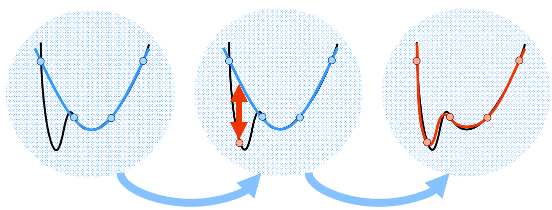
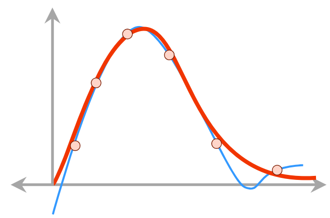




$$\begin{aligned} \min \quad & f(x,y) \\ \text{s.t.} \quad & g(x,y) \leq 0 \\ & h(x,y) = 0 \end{aligned}$$

ALAMO: Machine learning from data and first principles

Nick Sahinidis

Georgia Institute of Technology

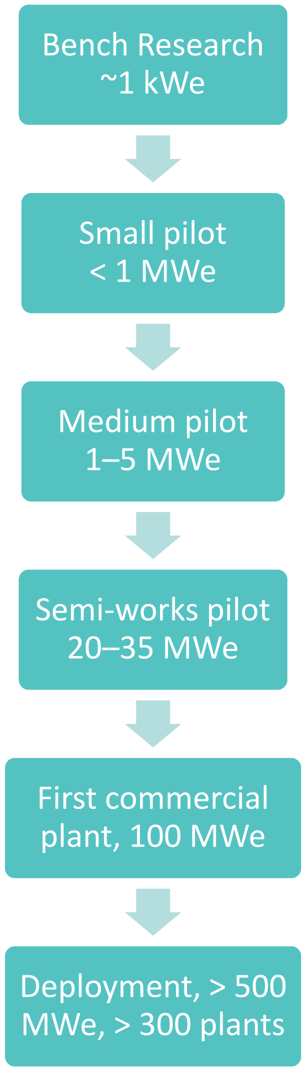
Industrial & Systems Engineering and Chemical & Biomolecular Engineering

Acknowledgements:

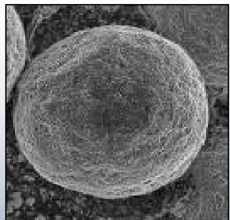
Alison Cozad, Owais Sarwar, Zach Wilson; David Miller

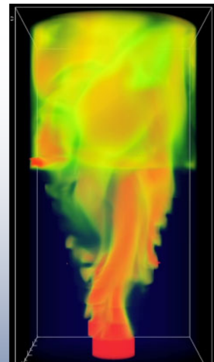
CARBON CAPTURE CHALLENGE

- **The traditional pathway from discovery to commercialization of energy technologies can be quite long, i.e., ~2-3 decades**
- **President Obama's plan required that barriers to the widespread, safe, and cost-effective deployment of CCS be overcome within 10 years**
- **New approaches are needed for taking carbon capture concepts from lab to power plant, <u>quickly</u>, and at low cost and risk**
- **CCSI was launched to accelerate the development of carbon capture technology, from discovery through deployment, with the help of science-based simulations**



CARBON CAPTURE SIMULATION INITIATIVE






Identify promising concepts → Reduce the time for design & troubleshooting → Quantify the technical risk, to enable reaching larger scales, earlier → Stabilize the cost during commercial deployment

National Labs

Berkeley Lab
Pacific Northwest National Laboratory
NETL
Los Alamos National Laboratory
Lawrence Livermore National Laboratory

