SAND2012-2576C

Evaluation of Mixed Continuous-Discrete Surrogate Approaches

Patricia D. Hough and Laura P. Swiler
Sandia National Laboratories

Peter Qian and Xu Xu
University of Wisconsin

Herbert Lee
University of California, Santa Cruz

Curtis Storlie
Los Alamos National Laboratory

SIAM Conference on Uncertainty Quantification
Raleigh, North Carolina, April 2-5, 2012

National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Acknowledgements

- Genetha Gray, Sandia
- Bill Hart, Sandia
- Ken Perano, Sandia
- Cindy Phillips, Sandia
- Ali Pinar, Sandia
- John Siirola, Sandia
- Jean-Paul Watson, Sandia
- David Woodruff, UC Davis

Mission needs motivate research in simulation-based analysis

http://safetycampus.files.wordpress.com/2008/12/forklift_accident_with_bomb.jpg

System Engineers

- Probability of Loss of Assured Safety if dropped?
- Adjust handling height?

Analysts

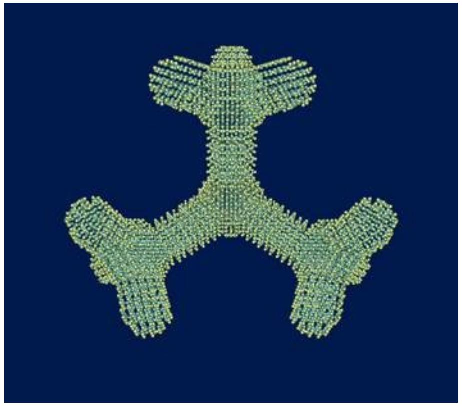
- Use simulation with optimization and UQ tools?
- Most info from limited number of simulations?

Algorithm Developers

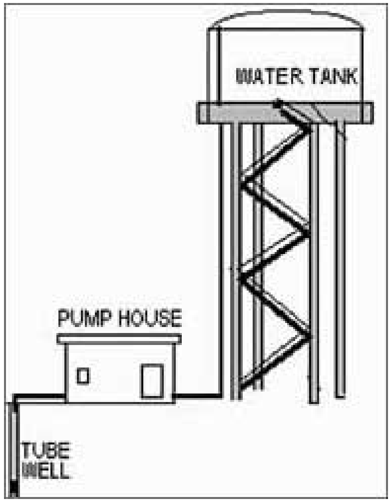
- New mathematics and statistics algorithms?
- Efficient implementations?

Categorical/discrete variables are finding their way into engineering and science simulations

System bolts
(Picture courtesy of J. Crowell)



Branched tetrapod nanocrystal generated by NREL "tetra" code
(Picture courtesy of P. Graf)



Water management system
(Picture courtesy of G. Gray)

- Modeling choices
  - Alternate plausible models
  - Choice between multiple materials
- Design choices
  - Material design
  - Operational settings

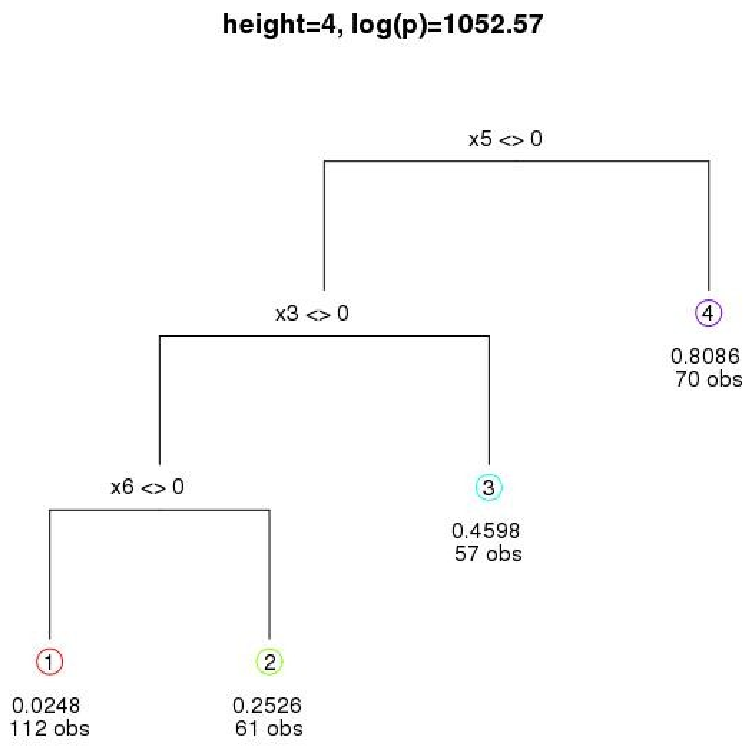
Key analysis characteristics
- Small number of variables
- Simulation based on computationally expensive equation solver

Would like to use surrogate models to improve tractability of simulation-based analysis

- Optimization and UQ methods are computationally expensive
- However, in mixed variable spaces...
  - Usual surrogate assumptions no longer hold
    - Continuous inputs
    - How output varies as input varies
  - Mixed variable surrogate approaches untested
  - Need a good testbed

Goal: Evaluate and compare mixed variable surrogate modeling approaches.

Approach 1: Treed Gaussian Process (TGP)



- Gaussian process is a specified by mean and covariance
- TGP partitions space and constructs GP in each partition
- Mixed variable variant allows partitioning over categorical/discrete variables
  - Transforms each variable-level pair into binary variable
  - GP constructed at "leaf" nodes over only continuous variables

R. B. Gramacy and M. Taddy. "Categorical inputs, sensitivity analysis, optimization and importance tempering with tgp version 2, an R package for treed Gaussian process models." R manual available at http://cran.r-project.org/, 2009.

Approach 2: Adaptive COmponent Selection and Smoothing Operator (ACOSSO)

- Univariate smoothing spline estimate

$$\frac{1}{n}\sum_{i=1}^{n}[y_i - f(x_i)]^2 + \lambda\int_0^1 [f''(x)]^2 dx$$

Term which penalizes roughness

- ACOSSO Estimate: f is an additive function

$$\boxed{f(x) = \sum_{j=1}^{q} f_j(x_j)}$$

$$f_j(1) = c_1, f_j(2) = c_2, \ldots f_j(m_j) = c_{m_j}, \sum_{x=1}^{m_j} c_j(x) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}[y_i - f(\mathbf{x}_i)]^2 + \sum_{j=1}^{q}\lambda_j\int_0^1 [f_j''(x_j)]^2 dx$$

Term which penalizes trend

Term which penalizes categorical predictors

$$\boxed{\frac{1}{n}\sum_{i=1}^{n}[y_i - f(\mathbf{x}_i)]^2 + \lambda\left(w_j \sum_{j=1}^{q}\left\{\left[\int_0^1 [f_j'(x_j)]dx_j\right]^2 + \int_0^1 [f_j''(x_j)]^2 dx_j\right\}^{1/2} + \sum_{j=q+1}^{p} w_j\left\{\sum_{x_j=1}^{m_j} f_j^2(x_j)\right\}^{1/2}\right)}$$

C.B. Storlie, J.C. Helton,, B. J. Reich, and L.P. Swiler. "Analysis of Computationally Demanding Models with Qualitative and Quantitative Inputs." Draft.

Approach 3: Gaussian Processes with Special Correlation Functions

- Covariance functions represent covariance between discrete points
- We consider four approaches
  - Individual Kriging: fits data at each qualitative level with distinct GP over quantitative variables
  - Exchangeable Correlation

    $\tau_{r,s} = c \ (0<c<1)$ for $r \neq s$
  - Multiplicative Correlation

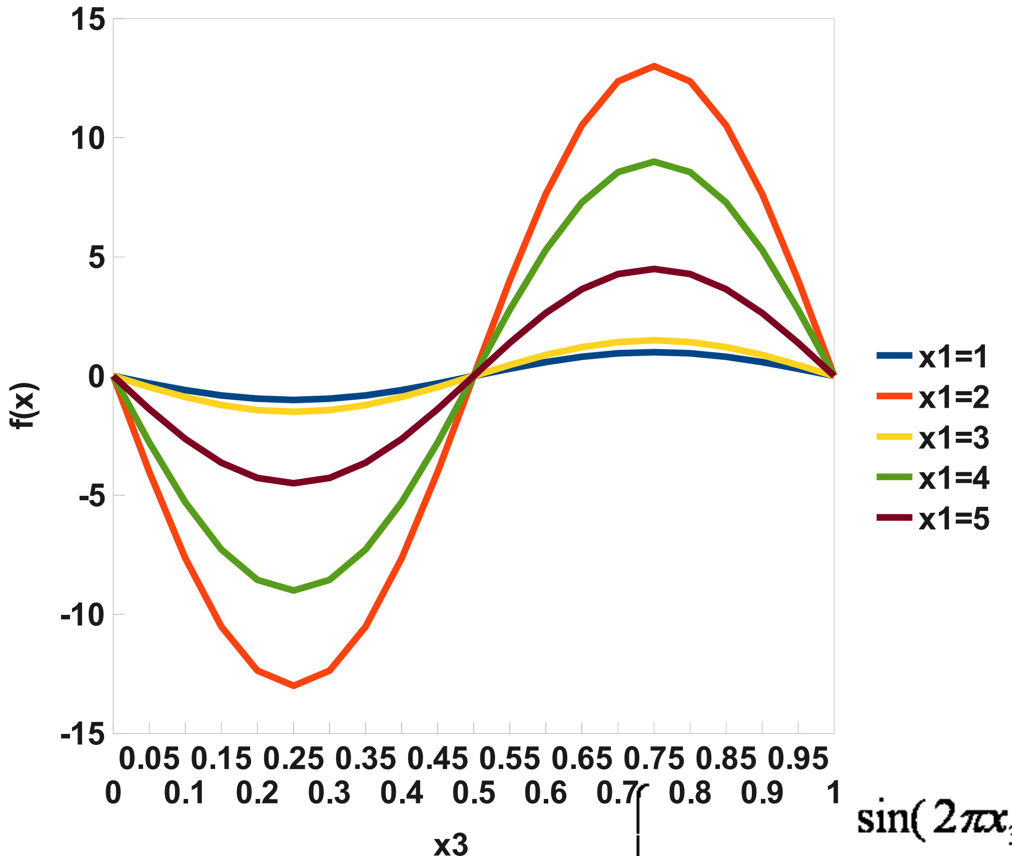
    $\tau_{r,s} = \exp\{-(\theta_r + \theta_s) I[r \neq s]\} \ (\theta_r, \theta_s > 0)$
  - Unrestricted Correlation

P. Z. G. Qian, H. Wu, and C. F. J. Wu, "Gaussian process models for computer experiments with qualitative and quantitative factors", *Technometrics,* 50(3):383-396, 2008.

We defined a set of tests and evaluation metrics

- Applied all approaches to various test functions with a range of characteristics
- Used mean squared error (MSE) over 200 Latin Hypercube points per qualitative level as a metric for comparison
- Performed 10 replicates of each experiment to to randomness in GPs
- Used 3 different sample designs for selecting build points
  - Standard Latin Hypercube: Latin Hypercube over continuous variables randomly assigned to discrete levels
  - k Latin Hypercube: Latin Hypercube generated over continuous variables at each discrete level
  - Sliced Latin Hypercube: Latin Hypercube over continuous variables is sliced into smaller Latin Hypercubes assigned to each discrete level

Test Function 1: Different sine curve over continuous variables for each categorical level

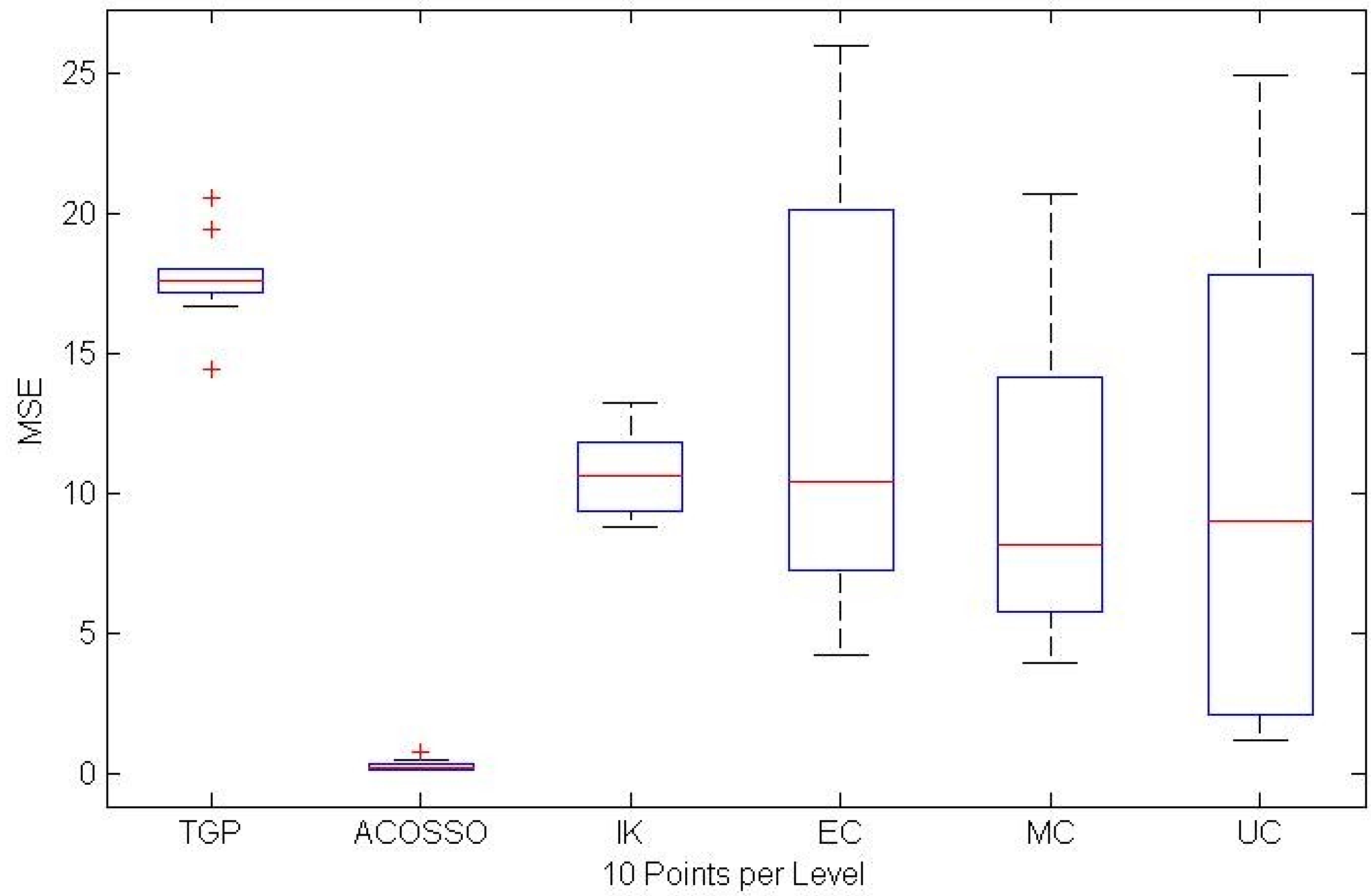


$x_1 \in \{1, 2, 3, 4, 5\}$

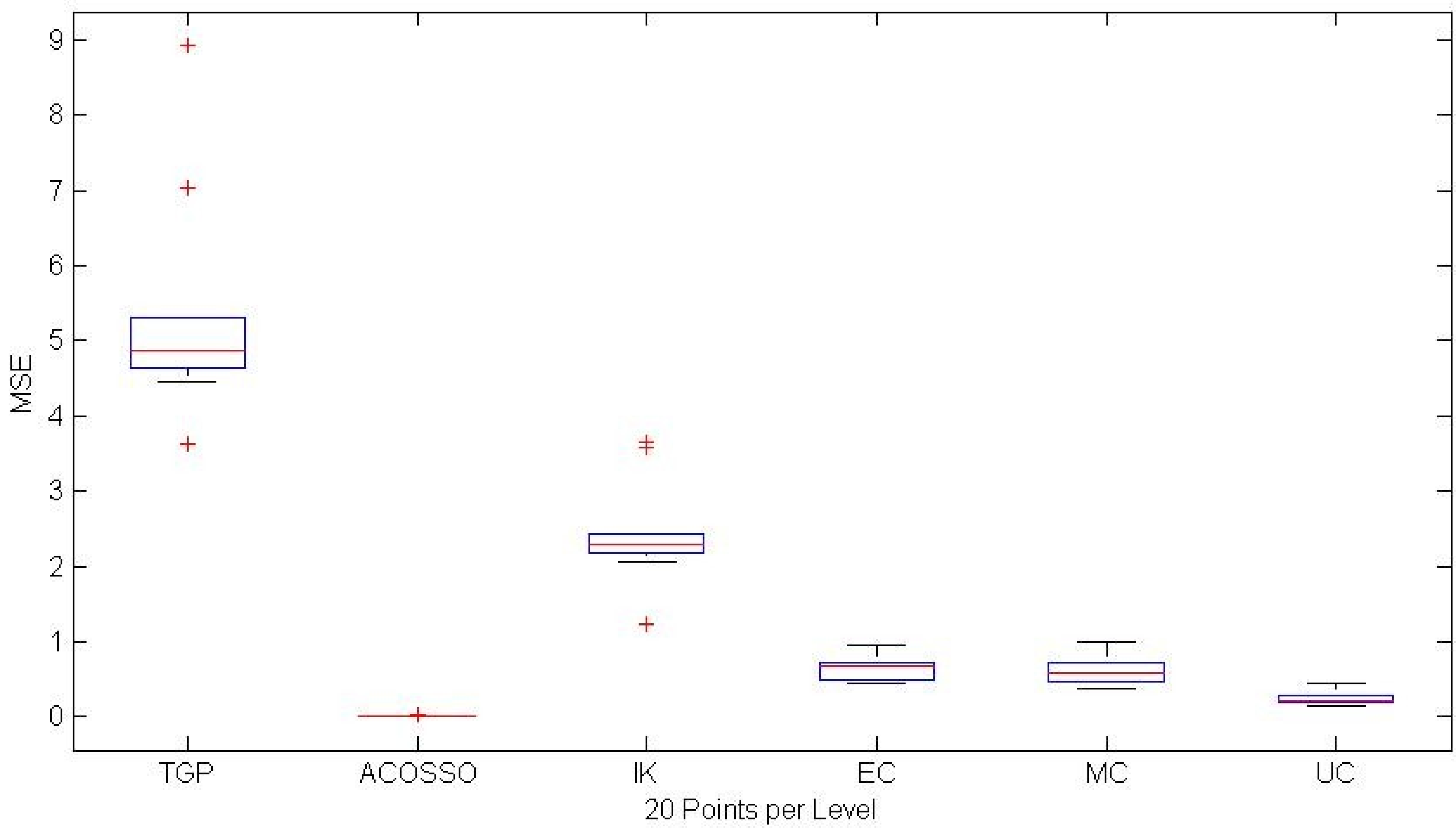
$x_2, x_3 \in (0, 1)$

$$y = f_2(x) = \begin{cases} \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) \text{ if } x_1 = 1 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 12\sin(2\pi x_3 - \pi) \text{ if } x_1 = 2 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 0.5\sin(2\pi x_3 - \pi) \text{ if } x_1 = 3 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 8\sin(2\pi x_3 - \pi) \text{ if } x_1 = 4 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 3.5\sin(2\pi x_3 - \pi) \text{ if } x_1 = 5 \end{cases}$$

