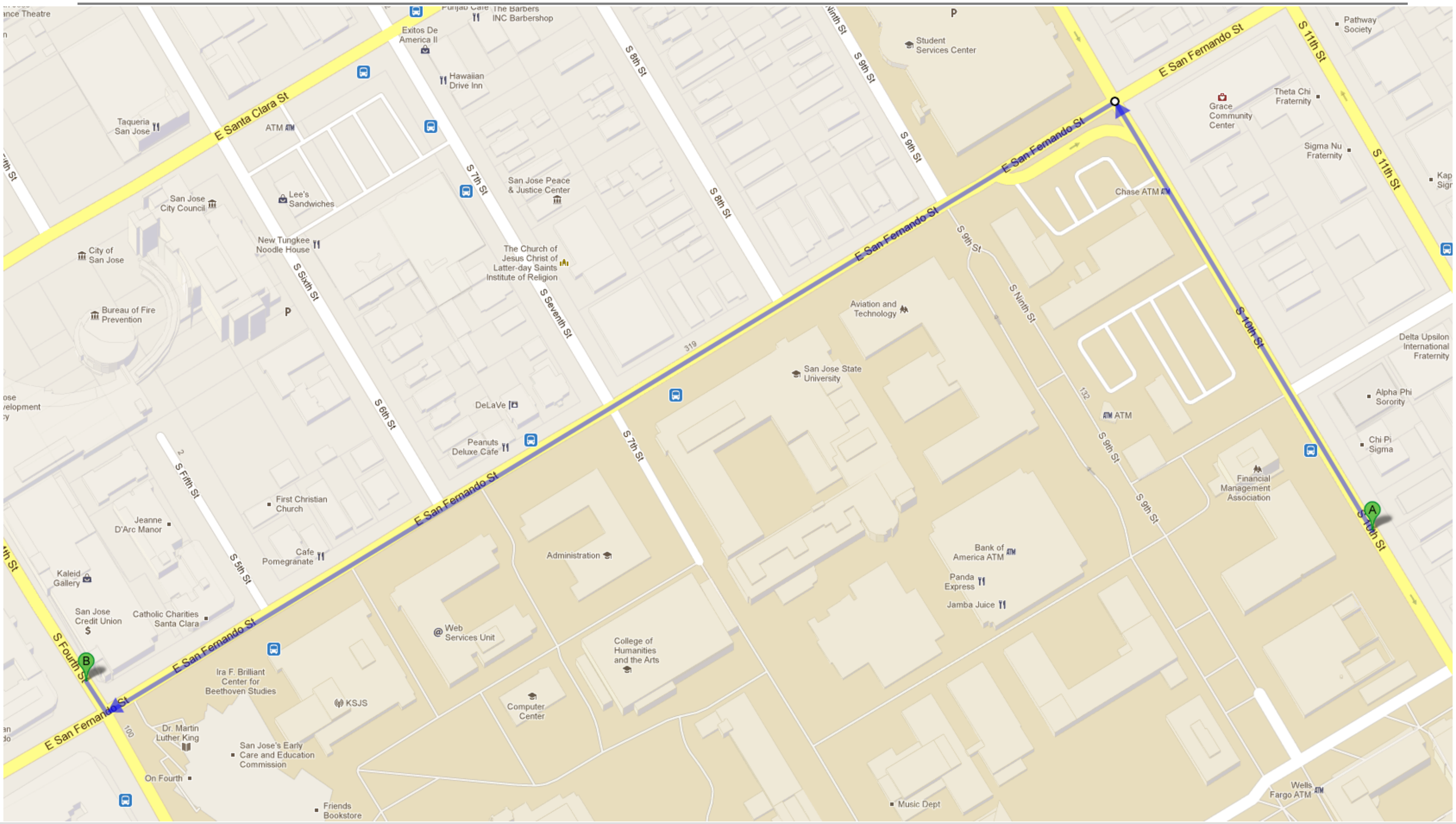
- Impose additional conditions to enforce smoothness
- Different experimental designs
- Different ways of implementing 2-stage
- How to pick design points
- Enforce positive estimated variance

Questions?

Thanks to

- IBM liaisons
  - Dr. Cheryl Kieliszewski
  - Dr. Peter Haas
  - Dr. Ignacio Terrizzano
- Professors
  - Dr. Bee Leng Lee
  - Dr. Bem Cayco
  - Drs. Martina Bremer, Steven Crunk, and Andrea Gottlieb
- CAMCOS
  - Dr. Slobodan Simić
- Our friends and family

Lunch! Flames Eatery & Bar, 88 S 4th Street



Courtesy of Google Maps.