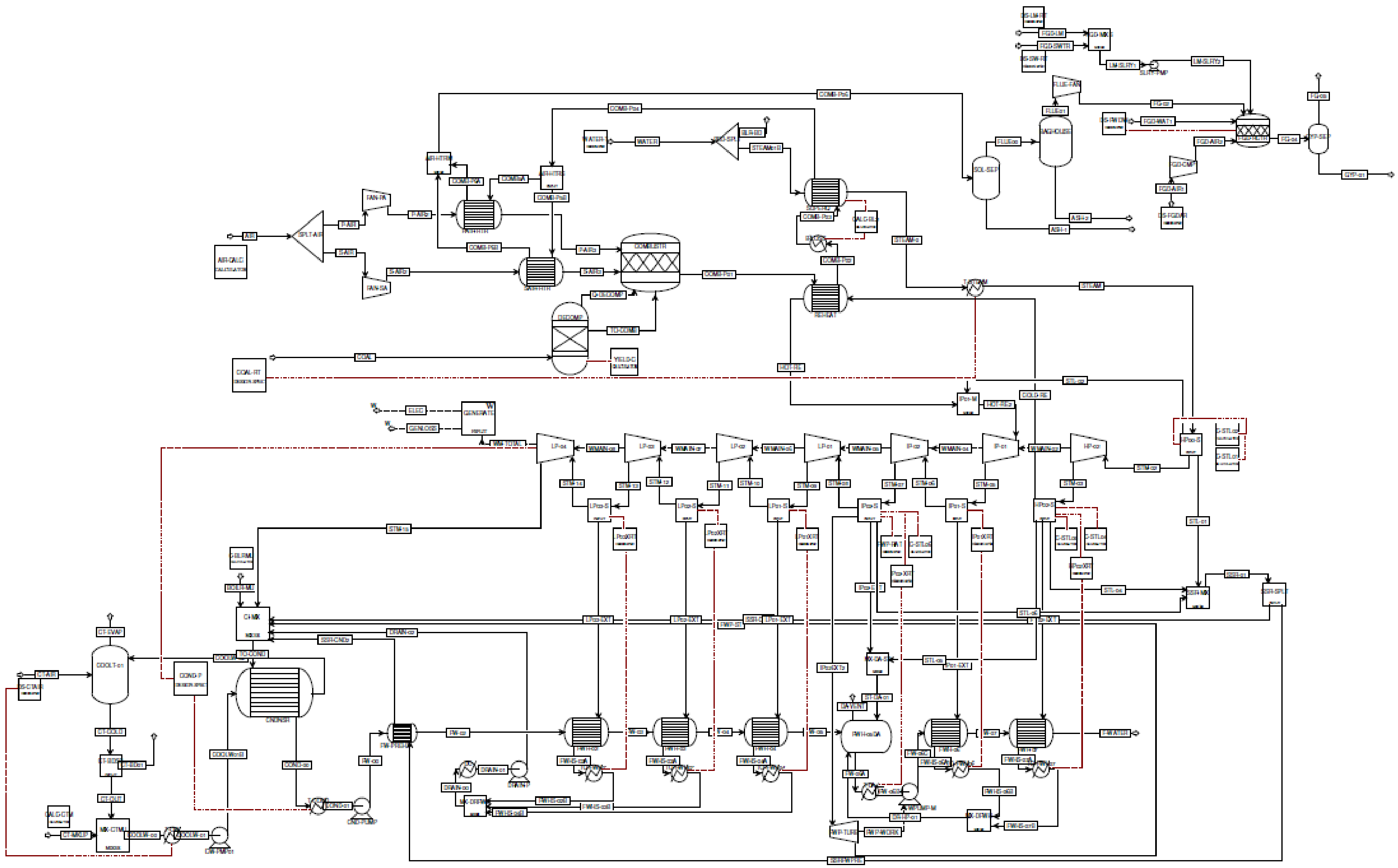
Academia

Carnegie Mellon
Princeton University
West Virginia University
Boston University
Institute for Clean and Secure Energy, The University of Utah

Industry

Fluor
ADA Environmental Solutions
B&W
GE
EPRI Electric Power Research Institute
Air Products
Burns & McDonnell
Southern California Edison
Alstom
Southern Company
Duke Energy
URS
Boeing
DuPont
ExxonMobil
Chevron
Phillips 66
WorleyParsons
Eastman

SIMULATION OPTIMIZATION

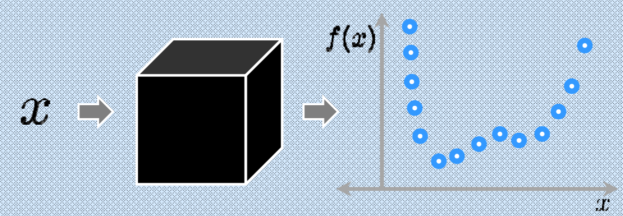


Pulverized coal plant Aspen Plus® simulation provided by the National Energy Technology Laboratory