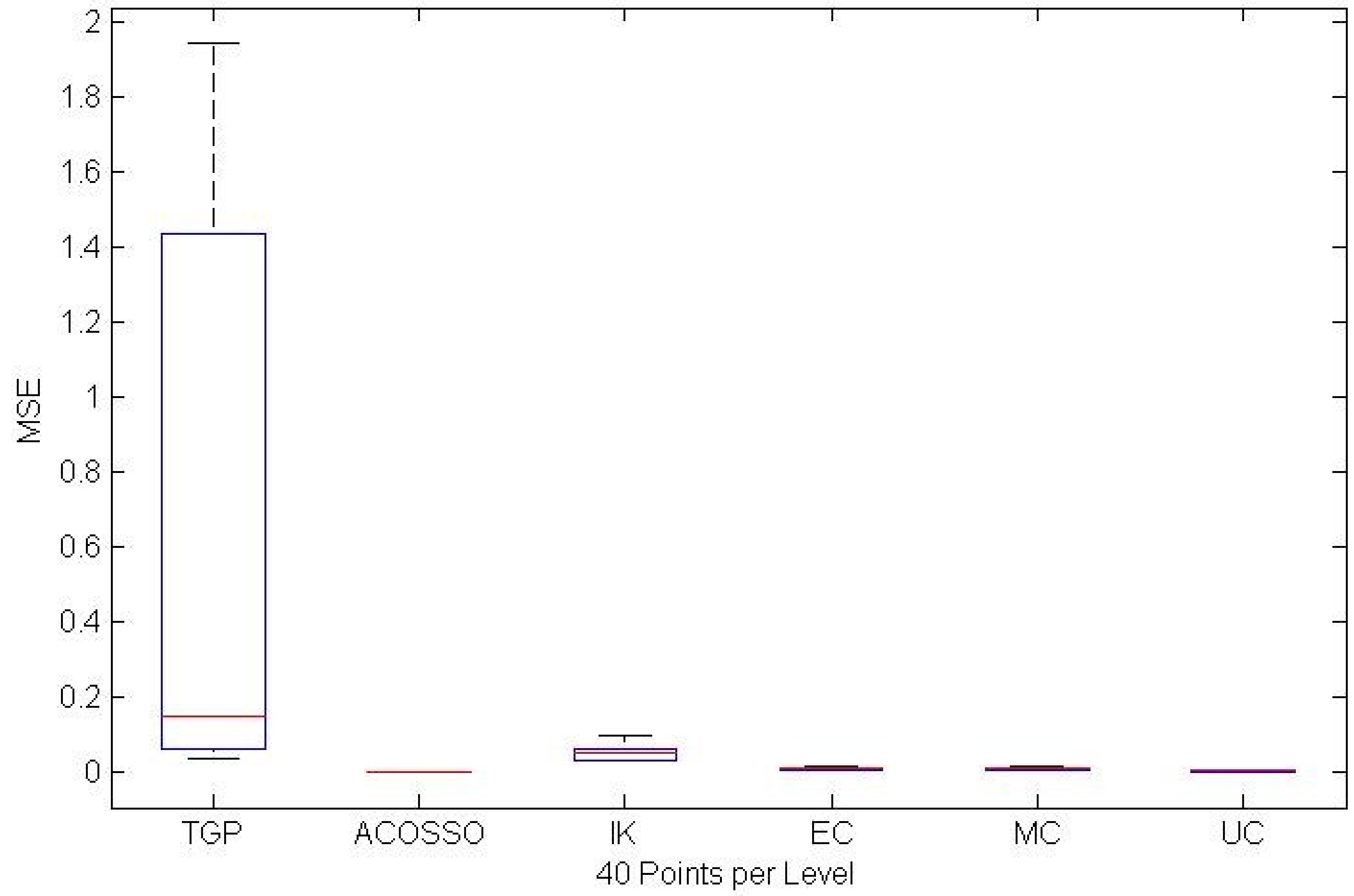
ACOSSO fits it well with only 10 build points per level due to separability



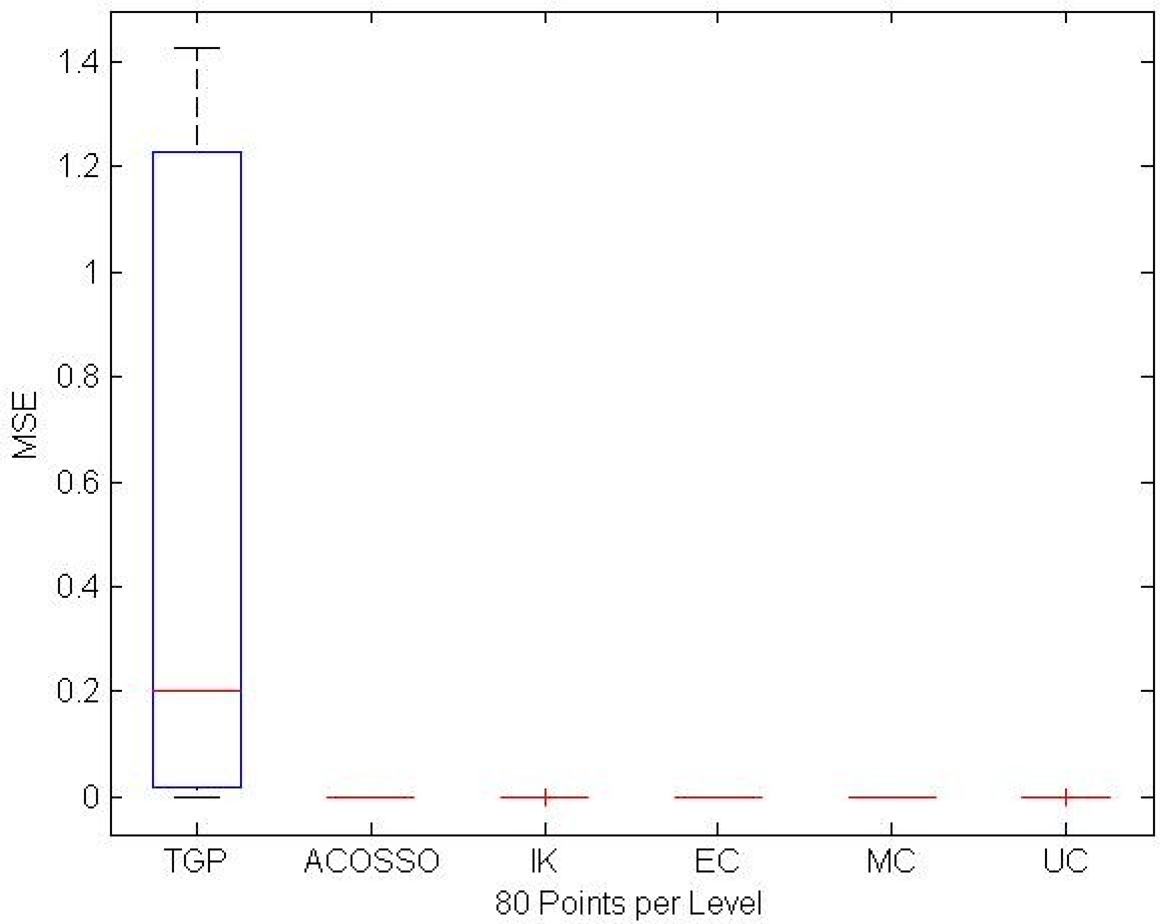
With 20 build points per level, GPs are catching up with ACOSSO



By 40 build points per level, all methods are comparable, though more variability with TGP

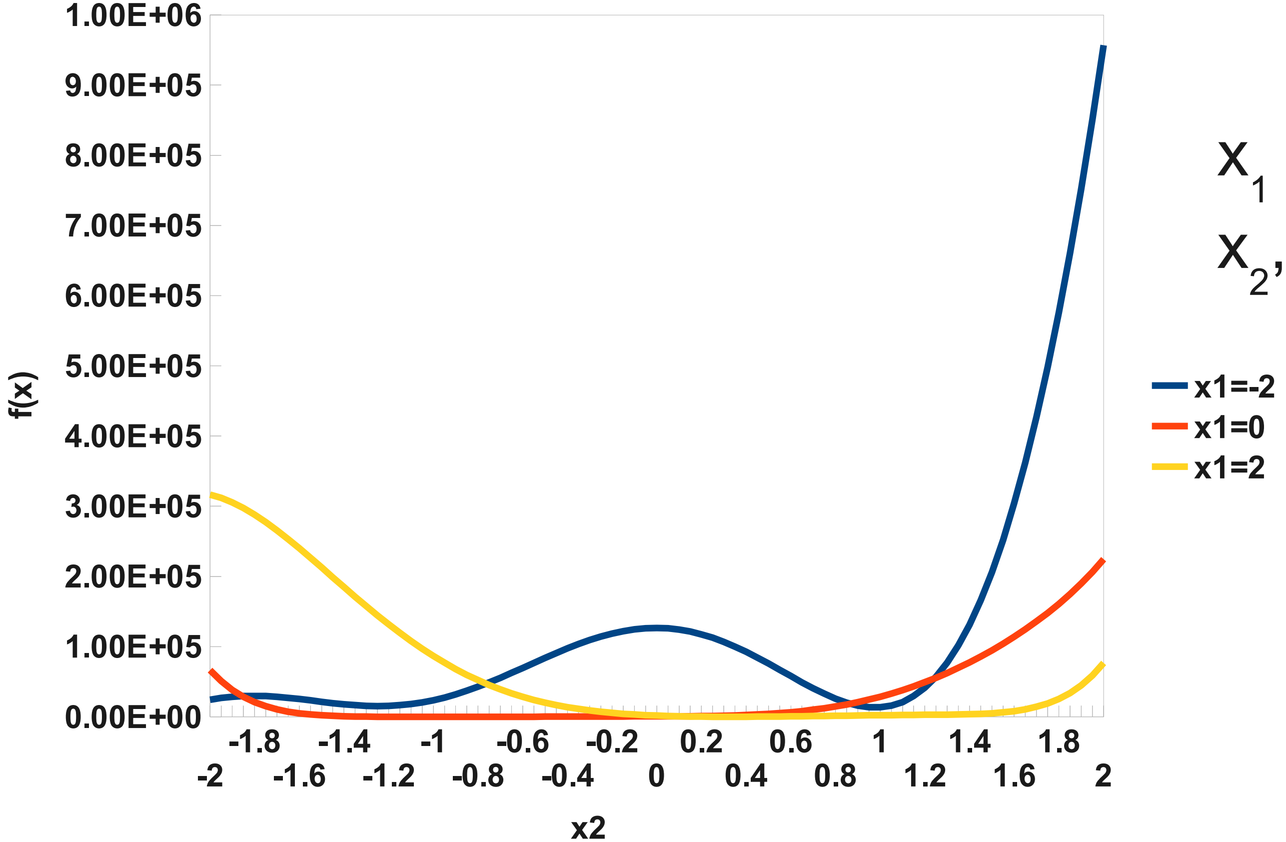


Very little gained by going to 80 build points per level



Test Function 2: Goldstein-Price Function

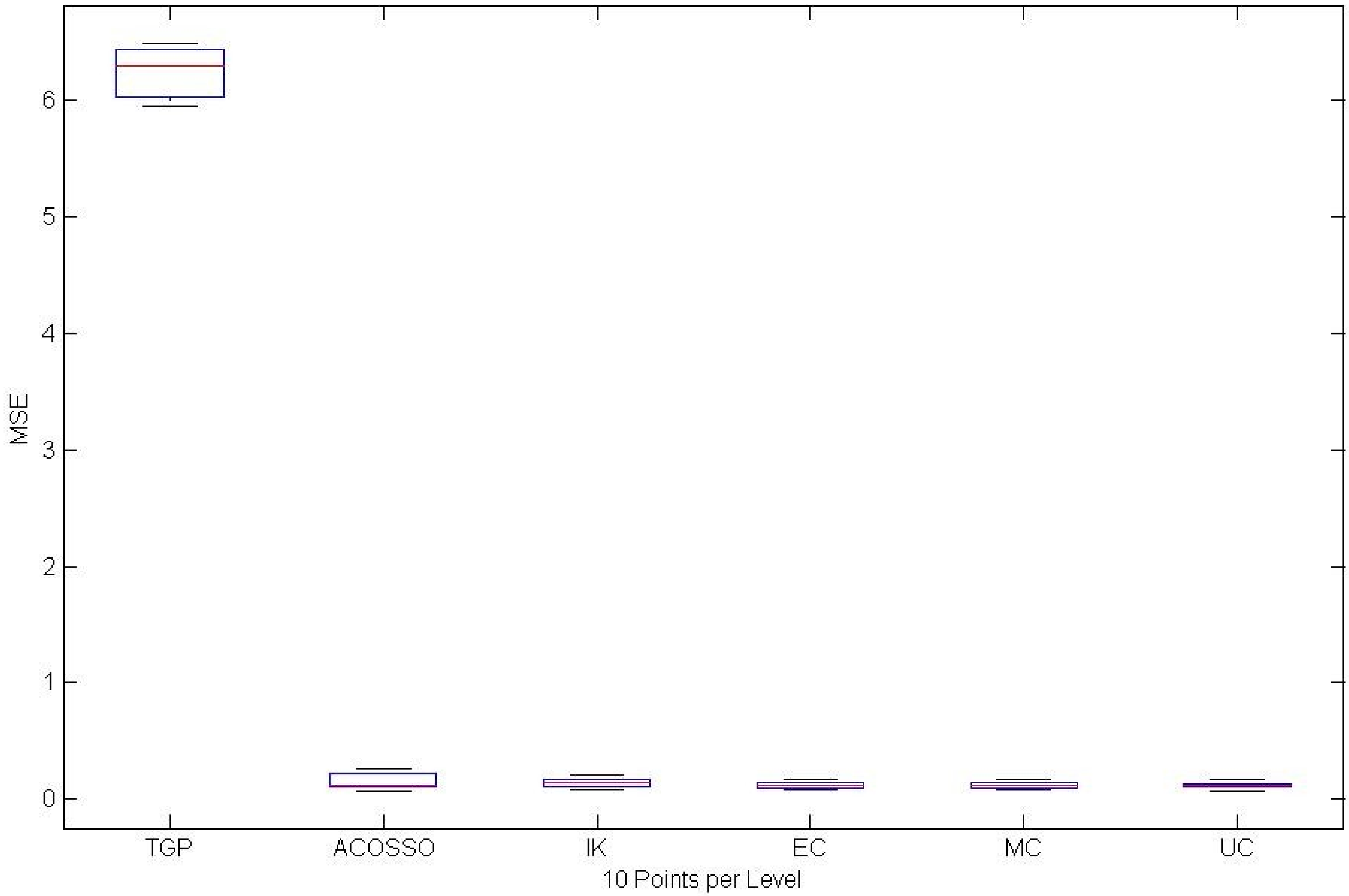
- Function of 2 variables ranging over six orders of magnitude
- $f(x) = (1 + (x_1 + x_2 + 1)^2 * (19 – 14x_1 + 3x_1^2 – 14x_2 + 6x_1x_2 + 3x_2^2) * (30 + (2x - 3x\ )^2 * (18 – 32x + 12x\ ^2 + 48x - 36x_1x_2 + 27x_2^2))$