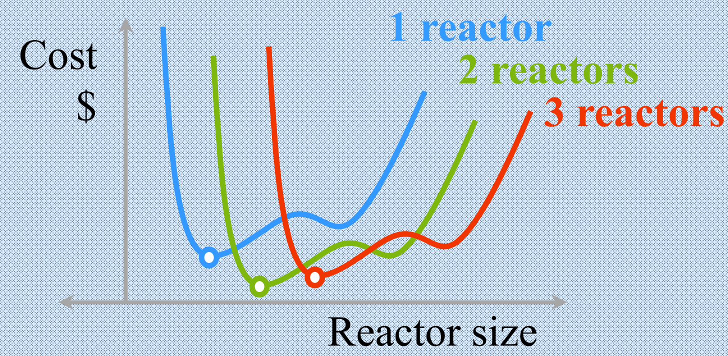
CHALLENGES

OPTIMIZER

No algebraic model

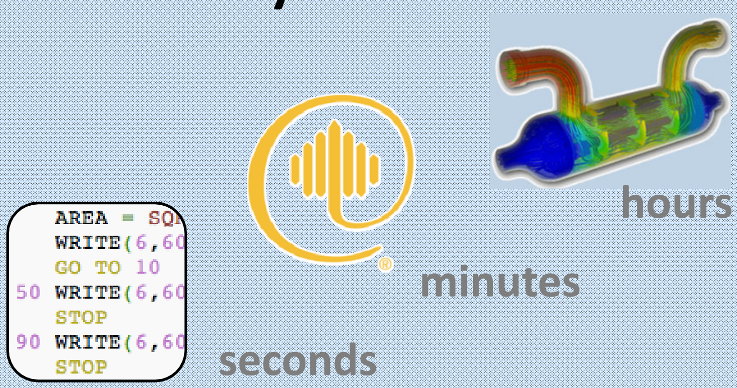


Complex process alternatives



SIMULATOR

Costly simulations

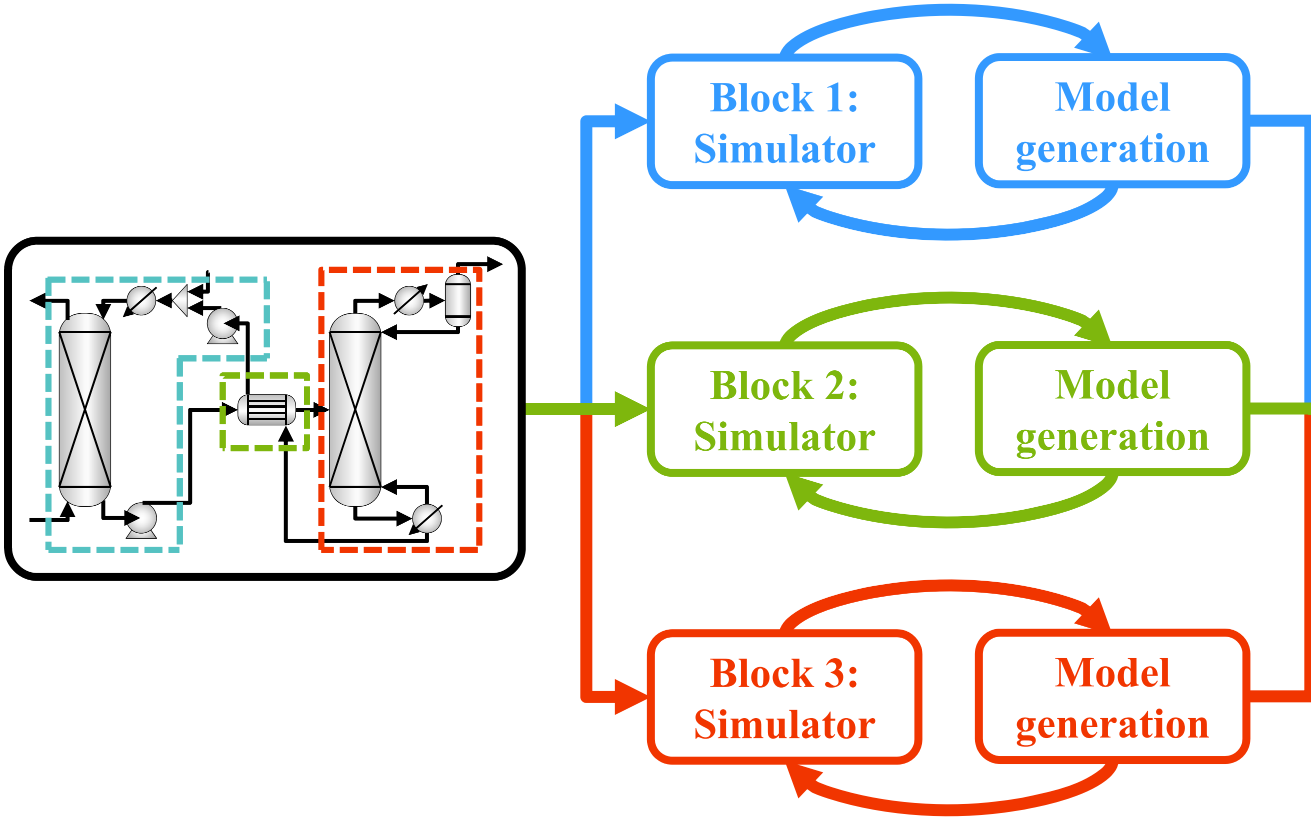


Scarcity of fully robust simulations

✗ Gradient-based methods

✗ Derivative-free methods

PROCESS DISAGGREGATION



$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0 \\ & h(x) = 0 \\ & x \in [x^l, x^u] \end{aligned}$$

Process Simulation

Disaggregate process into process blocks

Surrogate Models

Build simple and accurate models with a functional form tailored for an optimization framework

Optimization Model

Add algebraic constraints design specs, heat/mass balances, and logic constraints

MACHINE LEARNING PROBLEM

Build a model of output variables z as a function of input variables x over a specified interval