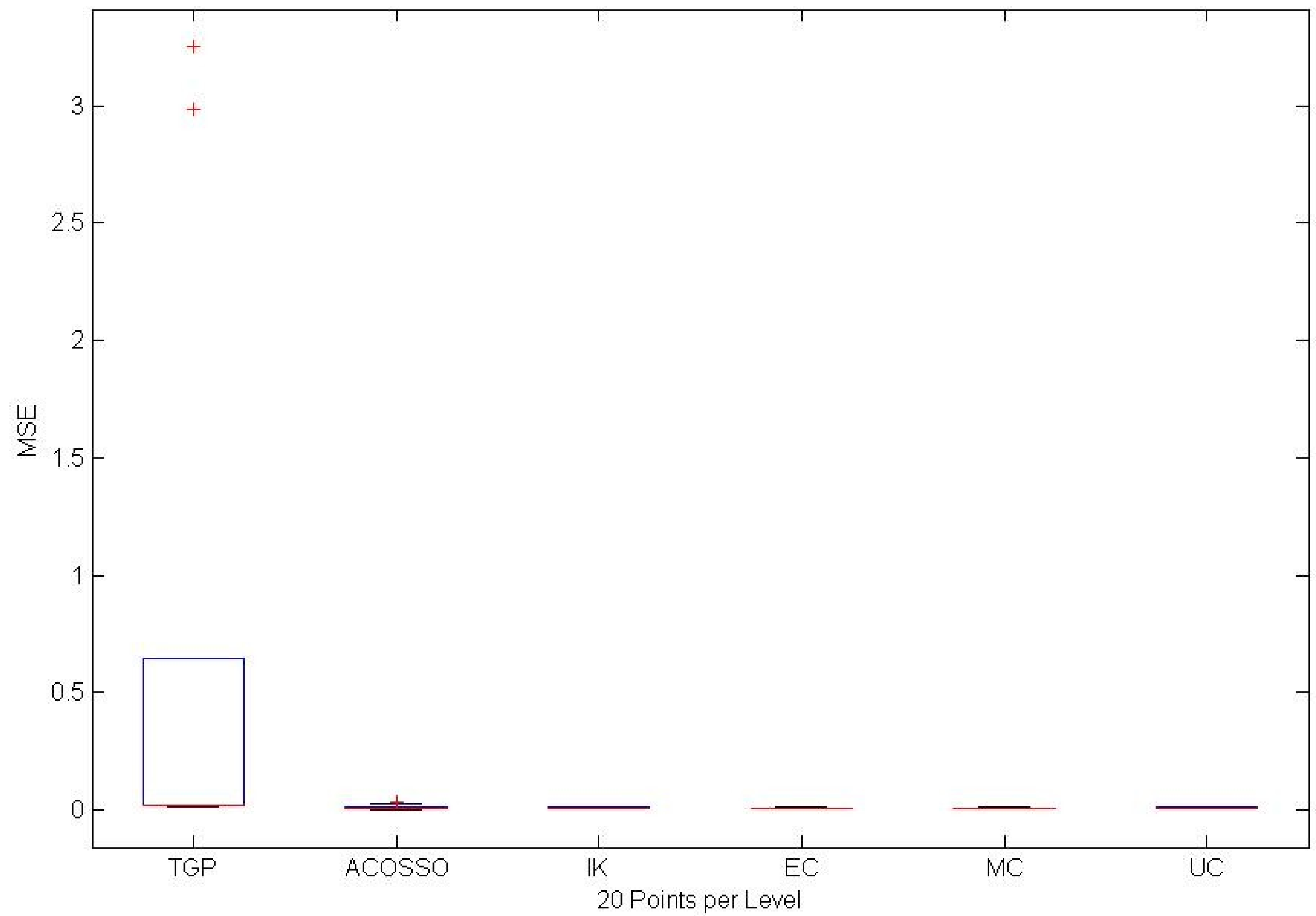
$x_1 \in \{-2, 0, 2\}$

$x_2, x_3 \in (-2, 2)$

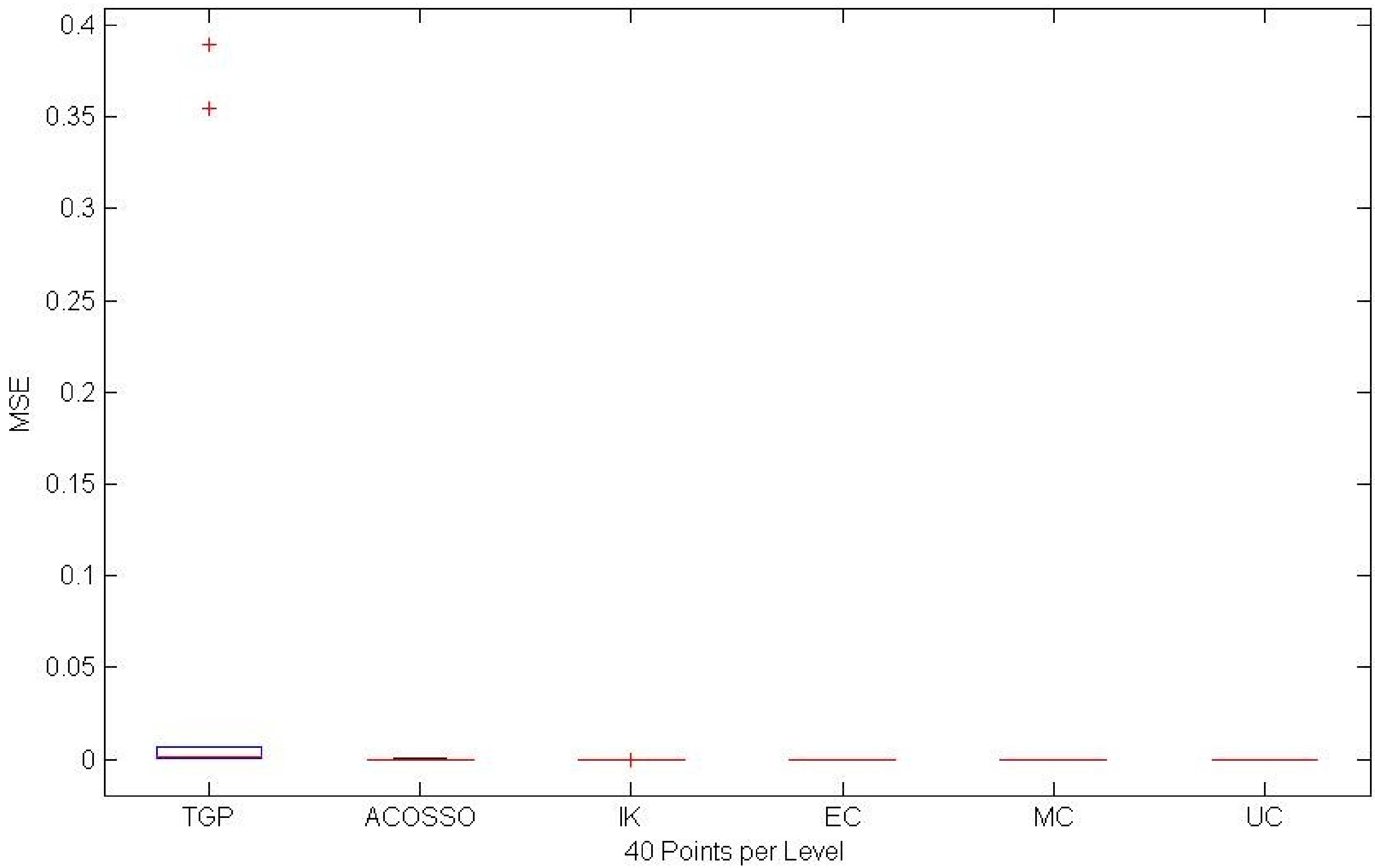
With 10 build points per level, all but TGP produce good fits



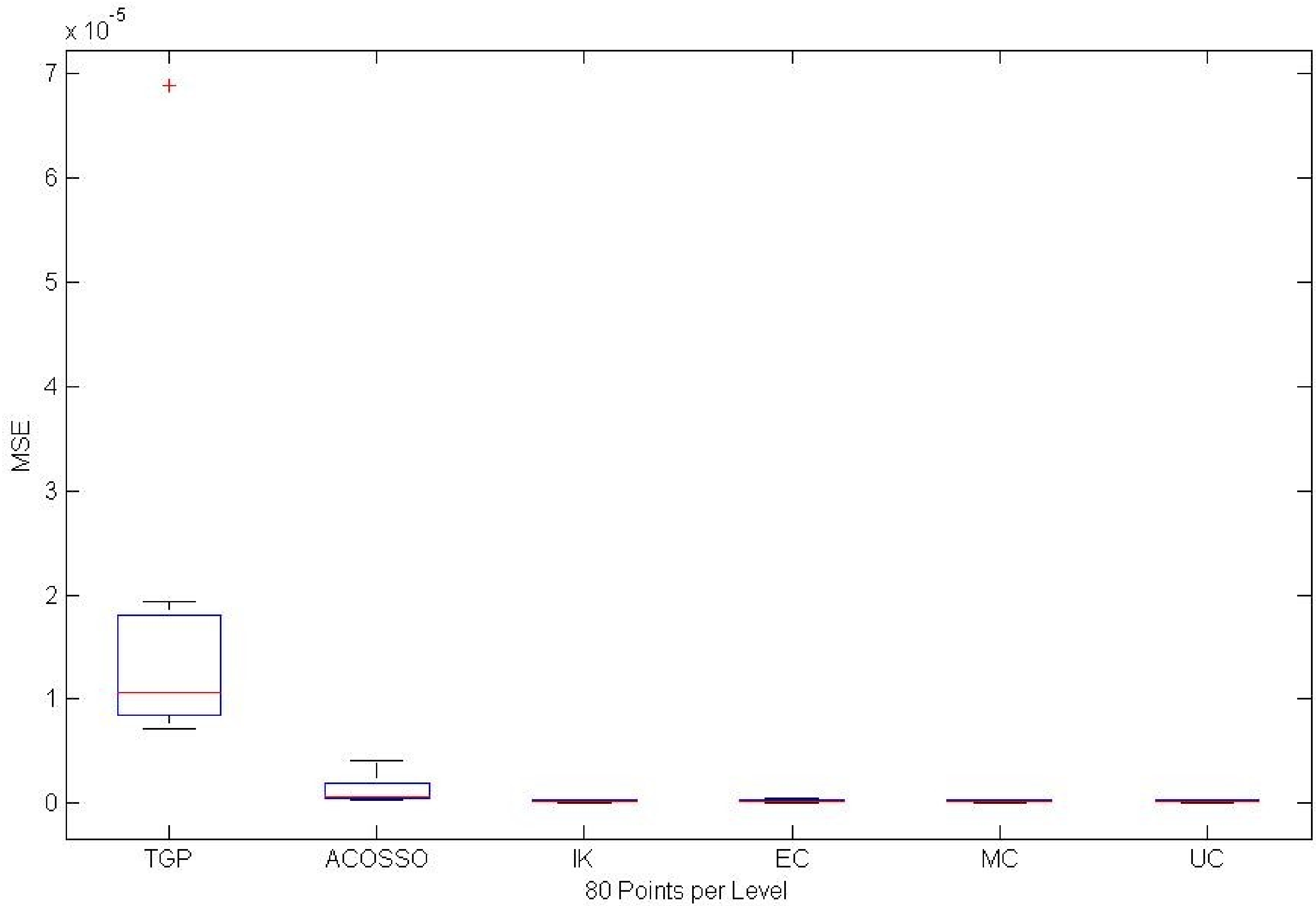
At 20 build points per level, TGP is on a par with the other methods



40 build points per level reduces the variance in TGP performance



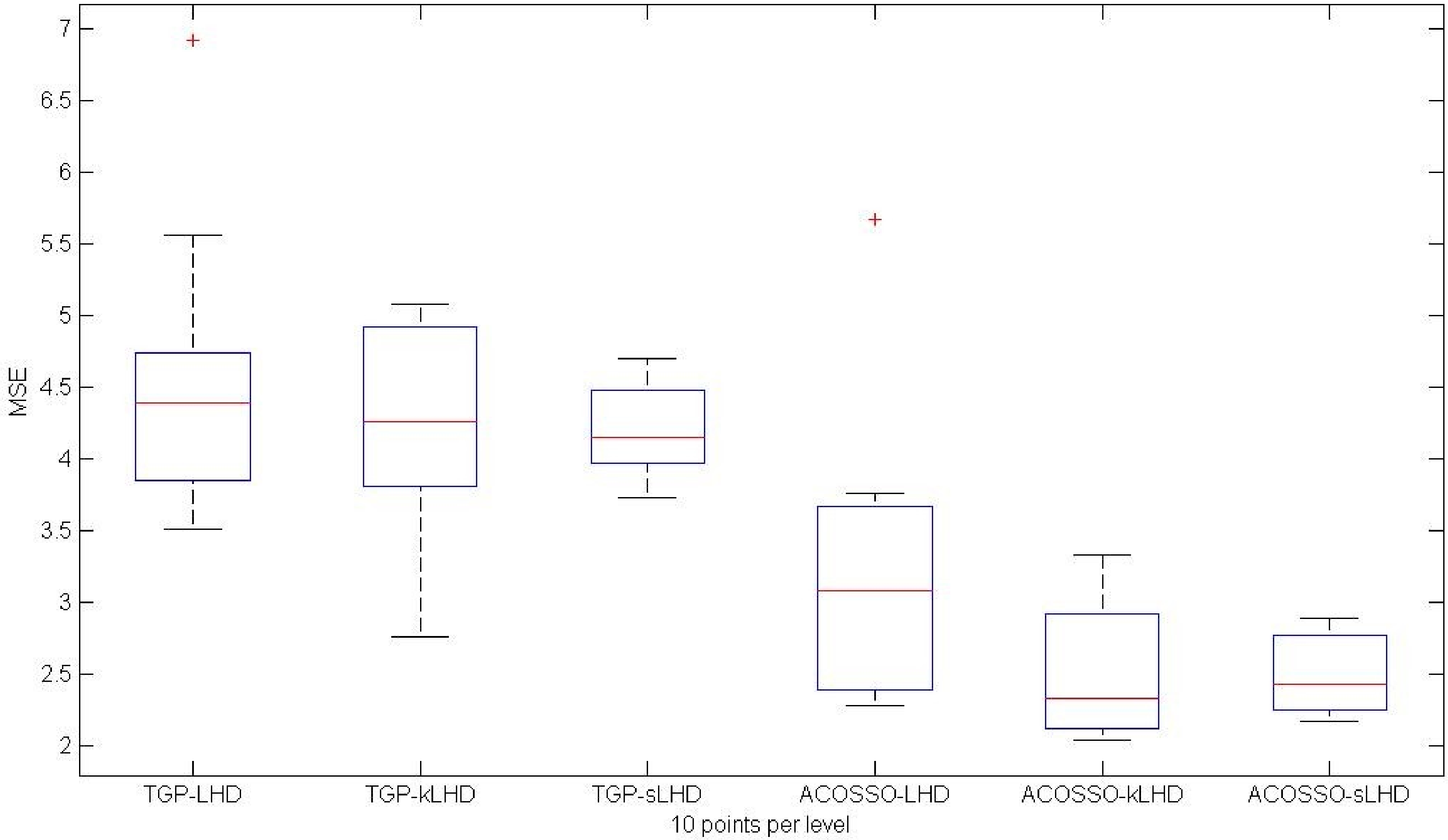
Little change results from going to 80 build points per level



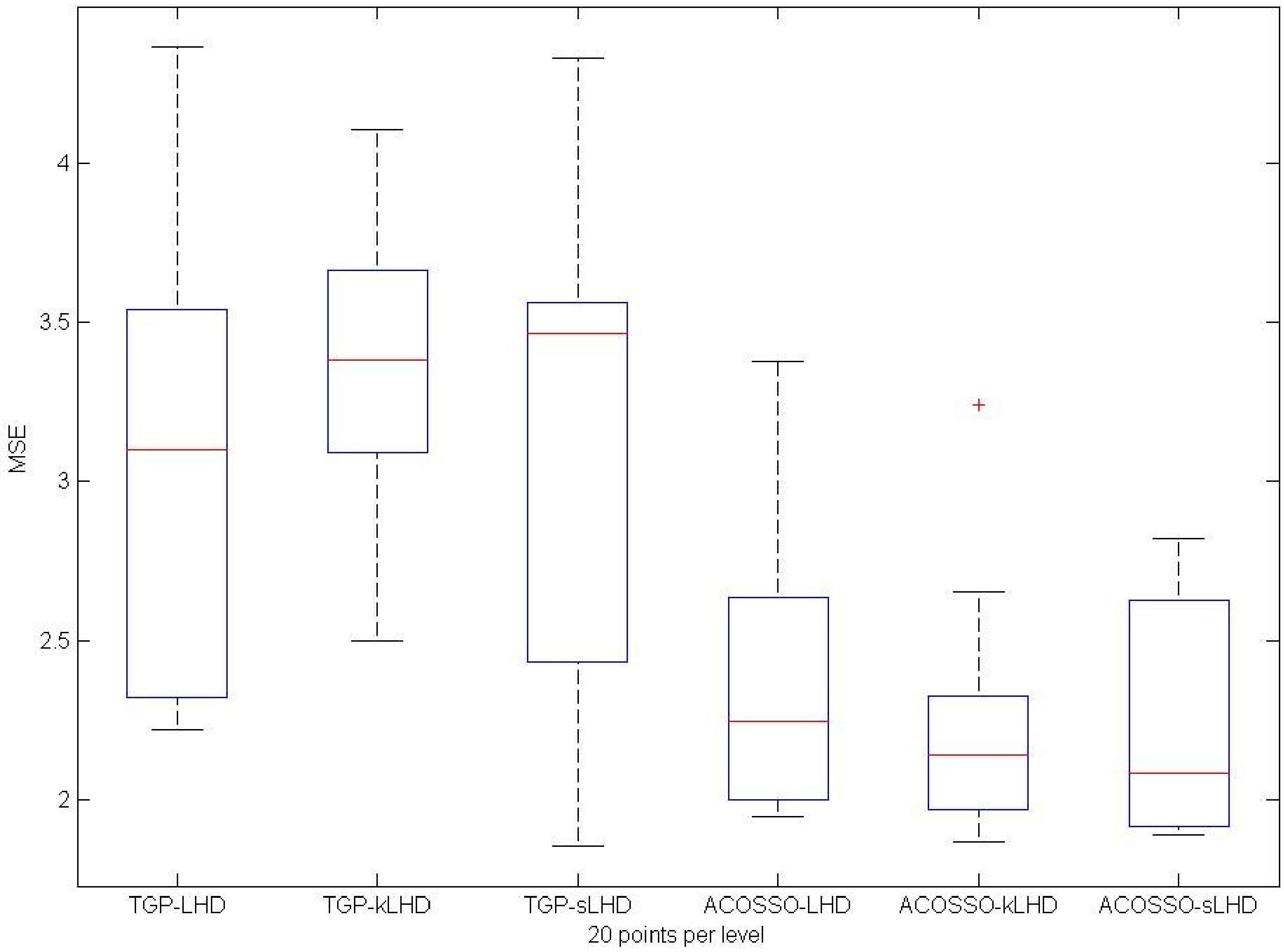
Test Function 3: 4th-order polynomial with 19 terms

- Generated using a random polynomial generator
  - 4 variables: 2 continuous, 2 discrete
    $x_1, x_2 \in \{20, 50, 80\}$; $x_3, x_4 \in (0, 100)$
- Generator uses a system of linear equations to solve for the random coefficients, described in:
  - McDaniel, W. R. and B. E. Ankenman, “A Response Surface Test Bed.” Qual. Reliab. Engng. Int. 2000; 16: 363–372
- Can control the degree of nonlinearity, range of polynomial values, etc.