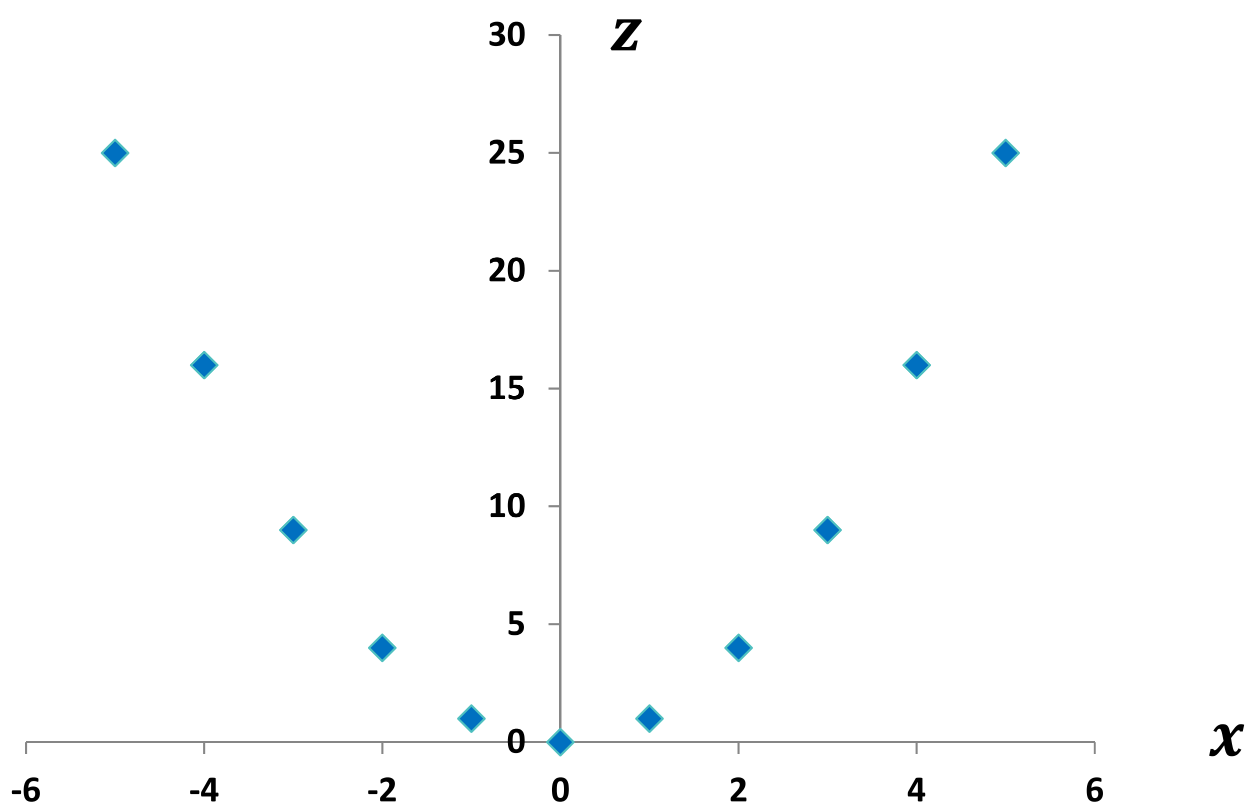
$$x \in \mathbb{R}^k \qquad x^l \le x \le x^u \qquad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_k \end{pmatrix}$$

$$\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_l \\ \vdots \\ z_m \end{pmatrix} \qquad z \in \mathbb{R}^m \qquad z = f(x)$$

Independent variables:
Operating conditions, inlet flow properties, unit geometry, molecular descriptors, etc.

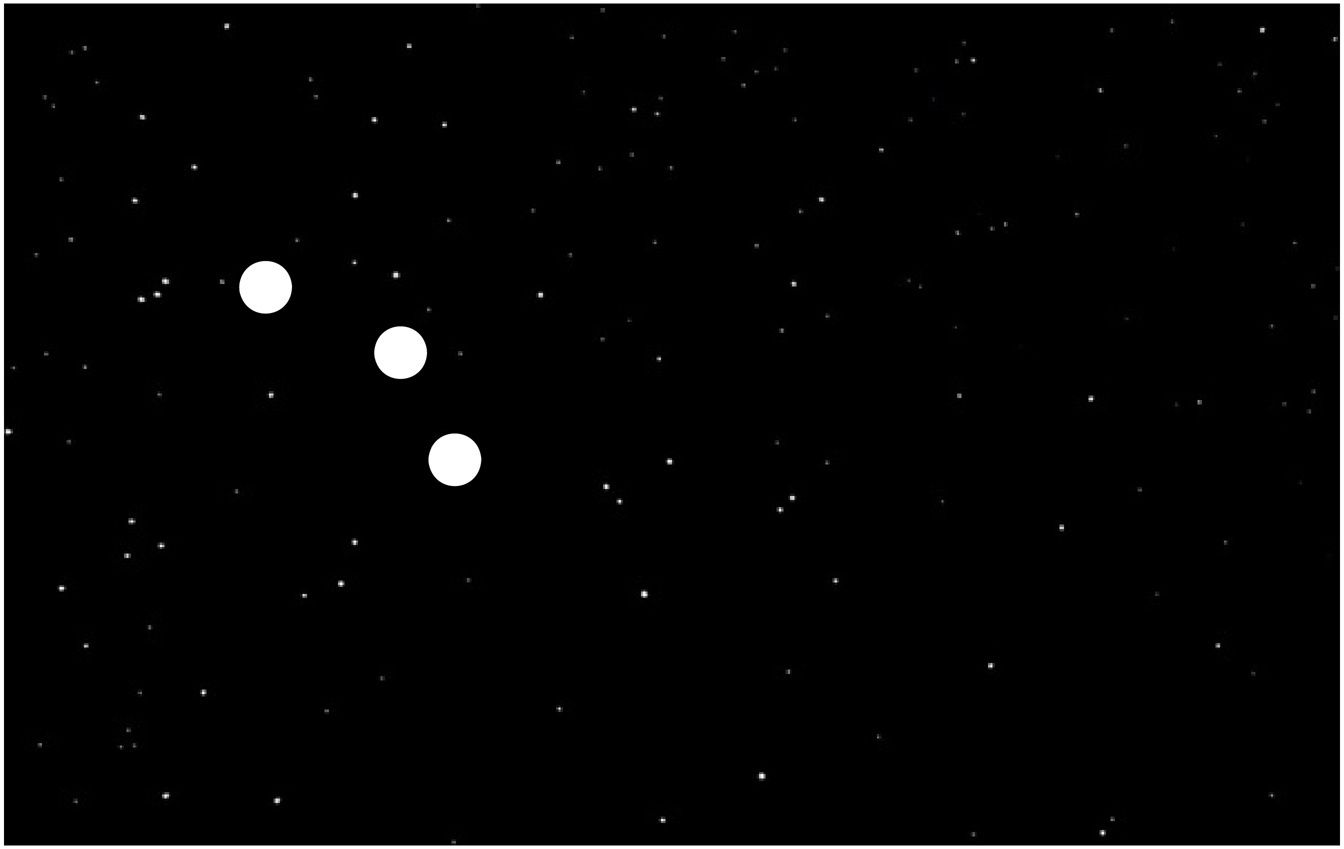
Dependent variables:
Efficiency, outlet flow conditions, conversions, heat flow, chemical potential, etc.