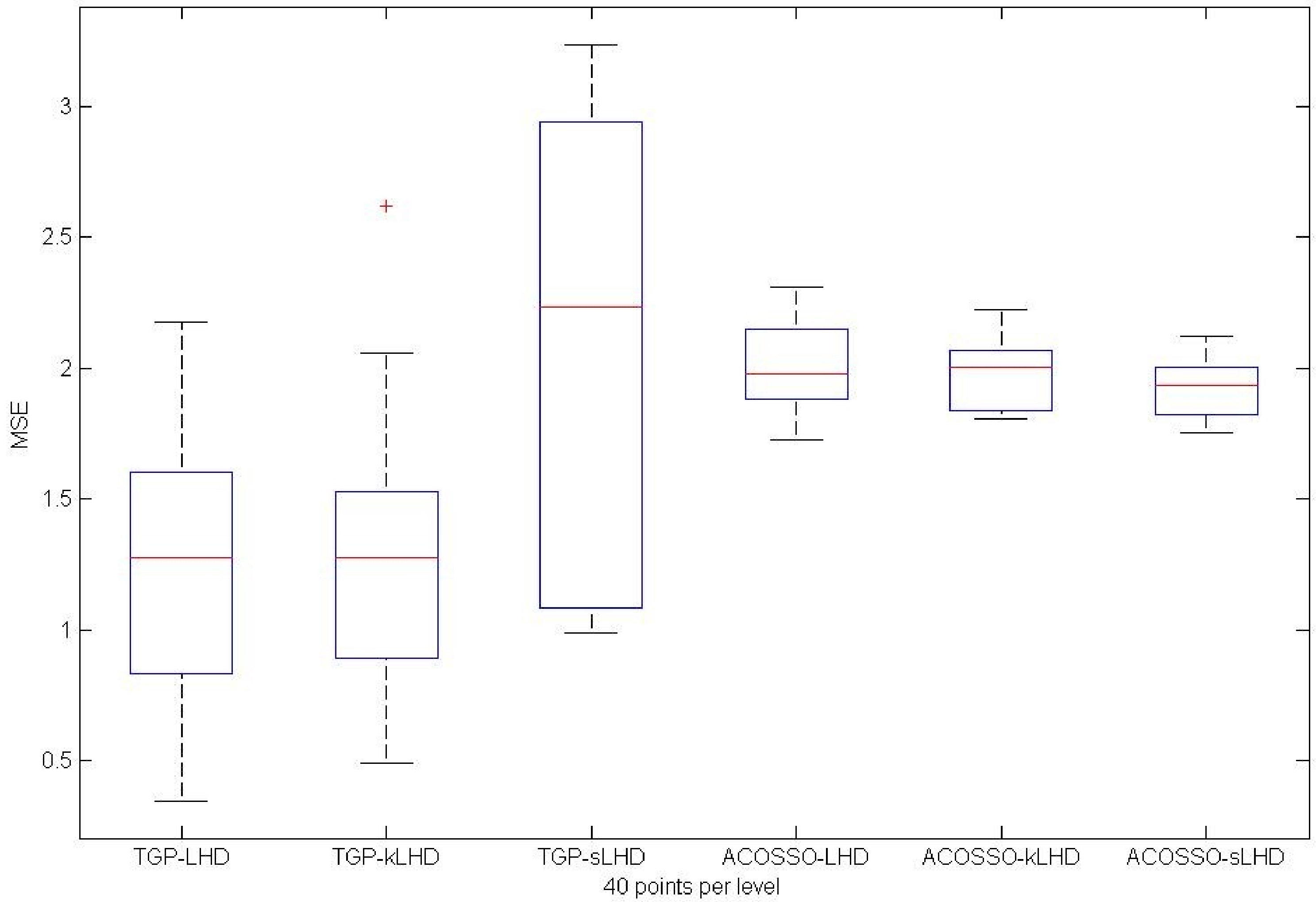
At 10 build points per level, ACOSSO starts out ahead



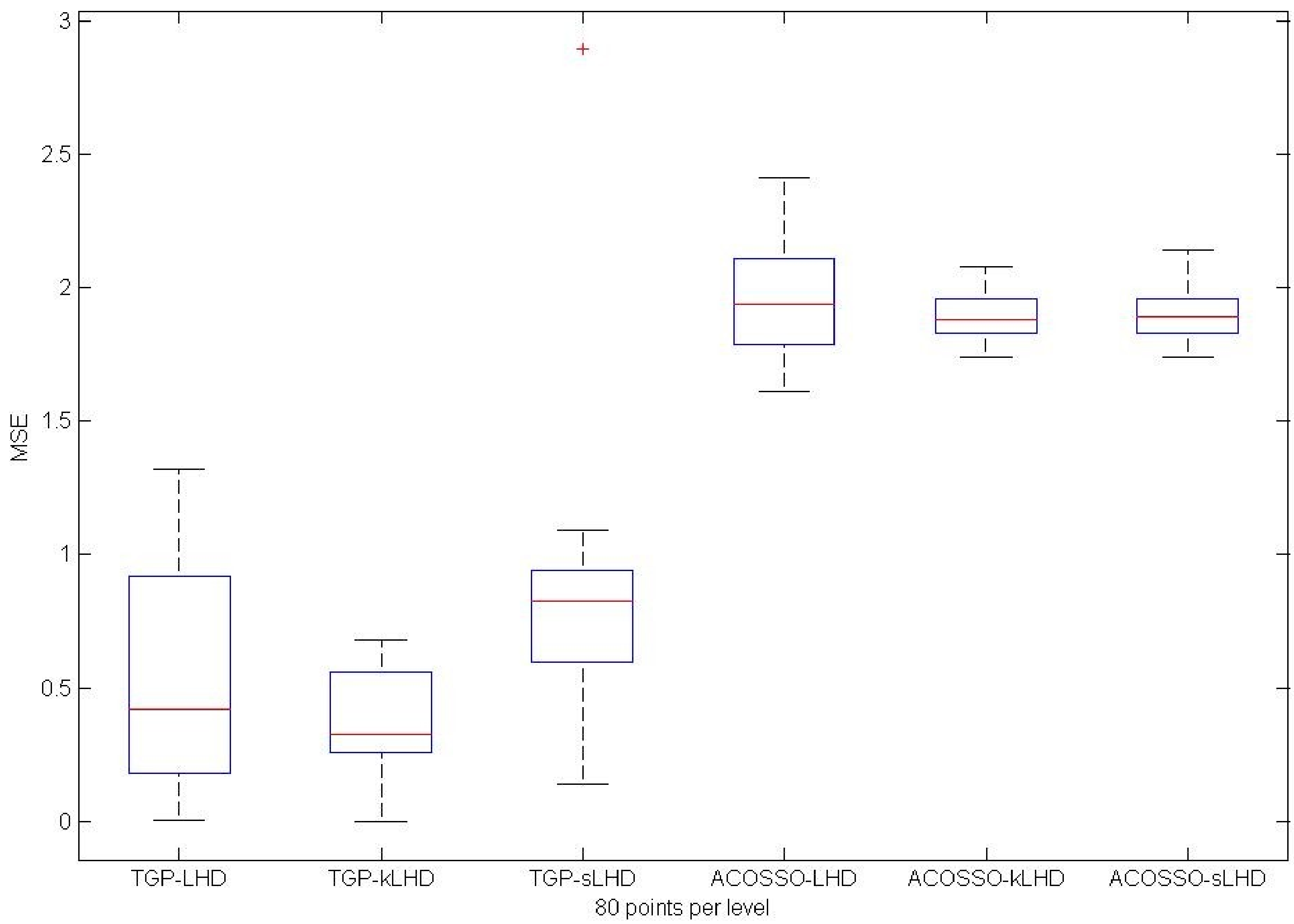
At 20 build points per level, GP is catching up



GP is pulling ahead at 40 build points per level



GP is doing well at 80 build points per level; ACOSSO has not improved at all



Closing Thoughts

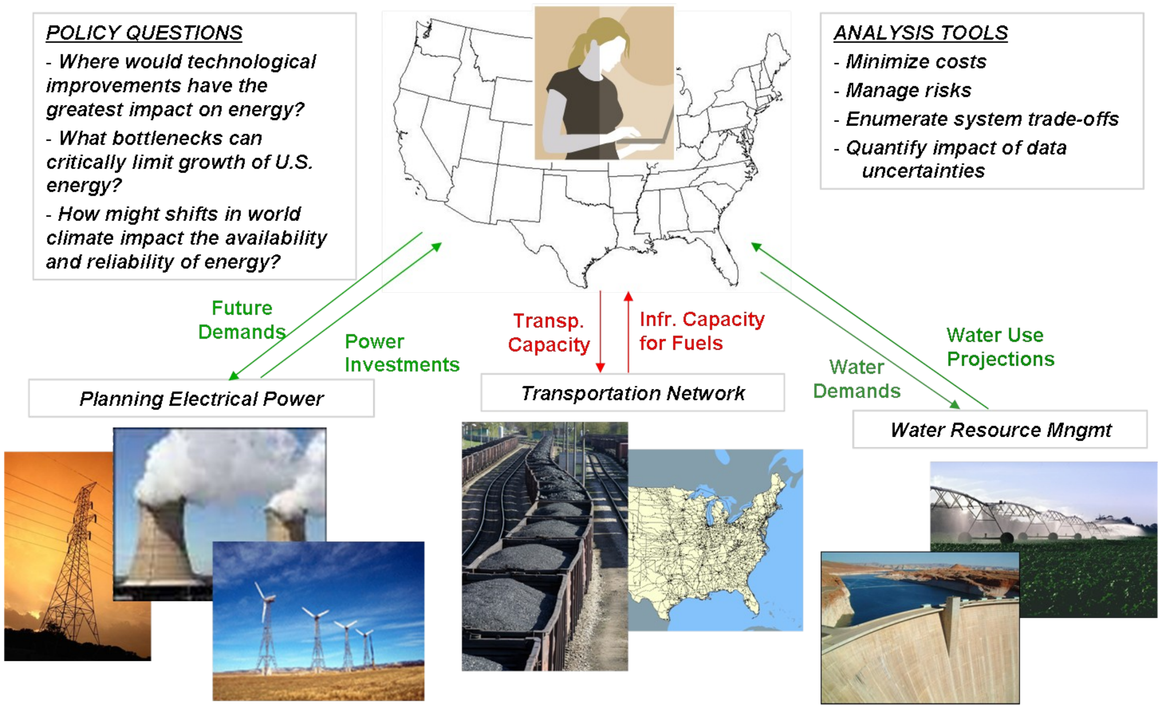
- Mixed-variable surrogates are essential to meet a growing need in simulation-based analysis
- We evaluated three candidate approaches
  - GP with special correlation functions appears most consistent
  - ACOSSO is best for separable functions
  - TGP is best for poorly scaled functions
  - All methods are viable
- Still work to be done
  - Evaluate on real problems
  - More efficient implementations

References

- R.B. Gramacy and H. K. H. Lee. "Bayesian treed Gaussian process models with an application to computer modeling." *Journal of the American Statistical Association*, 103:1119-1130, 2008.
- R. B. Gramacy and H. K. H. Lee. "Gaussian processes and limiting linear models." *Computational Statistics and Data Analysis,* 53:123-136, 2008.
- R. B. Gramacy and M. Taddy. "Categorical inputs, sensitivity analysis, optimization and importance tempering with tgp version 2, an R package for treed Gaussian process models." R manual available at http://cran.r-project.org/, 2009.
- McDaniel, W. R. and B. E. Ankenman, "A Response Surface Test Bed." *Qual. Reliab. Engng. Int.* 2000; 16: 363–372.
- P. Qian, H. Wu, and C.F.J. Wu. "Gaussian process models for computer experiments with qualitative and quantitative factors. " *Technometrics,* 50(3):383–396, 2008.
- B. J. Reich, C. B. Storlie, and H.D. Bondell. "Variable selection in Bayesian smoothing spline ANOVA models: Application to deterministic computer codes." Technometrics, 51, 110-120, 2009.
- C.B. Storlie and J.C. Helton. "Multiple predictor smoothing methods for sensitivity analysis: Description of techniques." *Reliability Engineering and System Safety,* 93(1):28–54, 2008.
- C.B. Storlie, L.P. Swiler, J.C. Helton, and C.J. Sallaberry. "Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models." *Reliability Engineering and System Safety,* 94 (2009) 1735–1763.
- C.B. Storlie, J.C. Helton,, B. J. Reich, and L.P. Swiler. "Analysis of Computationally Demanding Models with Qualitative and Quantitative Inputs." Draft manuscript.

Backup Slides

Additionally, large-scale "system of systems" simulations have many discrete variables



- Large-scale models composed of many constitutive system models
- Variety of variable types
  - Inventory
  - Yes/no siting
  - Control settings

Key analysis characteristics
- Large number of discrete and continuous variables
- Objective based on composition of heterogeneous models

Surrogate models improve computational tractability

(Queipo, Haftka, Shyy, Goel, Vaidyanathan, and Tucker (2005))

- Response surface models
  - Draw data from simulation
  - Fit fast approximation to data
- Reduced-order models
  - Reduce the number of variables
  - Principal component analysis, proper orthogonal decomposition, dimensionality reduction
- Multi-fidelity models
  - Coarsen the discretization
  - Reduce the amount of geometric detail
  - Reduce the amount of physics included
- Stochastic expansion
  - Build global approximation as function of uncertain variables
  - Polynomial chaos, stochastic collocation

Approach 1: Categorical Regression

- Uses indicator functions for categorical variable levels
  - $Y=\beta_0 + \beta_1X_1 + \beta_2X_2$ where X_1 is continuous, X_2 is binary
    - $Y=\beta_0 + \beta_1X_1$ for $X_2=0$, $Y=\beta_0 + \beta_1X_1 + \beta_2$ for $X_2=1$
    - Results in 2 different models
- Computationally expensive
  - Need enough samples over continuous variables for *EACH* discrete combination for accurate regression function
  - Increase number of discrete variables + increase number of “levels” per variable => combinatorial explosion

One of our favorite response surfaces is the Gaussian process

- Specified by mean and covariance
- Vanilla covariance function

$$C_{12}(\mathbf{x}^1,\mathbf{x}^2)=\sigma^2\exp\{-\sum_{i=1}^{n}\rho_i^2(\mathbf{x}_i^1-\mathbf{x}_i^2)^2\}$$

  - σ and ρ_i found by maximizing likelihood function

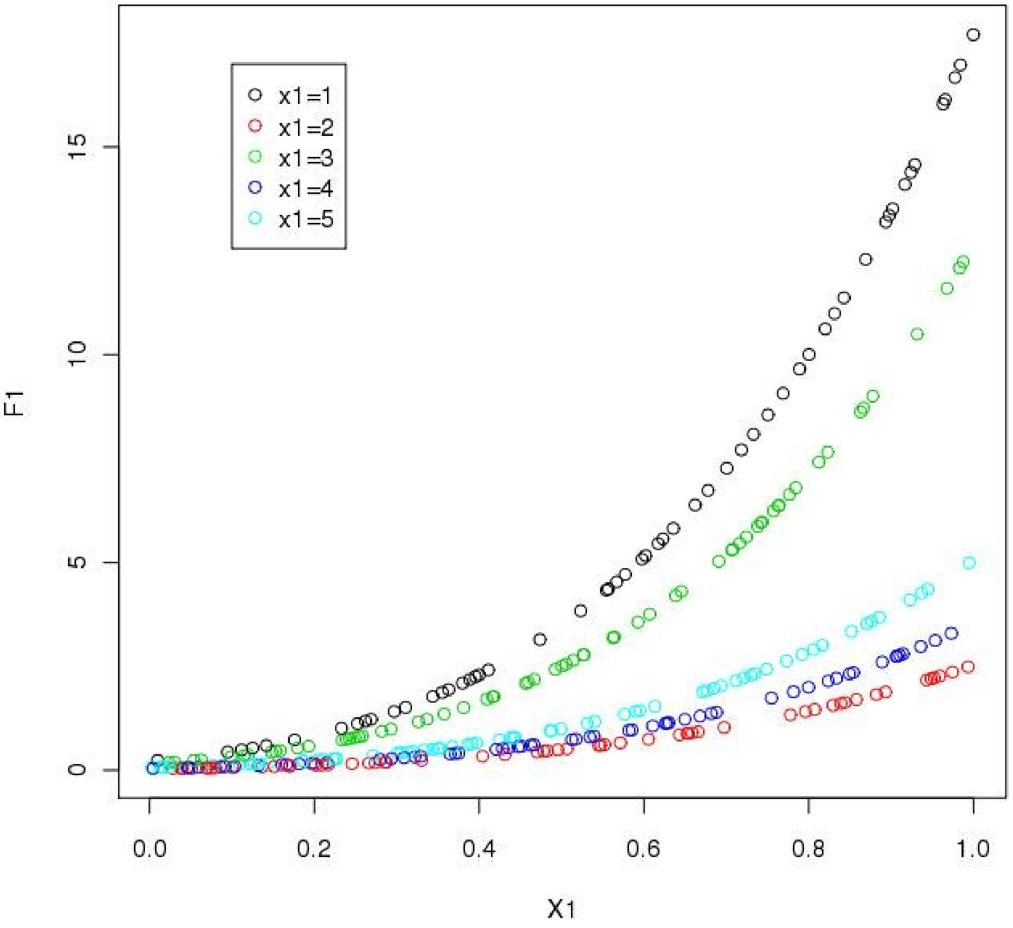
$$L=\frac{-n}{2}\log(2\pi)-\frac{1}{2}\log(\det(C))-\frac{1}{2}\mathbf{z}^T C^{-1}\mathbf{z}$$

Key feature: Estimates both mean behavior and variance.

Approach 2: Treed Gaussian Process (TGP)

- Gaussian process is a random process specified by mean and covariance functions
- Mixed variable variant allows partitioning over categorical/discrete variables
  - Transforms each variable-level pair into binary variable
  - GP constructed at “leaf” nodes over only continuous variables
- Alternate approach has explicit representation of categorical variables in GP
  - P. Qian, H. Wu, and C.F.J. Wu. “Gaussian process models for computer experiments with qualitative and quantitative factors.” *Technometrics,* 50(3):383–396, 2008.
  - Recommend isotropic correlation structure

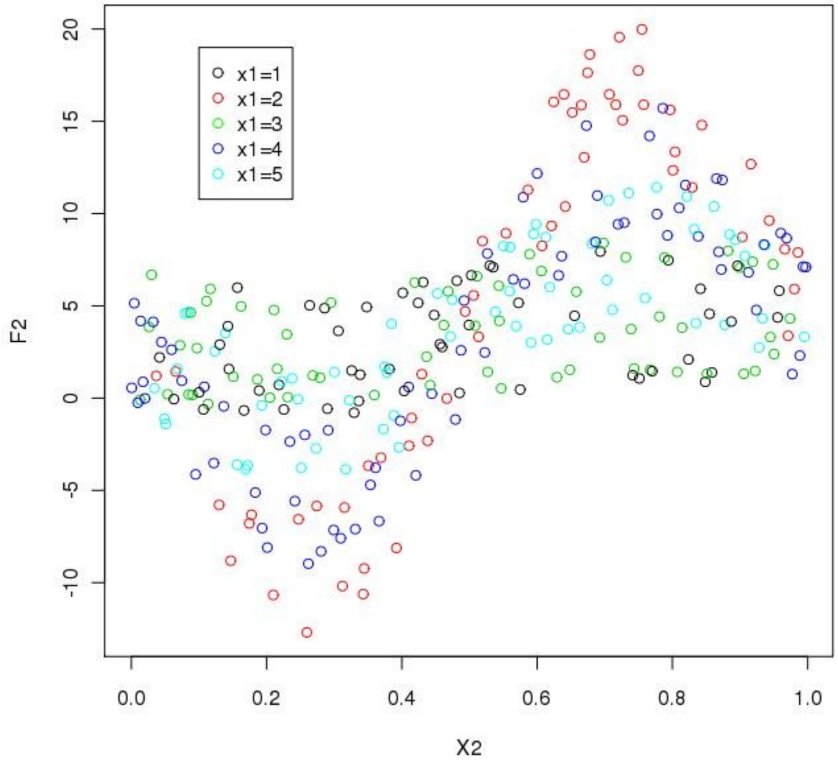
Testbed: Test Function 1



$$y = f_1(x) = \begin{cases} 3.5(x_2 + 0.5)^4 \text{ if } x_1 = 1 \\ 0.5(x_2 + 0.5)^4 \text{ if } x_1 = 2 \\ 2.5(x_2 + 0.5)^4 \text{ if } x_1 = 3 \\ 0.7(x_2 + 0.5)^4 \text{ if } x_1 = 4 \\ (x_2 + 0.5)^4 \text{ if } x_1 = 5 \end{cases}$$

Testbed: Test Function 2

$$y = f_2(x) = \begin{cases} \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) \text{ if } x_1 = 1 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 12\sin(2\pi x_3 - \pi) \text{ if } x_1 = 2 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 0.5\sin(2\pi x_3 - \pi) \text{ if } x_1 = 3 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 8\sin(2\pi x_3 - \pi) \text{ if } x_1 = 4 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 3.5\sin(2\pi x_3 - \pi) \text{ if } x_1 = 5 \end{cases}$$

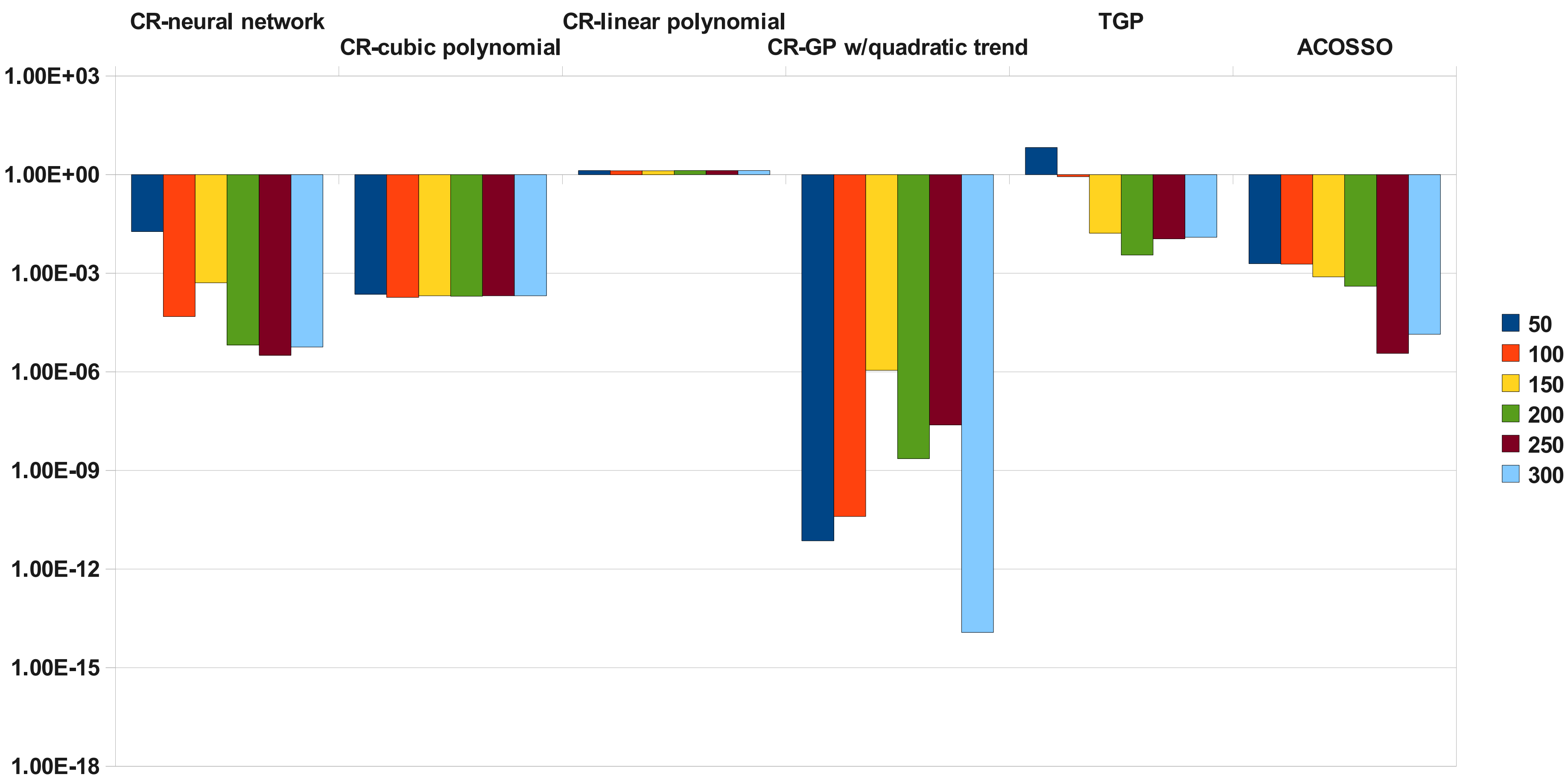


Testbed: Test Function 3

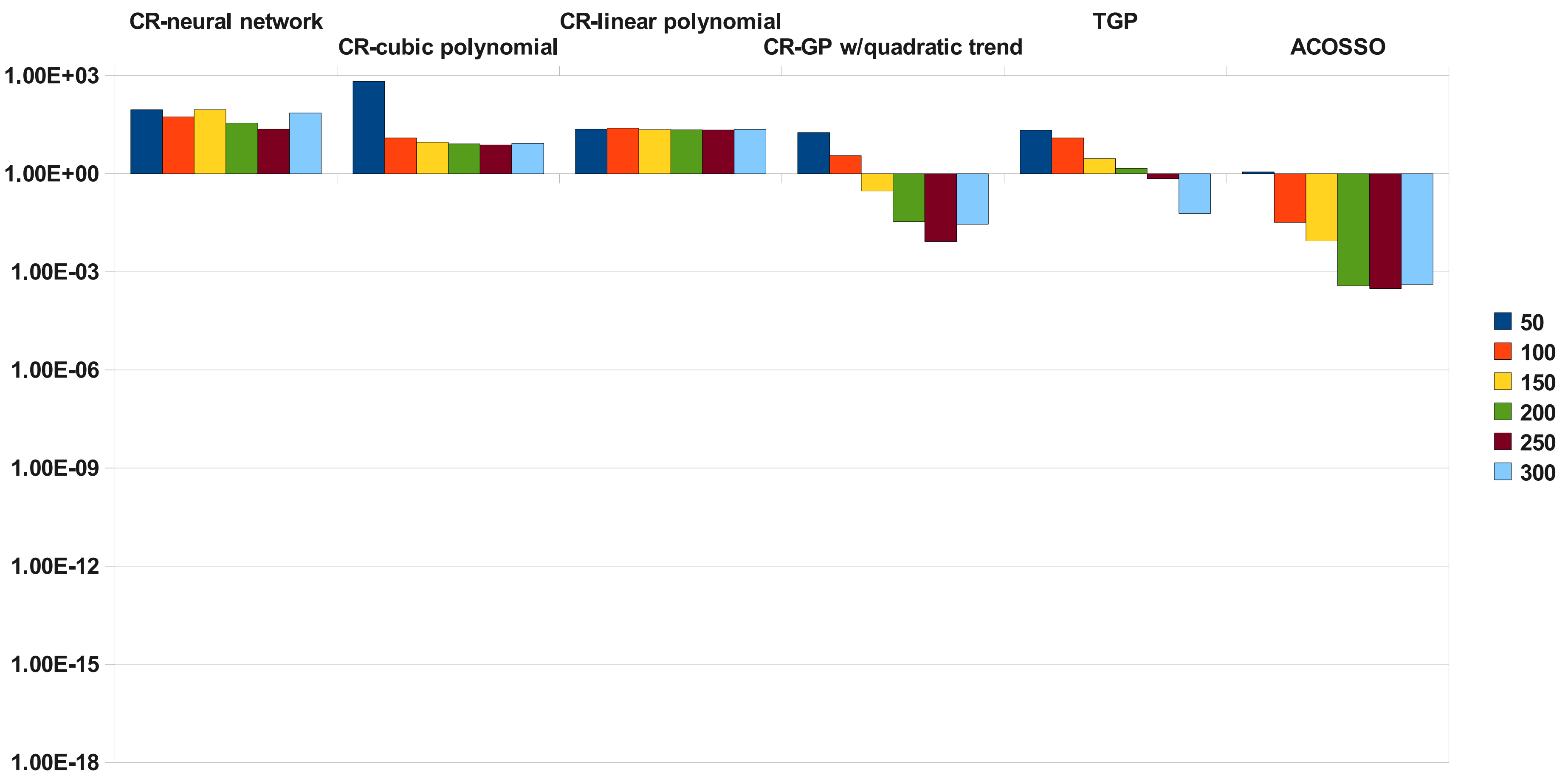
$$y = f_3(x) = \sum_{i=1}^{n} (x_i - 1)^4$$

- Initially started with 4 variables
  - 2 continuous on [0,2]
  - 2 discrete with values [0,1,2]
- Easy to scale up number of levels
  - Scaled number of levels to 5, with values [-1,0,1,2,3]
- Easy to scale up number of discrete variables
  - Scaled up to 5 discrete variables, with 3 and 5 levels
- Can also explore symmetry and function separability

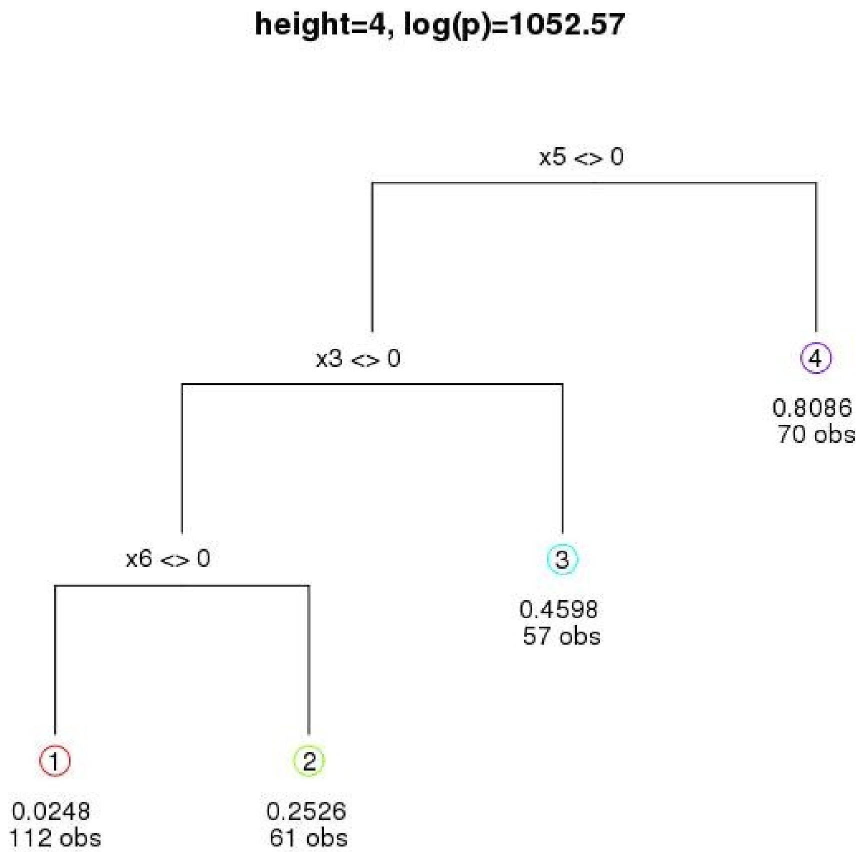
Test Function 1: Categorical regression works quite well, especially using GP