FITTING MODELS TO DATA



$$z = x^2$$

EUROPE IN 1801



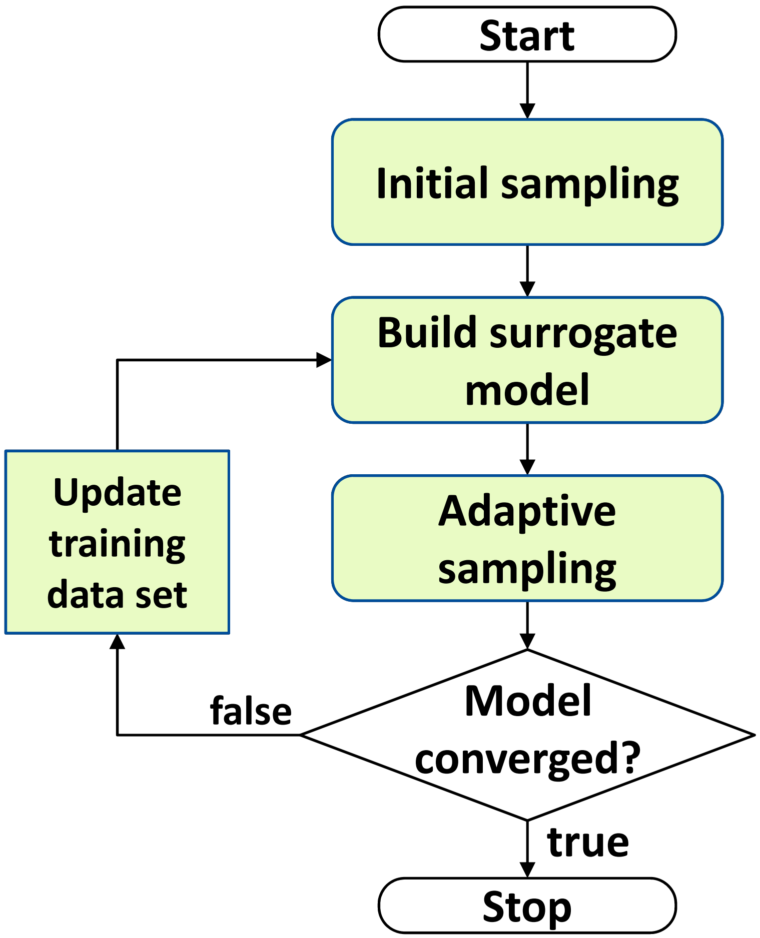
- **Piazzi observed positions of Ceres**
- **Gauss: Least squares**
  - Used observations *and* Kepler's conjecture

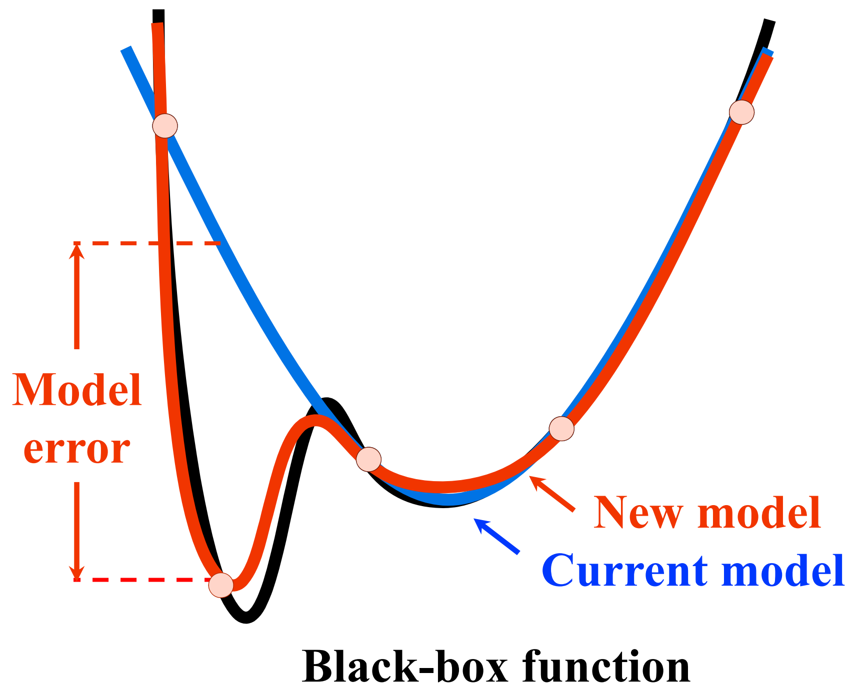
DESIRED MODEL ATTRIBUTES

1. **Accurate**
   - We want to reflect the true nature of the system
2. **Simple**
   - Interpretable
   - Usable for algebraic optimization
3. **Generated from a minimal data set**
   - Reduce experimental and simulation requirements
4. **Obeys physics and user insights**
   - Increase fidelity and validity in regions with no measurements