Test Function 2: All approaches have trouble resolving the categorical levels

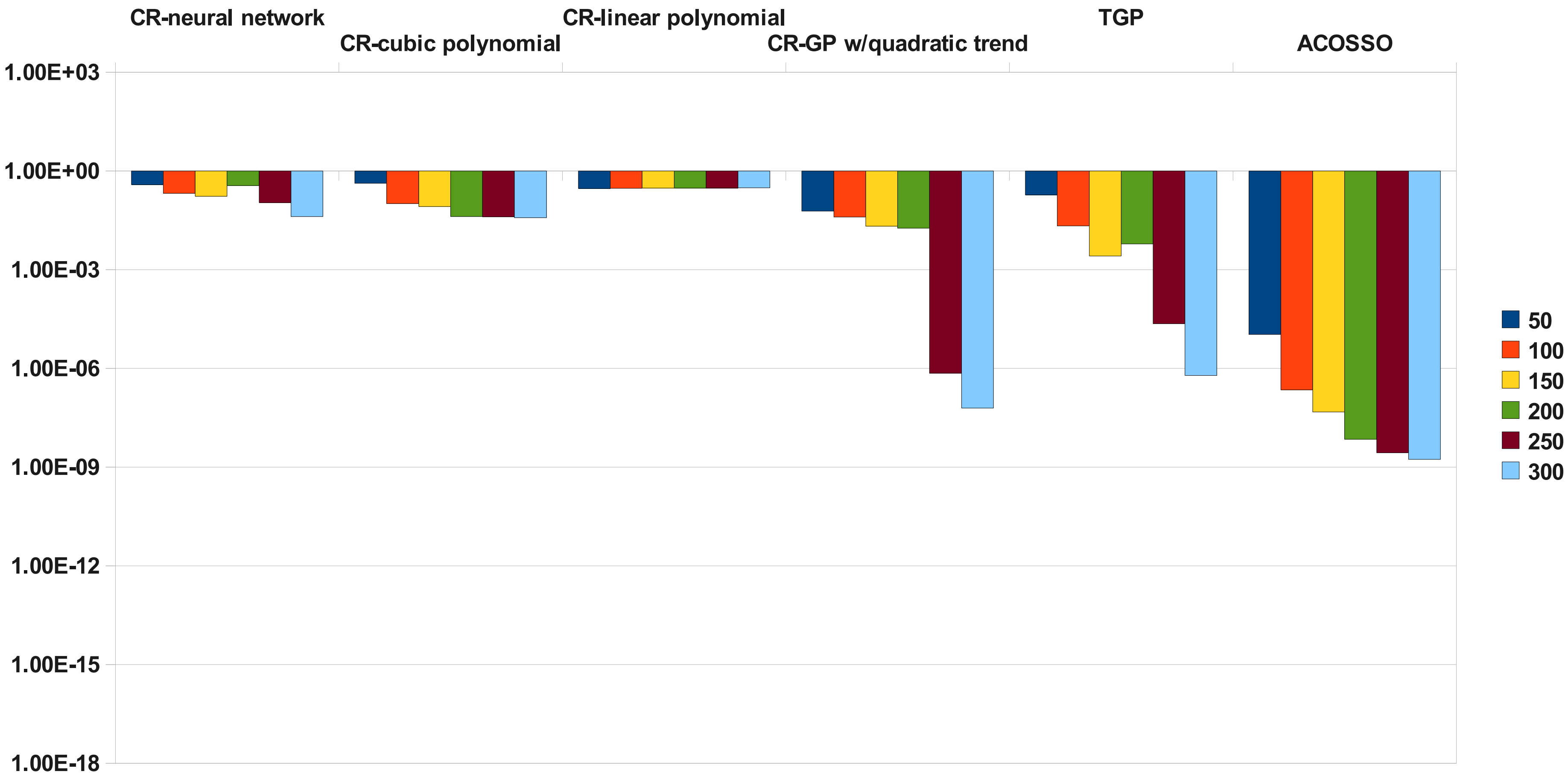


Aggregation of categorical levels in TGP may be a disadvantage



- TGP does not fully partition over all discrete variables
- Premise was that it would be sufficient to create surrogates over partitions which aggregate the discrete variables
- Perhaps too coarse, resulting in inaccurate surrogates

Test Function 3: Categorical regression starts to degrade with slight increase in dimension



Test Function 3: TGP scales better with number of discrete variables than number of levels

- Scaling up discrete levels from 3 to 5
- Scaling up discrete variables from 2 to 5

Test Function 3	SYMMETRIC - TGP			
Discrete	2 [0-1-2]	2 [-1-0-1-2-3]	5 [0-1-2]	5 [-1-0-1-2-3]
Continuous	2 [0,2]	2 [0,2]	2 [0,2]	2 [0,2]
50	0.7217199	119.75	1.38	319.22
100	0.03391995	57.15	0.79	300.08
150	0.01617074	25.94	0.87	272.76
200	0.00631333	25.26	0.72	265.96
300	4.45E-05	17.91	0.52	231.41
500	1.74E-06	1.27	0.32	223.68

MSE decreases more quickly for more discrete variables vs. an increased number of levels per variable.

Test Function 3: Same trend holds for ACOSSO, though it still performs well

- Scaling up discrete levels from 3 to 5
- Scaling up discrete variables from 2 to 5

Test Function 3	SYMMETRIC - ACOSSO			
Discrete	2 [0-1-2]	2 [-1-0-1-2-3]	5 [0-1-2]	5 [-1-0-1-2-3]
Continuous	2 [0,2]	2 [0,2]	2 [0,2]	2 [0,2]
50	1.20E-04	8.15E-04	2.24E-04	2.56E-01
100	9.31E-06	1.55E-03	6.27E-06	4.06E-06
150	1.50E-06	1.34E-03	2.01E-06	2.56E-04
200	1.75E-07	3.20E-06	6.97E-07	1.99E-03
300	3.17E-07	4.68E-05	1.31E-07	5.24E-05
500	7.69E-08	3.08E-04	8.56E-08	2.14E-05

MSE decreases more quickly for more discrete variables vs. an increased number of levels per variable.

Test Function 3: Asymmetry has more adverse effects on TGP than ACOSSO

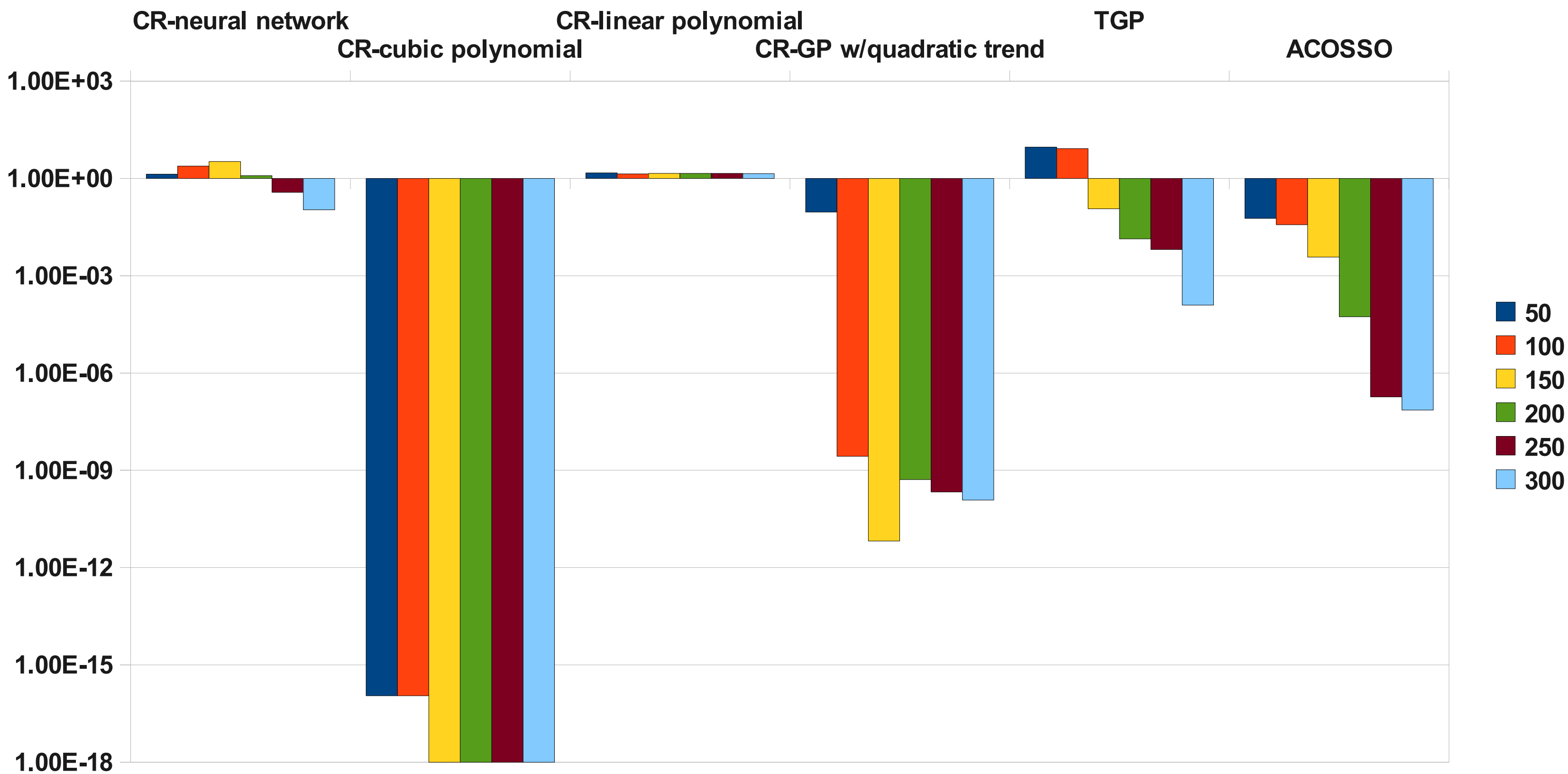
- Scaling up discrete levels from 3 to 5
- Scaling up discrete variables from 2 to 5
- Function is now asymmetric

Test Function 3	TGP		ACOSSO	
Discrete	2 [1-2-3]	2 [1-2-3-4-5]	2 [1-2-3]	2 [1-2-3-4-5]
Continuous	2 [0,2]	2 [0,2]	2 [0,2]	2 [0,2]
50	53.53	11843.35	3.67E-03	0.34
100	0.62	2444.52	1.96E-03	0.18
150	0.26	3402.73	3.27E-04	0.12
200	0.15	4494.97	9.94E-04	0.07
300	0.02	2382.42	3.55E-06	0.03
500	0.01	0.39	5.78E-04	0.06

Separability and lack of variable interactions, particularly between discrete and continuous, may be playing to the strengths of ACOSSO.

Polynomial Function: Categorical regression performs quite well, followed by ACOSSO

- 2^{nd} order polynomial with 4 variables
  - 2 discrete variables at levels [20,50,80]
  - 2 continuous variables between 0 and 100



Polynomial Function: Complexity has more adverse effects on ACOSSO than TGP

- "Scaling" Problem Complexity
  - 2nd order polynomial with 14 terms
  - 3rd order polynomial with 24 terms
  - 4th order polynomial with 19 terms
- 10 discrete levels (instead of 3)

	Test Function Poly 2		Test Function Poly 3		Test Function Poly 4	
	TGP	**ACOSSO**	**TGP**	**ACOSSO**	**TGP**	**ACOSSO**
Discrete	2 [ten levels]	2 [ten levels]	2 [ten levels]	2 [ten levels]	2 [ten levels]	2 [ten levels]
Continuous	2[0,100]	2[0,100]	2[0,100]	2[0,100]	2[0,100]	2[0,100]
50	28.88	12.24	25.16	9.12	10.07	29.98
100	28.58	0.46	25.00	4.92	10.25	5.14
150	27.83	0.15	21.94	2.31	13.16	6.03
200	21.80	0.05	16.31	1.91	9.99	5.03
250	24.92	0.06	11.58	2.37	10.40	5.16
300	22.78	0.03	9.49	1.77	8.76	5.20

Possible increase in variable interactions and increased relative impact of continuous variables may be playing to the strengths of TGP.

Observations and Summary Thoughts

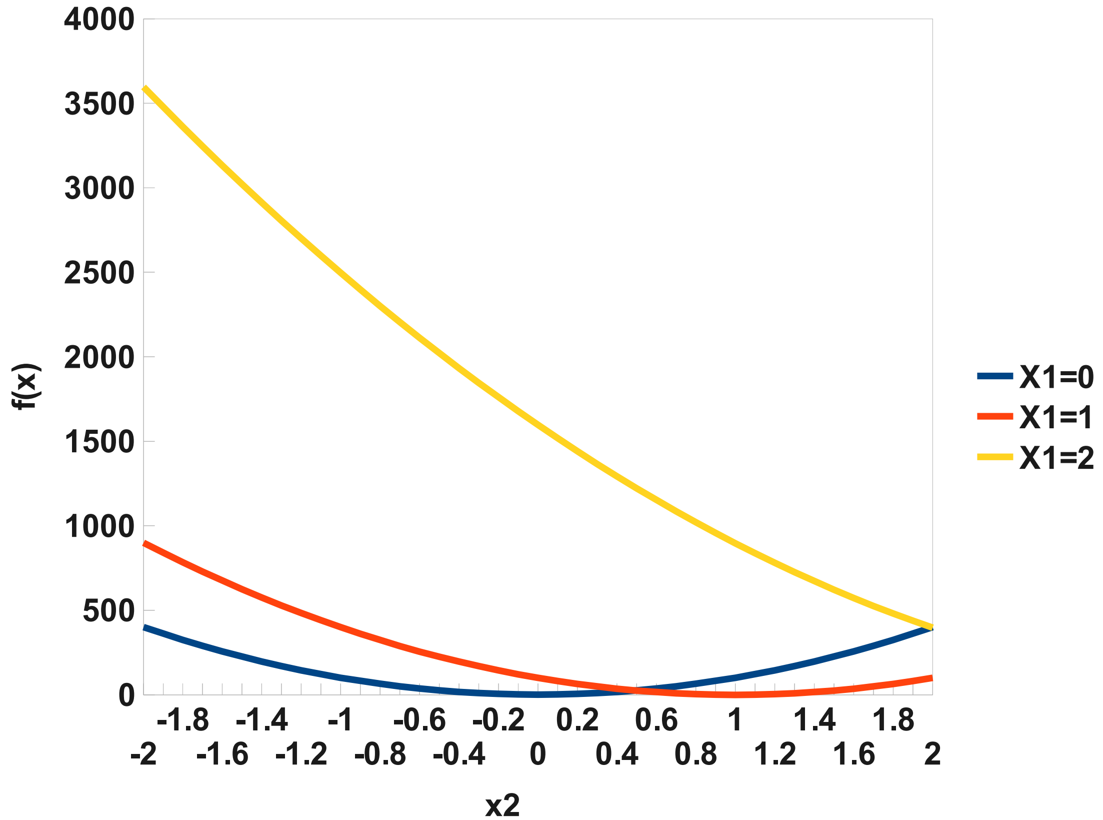
- Surrogate models are imperative for computational tractability of engineering analyses
- Categorical Regression performed very well on problems with small numbers of discrete variables/levels
- ACOSSO performs very well overall
- TGP performance is mixed
  - Functions 1, 2: performs well when it seems to get enough function evaluations (few hundred)
    - Ability to identify the splits
    - Not sufficient to aggregate across discrete levels
  - Functions 3, poly: performs poorly (i.e, high MSE)
    - Adaptive methods may lead to improvement

Observations and Summary Thoughts (2)

- Scalability
  - ACOSSO seems the most scalable
  - TGP suffers from too large an aggregation across discrete levels
  - Categorical regression is not scalable
  - Is there a difference between scalability across discrete variables vs. number of levels? Test function three suggests there might be
- Further work:
  - Amount of interaction between variables
  - Range/nonlinearity of function
  - Improving efficiency of TGP and ACOSSO implementations

We get similar results on a 4th-order Schittkowski test function

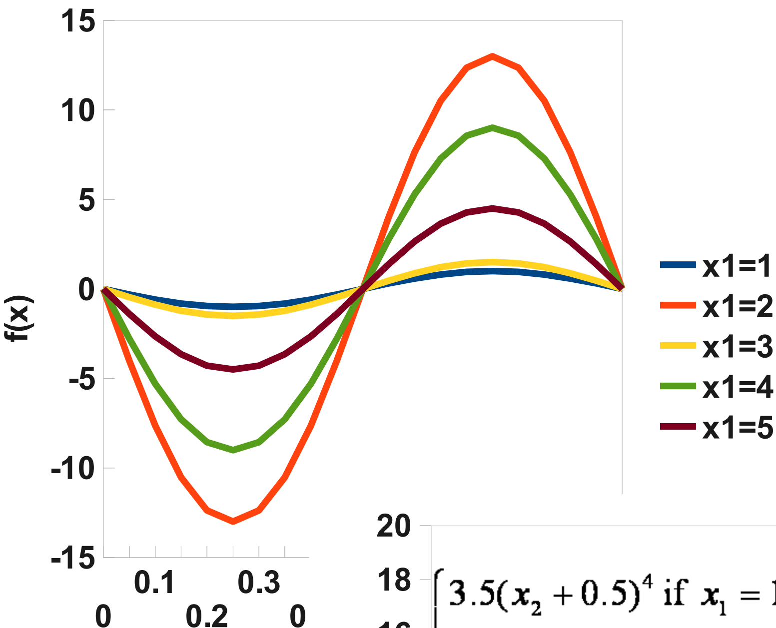
- $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1^2)$
- Shows both interactions and scale effects