ALAMO

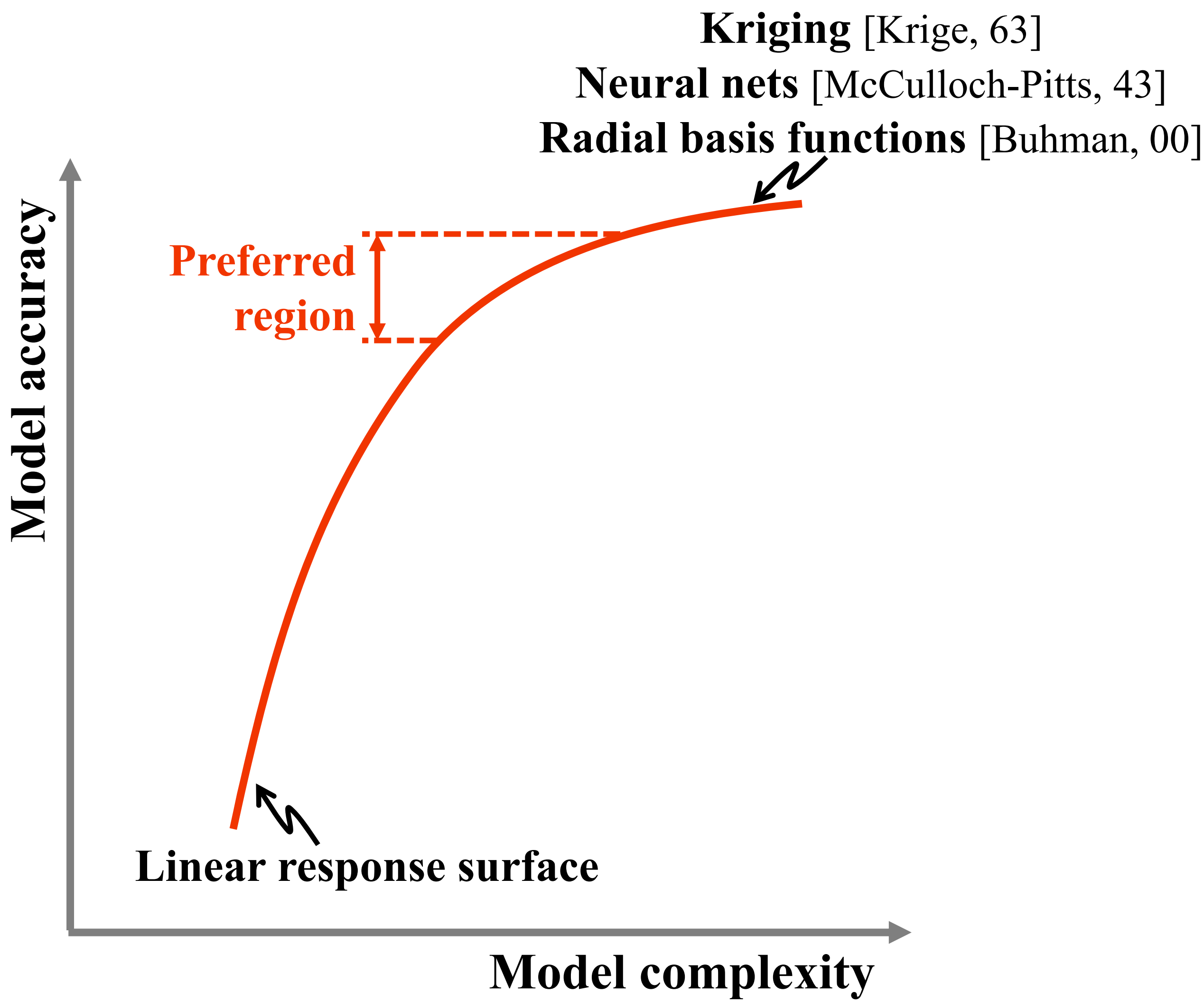
Automated Learning of Algebraic MOdels





Cozad, Sahinidis, Miller, 2014

MODEL COMPLEXITY TRADEOFF



MODEL IDENTIFICATION

- **Identify the functional form and complexity of the surrogate models $z = f(x)$**
- **Seek models that are linear combinations of sets of basis functions**

1. **Simple basis functions**

Category		$X_j(x)$
I.	Polynomial	$(x_d)^{\alpha}$
II.	Multinomial	$\prod_{d \in \mathcal{D}' \subseteq \mathcal{D}} (x_d)^{\alpha_d}$
III.	Exponential and logarithmic	$\exp\left(\frac{x_d}{\gamma}\right)^{\alpha}, \log\left(\frac{x_d}{\gamma}\right)^{\alpha}$

2. **User-specified basis functions**
3. **Radial basis functions**

OVERFITTING AND TRUE ERROR

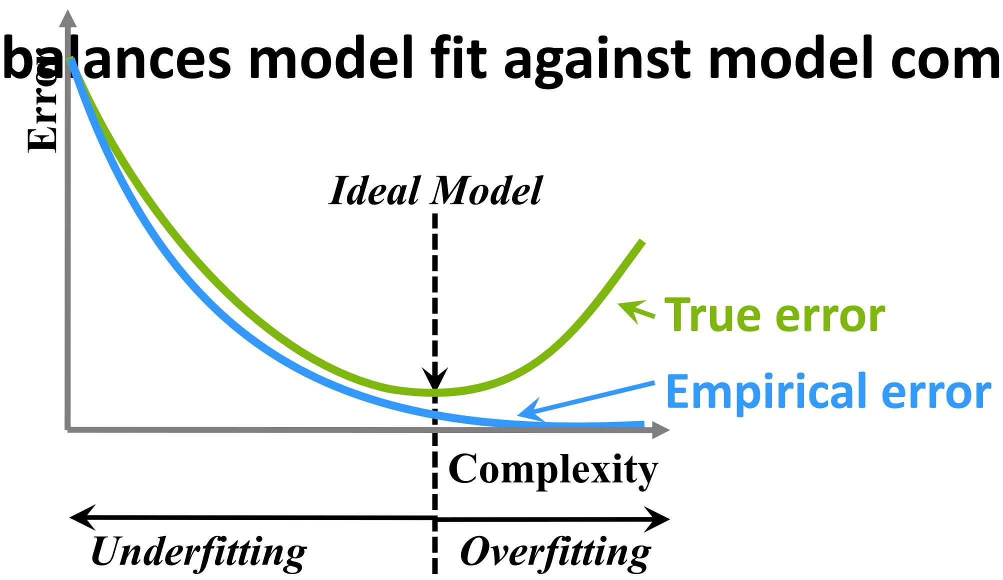
- **Step 1:** **Define a large set of potential building blocks**

$$\hat{z}(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 e^{x_1} + \beta_5 e^{x_2} + \ldots$$

- **Step 2:** **Model reduction**

$$\hat{z}(x) = 2 + x_2 + 5\, e^{x_1}$$

Select subset that balances model fit against model complexity



MODEL SELECTION CRITERIA

Balance fit (sum of square errors) with model complexity (number of terms in the model; denoted by ***p***)

Corrected Akaike information criterion

$$AIC_c = N \log\left(\frac{1}{N}\sum_{i=1}^{N}(z_i - X_i\beta)^2\right) + 2\boldsymbol{p} + \frac{2\boldsymbol{p}(\boldsymbol{p}+1)}{N-\boldsymbol{p}-1}$$

Mallows' Cp

$$C_p = \frac{\sum_{i=1}^{N}(z_i - X_i\beta)^2}{\widehat{\sigma^2}} + 2\boldsymbol{p} - N$$

Hannan-Quinn information criterion

$$HQC = N \log\left(\frac{1}{N}\sum_{i=1}^{N}(z_i - X_i\beta)^2\right) + 2\boldsymbol{p}\log(\log(N))$$

Bayes information criterion

$$BIC = \frac{\sum_{i=1}^{N}(z_i - X_i\beta)^2}{\widehat{\sigma^2}} + \boldsymbol{p}\log(N)$$