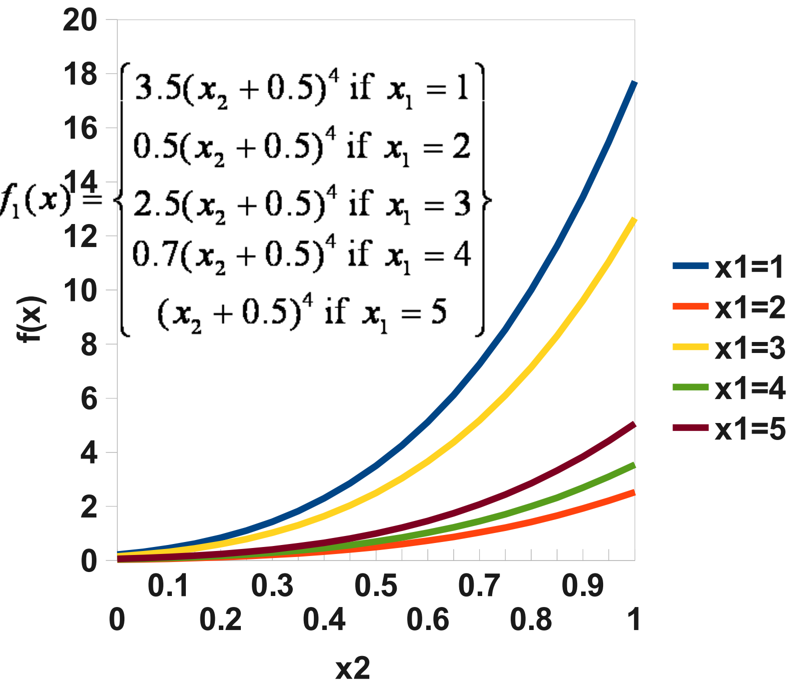
We established requirements for a testbed

- Fast running evaluations
- Easy to compile, cross-platform compatibility
- Extendable
- File input/output
- Scalability of function in terms of number discrete variables and/or levels per variable
- Ability to control problem complexity

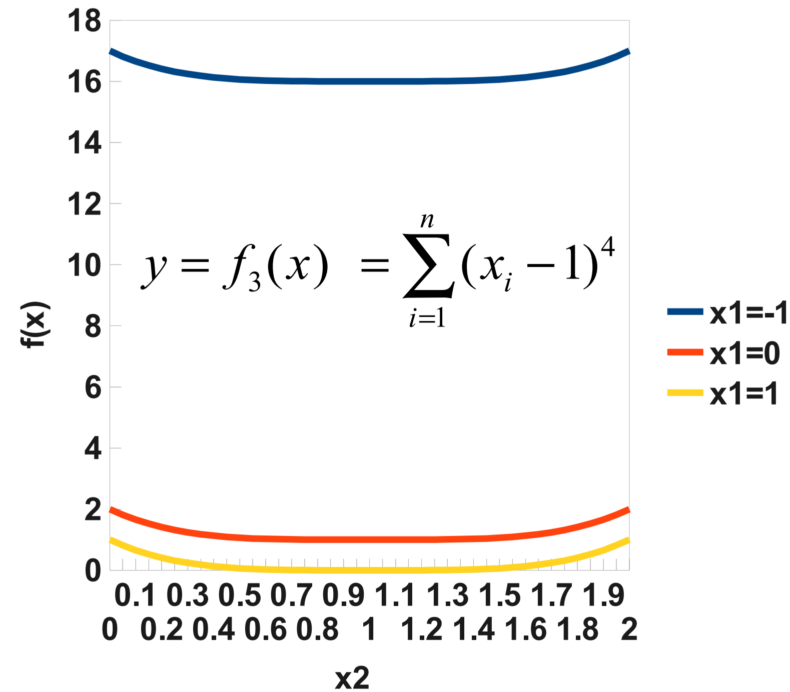
Testbed includes three defined functions



$$y = f_2(x) = \begin{cases} \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) \text{ if } x_1 = 1 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 12\sin(2\pi x_3 - \pi) \text{ if } x_1 = 2 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 0.5\sin(2\pi x_3 - \pi) \text{ if } x_1 = 3 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 8\sin(2\pi x_3 - \pi) \text{ if } x_1 = 4 \\ \sin(2\pi x_3 - \pi) + 7\sin^2(2\pi x_2 - \pi) + 3.5\sin(2\pi x_3 - \pi) \text{ if } x_1 = 5 \end{cases}$$



$$y = f_1(x) = \begin{cases} 3.5(x_2 + 0.5)^4 \text{ if } x_1 = 1 \\ 0.5(x_2 + 0.5)^4 \text{ if } x_1 = 2 \\ 2.5(x_2 + 0.5)^4 \text{ if } x_1 = 3 \\ 0.7(x_2 + 0.5)^4 \text{ if } x_1 = 4 \\ (x_2 + 0.5)^4 \text{ if } x_1 = 5 \end{cases}$$



$$y = f_3(x) = \sum_{i=1}^{n} (x_i - 1)^4$$

Testbed also includes a random polynomial generator

- Generates a random polynomial
  - Degree between 2 and 6
  - Number of variables between 1 and 15
- Uses a system of linear equations to solve for the random coefficients, described in:
  - McDaniel, W. R. and B. E. Ankenman, “A Response Surface Test Bed.” Qual. Reliab. Engng. Int. 2000; 16: 363–372
- Can control the degree of nonlinearity, range of polynomial values, etc.

We applied the same evaluation process to all candidate surrogates using these test problems

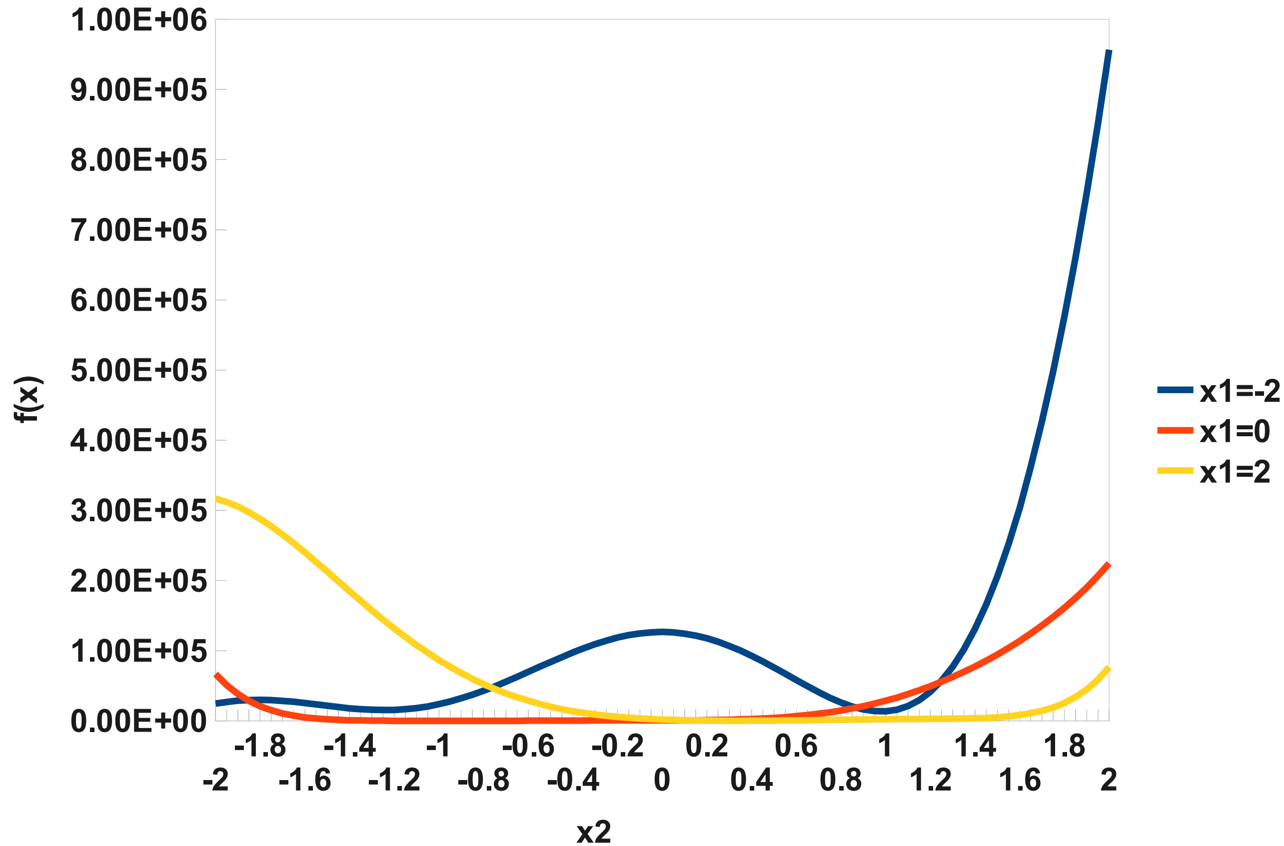
- Looked at surrogate performance over varying number of build points (LHS sample points)
- Used mean squared error (MSE) as a measure of goodness
  - Calculated over a grid (dimensioned based on the number of inputs)
- Categorical Regression run in DAKOTA
  - Generate a separate continuous surrogate for each combination of discrete variable values/levels
- TGP and ACOSSO run in R

Observations after first set of preliminary experiments were enlightening

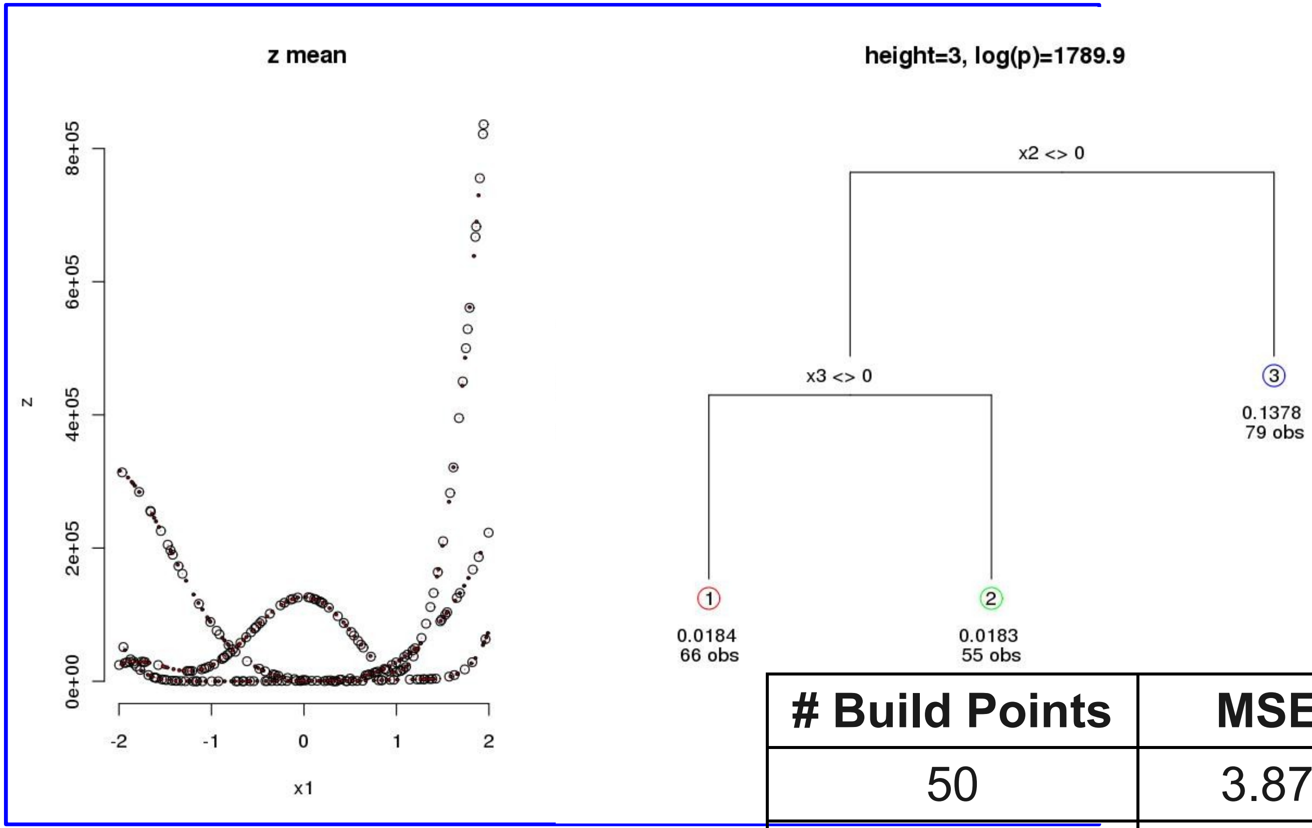
- Categorical Regression performed very well on problems with small numbers of discrete variables/levels
- ACOSSO performed very well overall
- TGP performance was mixed
  - Ability to identify the splits
  - Not sufficient to aggregate across discrete levels
  - Adaptive methods may lead to improvement
- ACOSSO seemed to be the most scalable with respect to number of variables and discrete levels
- Need to further investigate the effects of parameter interactions

Goldstein-Price function has parameter interactions and widely varying scale

- Function of 2 variables ranging over six orders of magnitude
- $f(x) = (1 + (x_1 + x_2 + 1)^2 * (19 – 14x_1 + 3x_1^2 – 14x_2 + 6x_1x_2 + 3x_2^2) * (30 + (2x_1 - 3x_2)^2 * (18 – 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2))$



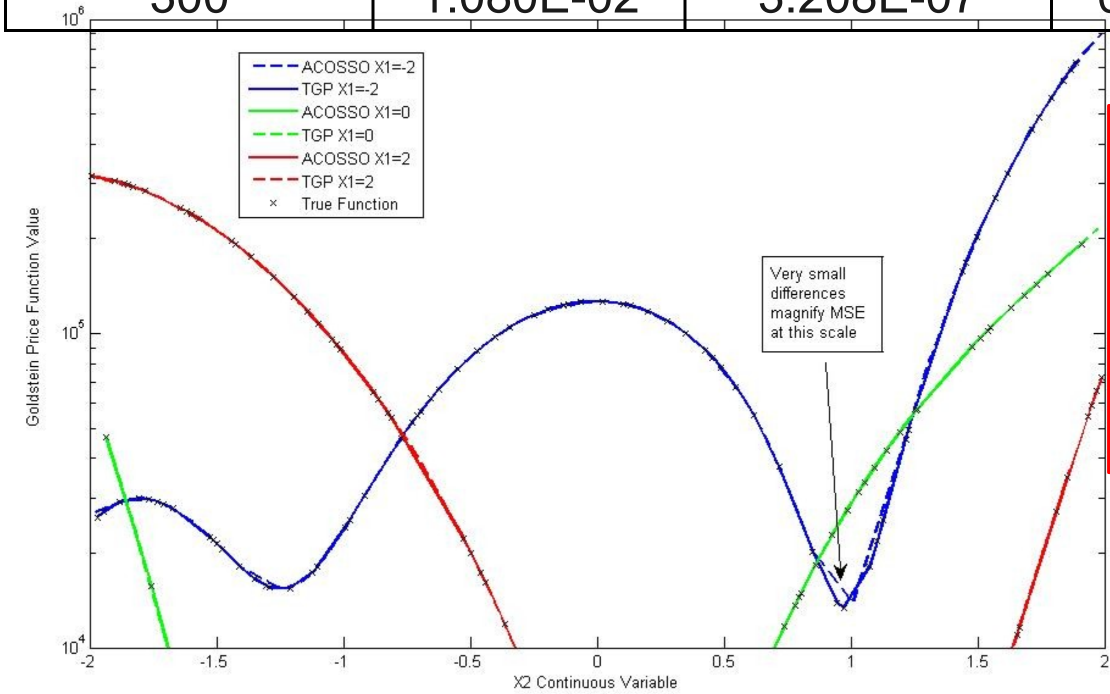
Partitioning enables TGP to perform better when response range is large



# Build Points	MSE TGP	MSE ACOSSO
50	3.87E+07	2.42E+07
100	3.89E+04	9.67E+06
150	2.50E+03	6.67E+06
200	9.51E+01	5.33E+06
250	4.39E+01	1.55E+06
300	2.20E+03	2.14E+04

Log scaling of response and normalized metrics put the differences in perspective

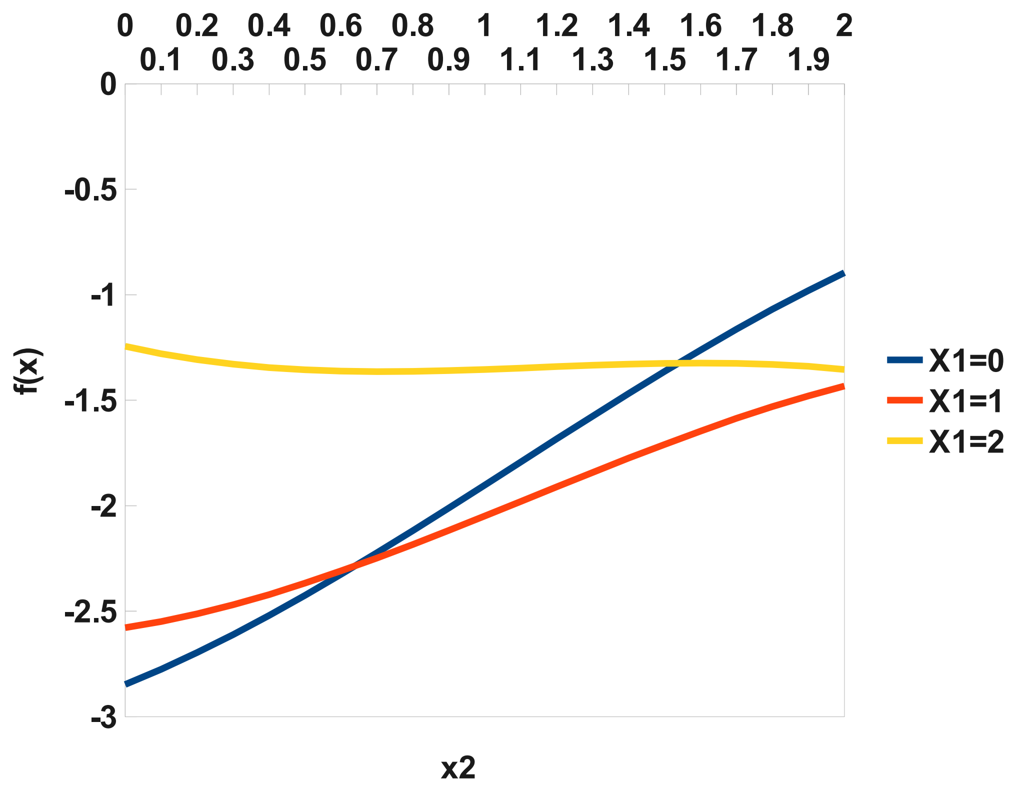
# Build Points	MSE TGP (log-scaled)	MSE ACOSSO (log-scaled)	R-sq TGP	R-sq ACOSSO
50	4.051E-01	1.164E-01	0.99738	0.99821
100	1.036E-02	5.991E-03	0.99999	0.99886
150	7.236E-03	8.764E-03	0.99999	0.99953
200	1.778E-04	3.718E-04	1.00000	0.99968
250	5.379E-03	1.803E-04	1.00000	0.99988
300	1.080E-02	3.208E-07	0.99999	1.00000



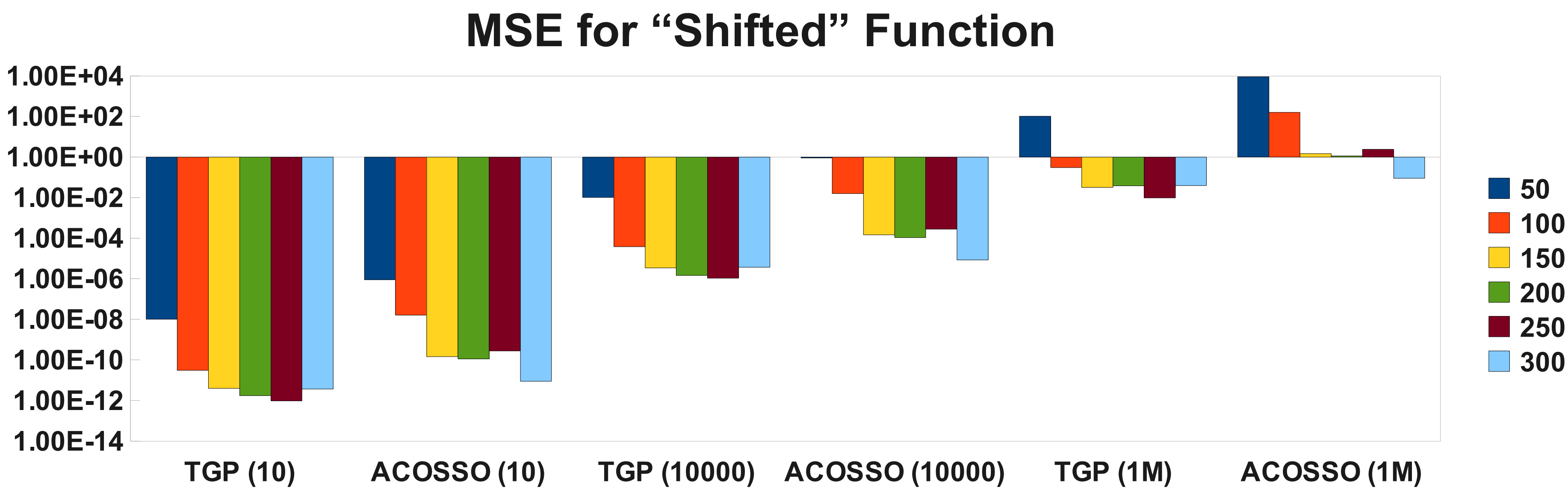
Whether or not these differences are important is application dependent. Should consider multiple metrics and drill down.

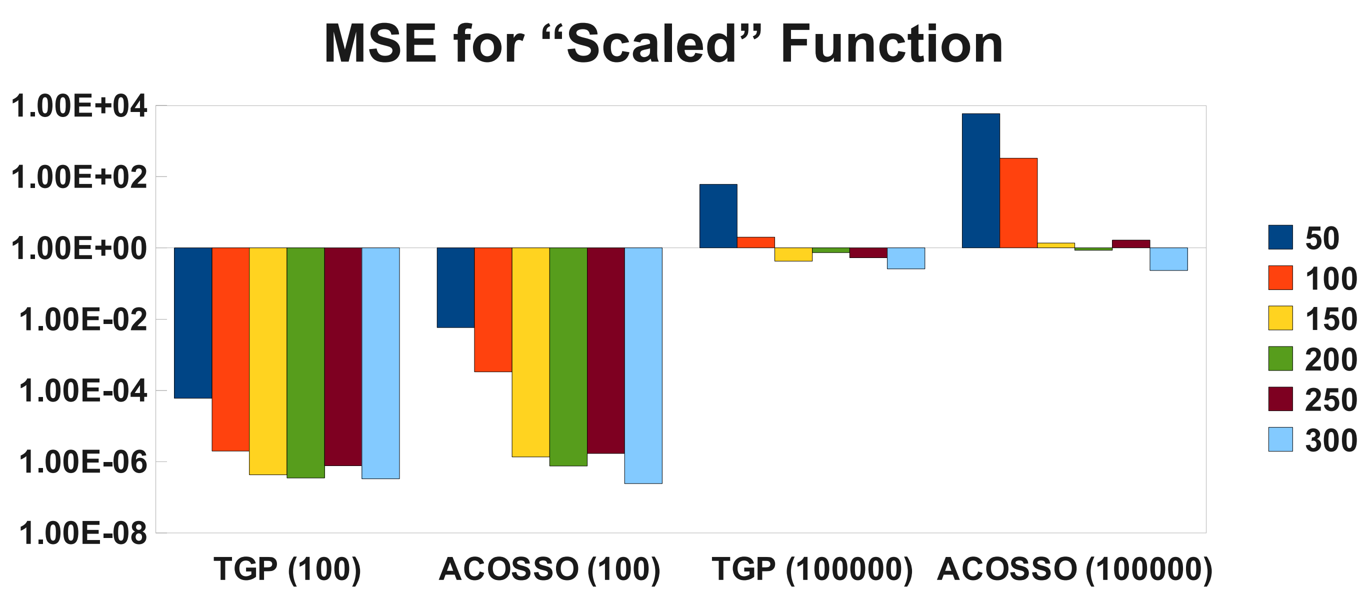
We used a 3rd-order polynomial to study effect of response magnitude and range

- 10 terms, significant interaction
- Looked at two aspects
  - increasing the magnitude of the overall function
  - increasing the range



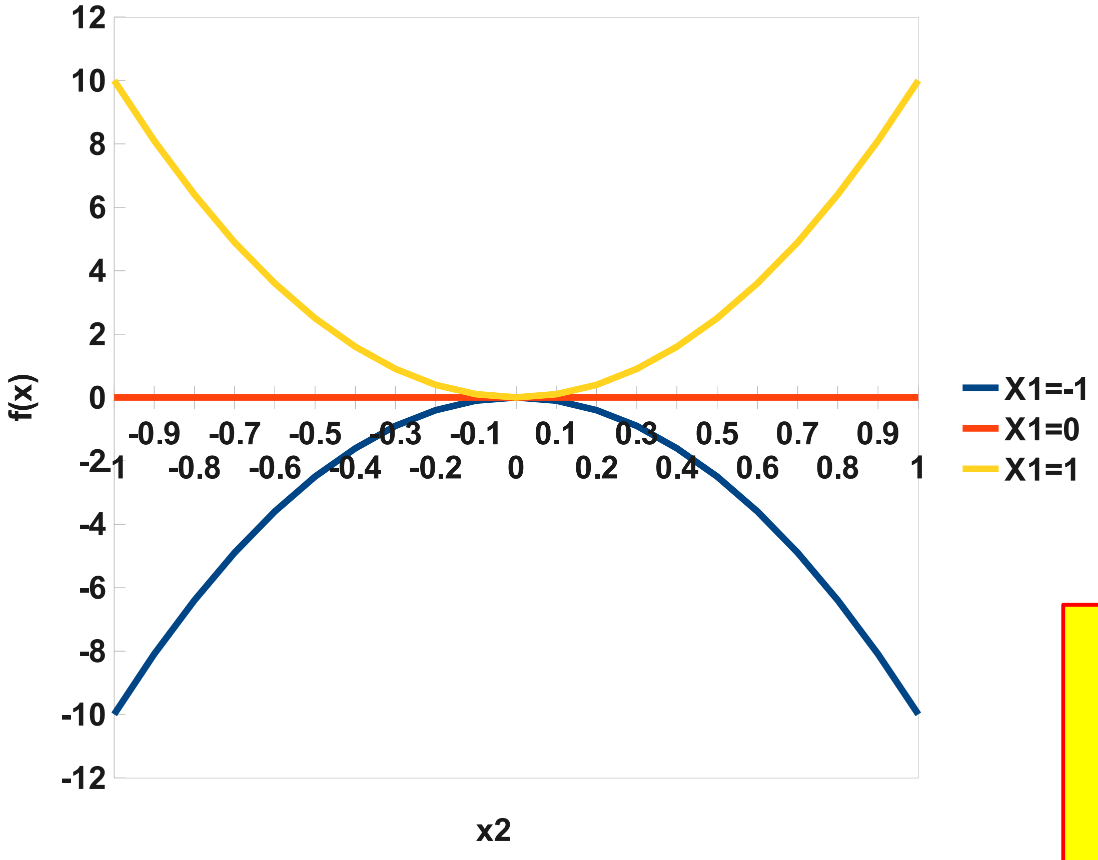
TGP performs better when magnitude or range is increased, especially with fewer build points

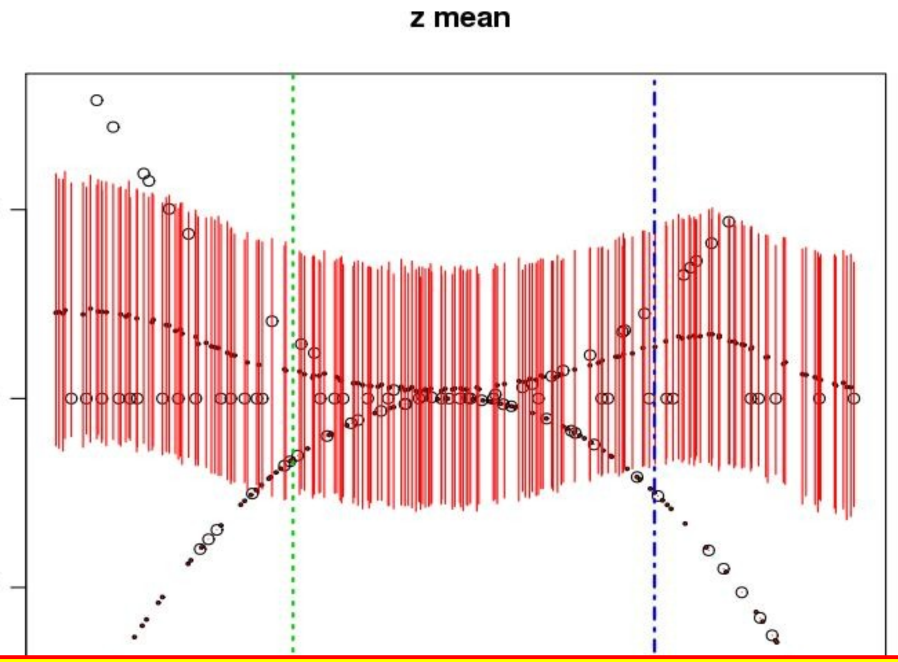




The simple function $f(x)=10x_1x_2^2$ proved more interesting than expected

- We thought this would be trivial
- Constant line at $x_1=0$ proved to be pathological for TGP





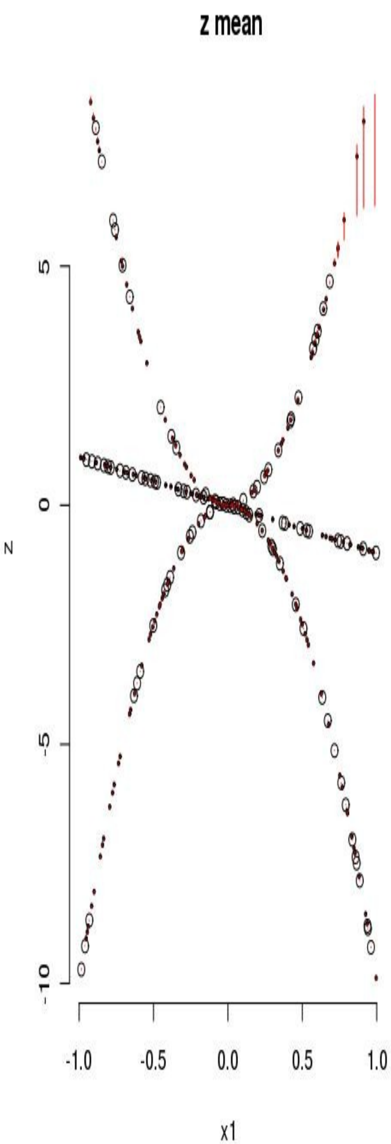
TGP cannot resolve the difference between $x_1=0$ and $x_1=1$

TGP failure is caused by lack of predictivity in continuous variable

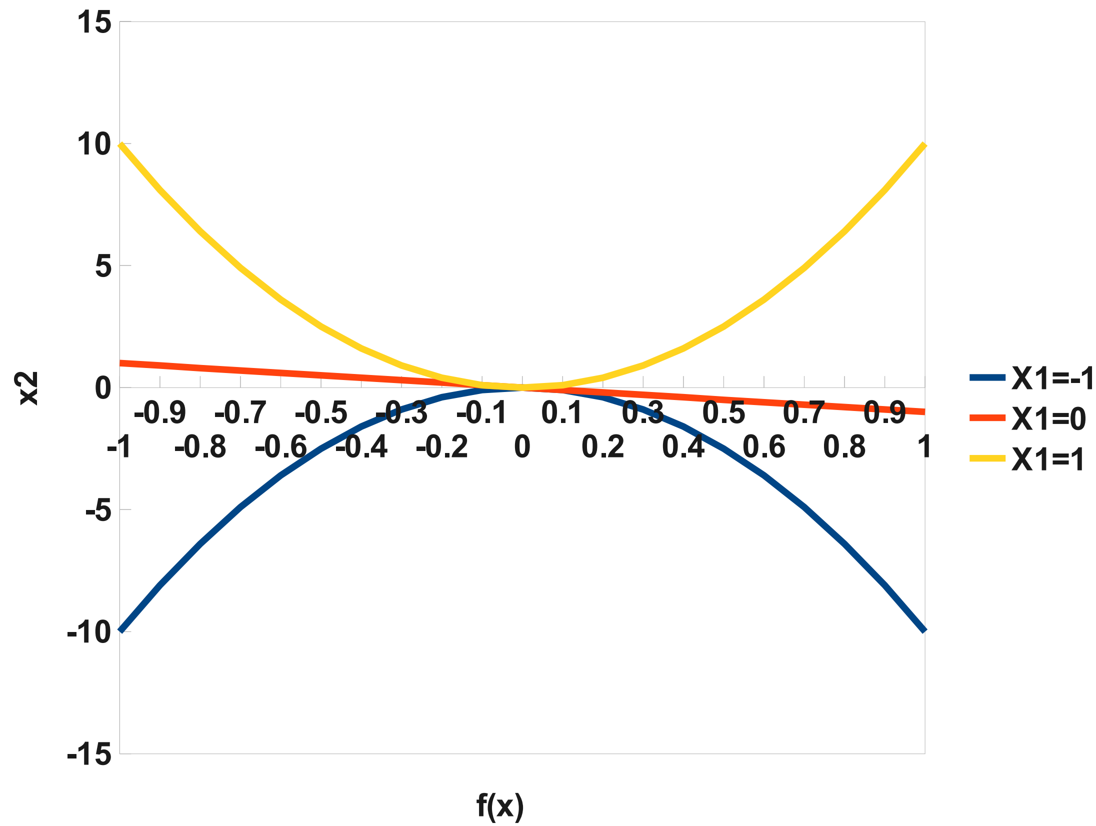
- x_2 is not predictive when $x_1 = 0$
- Constructing GP over both treats the categorical as continuous

# Build Points	MSE TGP (continuous)	MSE TGP (both)	MSE ACOSSO
50	4.28E+00	5.97E-06	4.71E-05
100	4.20E+00	5.04E-08	2.80E-06
150	3.71E+00	1.55E-08	2.84E-05
200	3.45E+00	1.32E-08	3.62E-07
250	3.56E+00	1.91E-08	1.71E-07
300	3.88E+00	**	1.49E-07

Slight variation on $10x_1x_2^2$ causes no trouble for TGP



- Change line f(x) = 0 at $x_1 = 0$ to f(x) = $-x_2$
- TGP has no problem in this case



Observations, the sequel

- Surrogate models are imperative for computational tractability of engineering analysis
- ACOSSO continues to have the most consistent performance
- TGP is best performer for problems with responses that vary over large orders of magnitude (when unscaled)
- Still need to investigate
  - Effects of variable interactions
  - Degree of variable predictivity needed for TGP
  - Effects of degree/type of nonlinearity
  - Improving efficiency of TGP and ACOSSO implementations