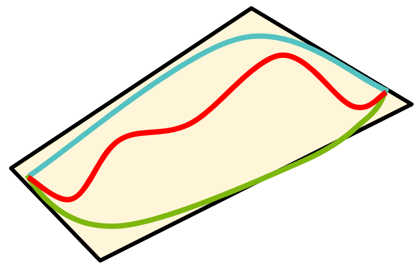
Mean squared error

$$MSE = \frac{\sum_{i=1}^{N}(z_i - X_i\beta)^2}{N-\boldsymbol{p}-1}$$

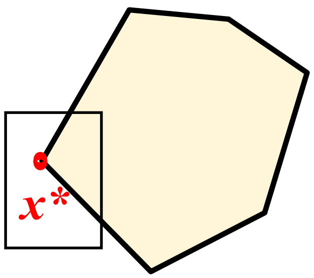
Mixed-integer nonlinear programming problems

BRANCH-AND-REDUCE

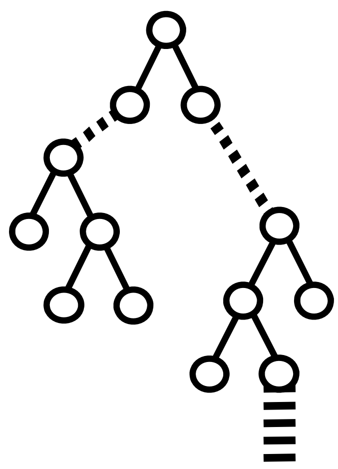
Convexification



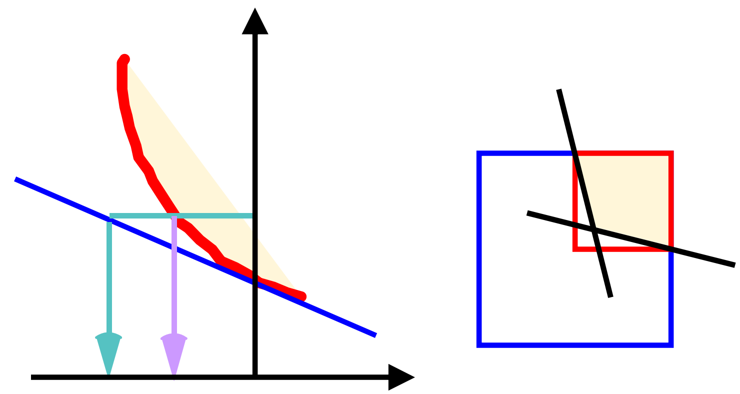
Finiteness



Branch-and-bound

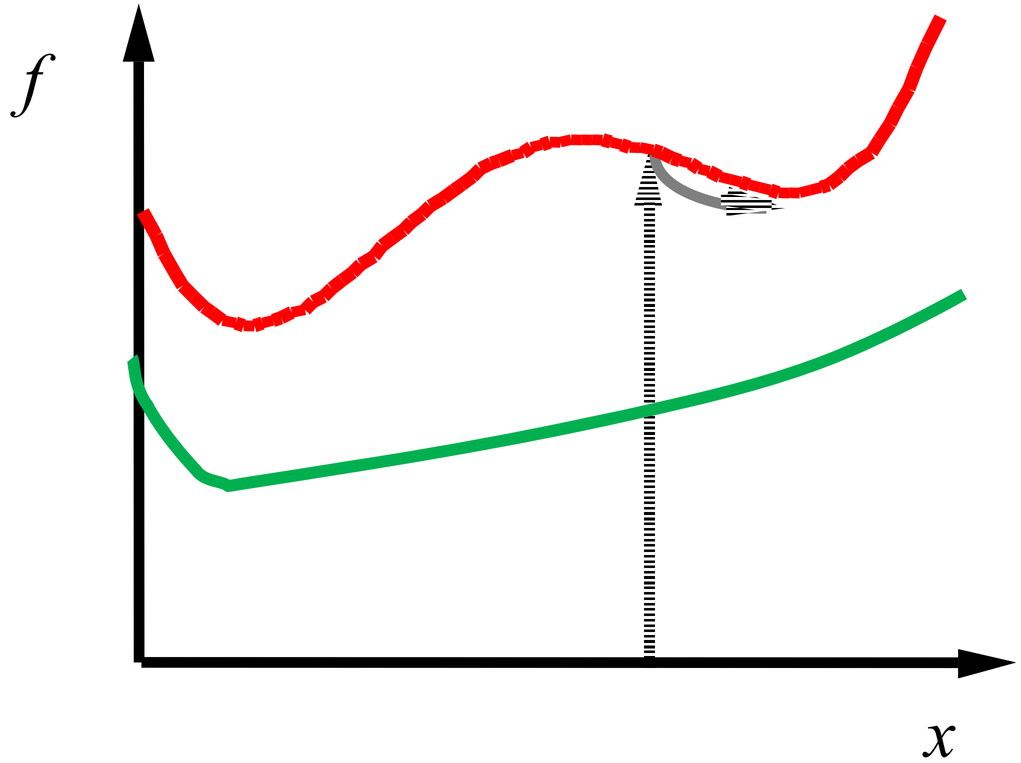


Range reduction



Ryoo and S., 1996; Tawarmalani and S., 2004; Khajavirad and S., 2018; Kilinc and S., 2018

CONVEXIFICATION



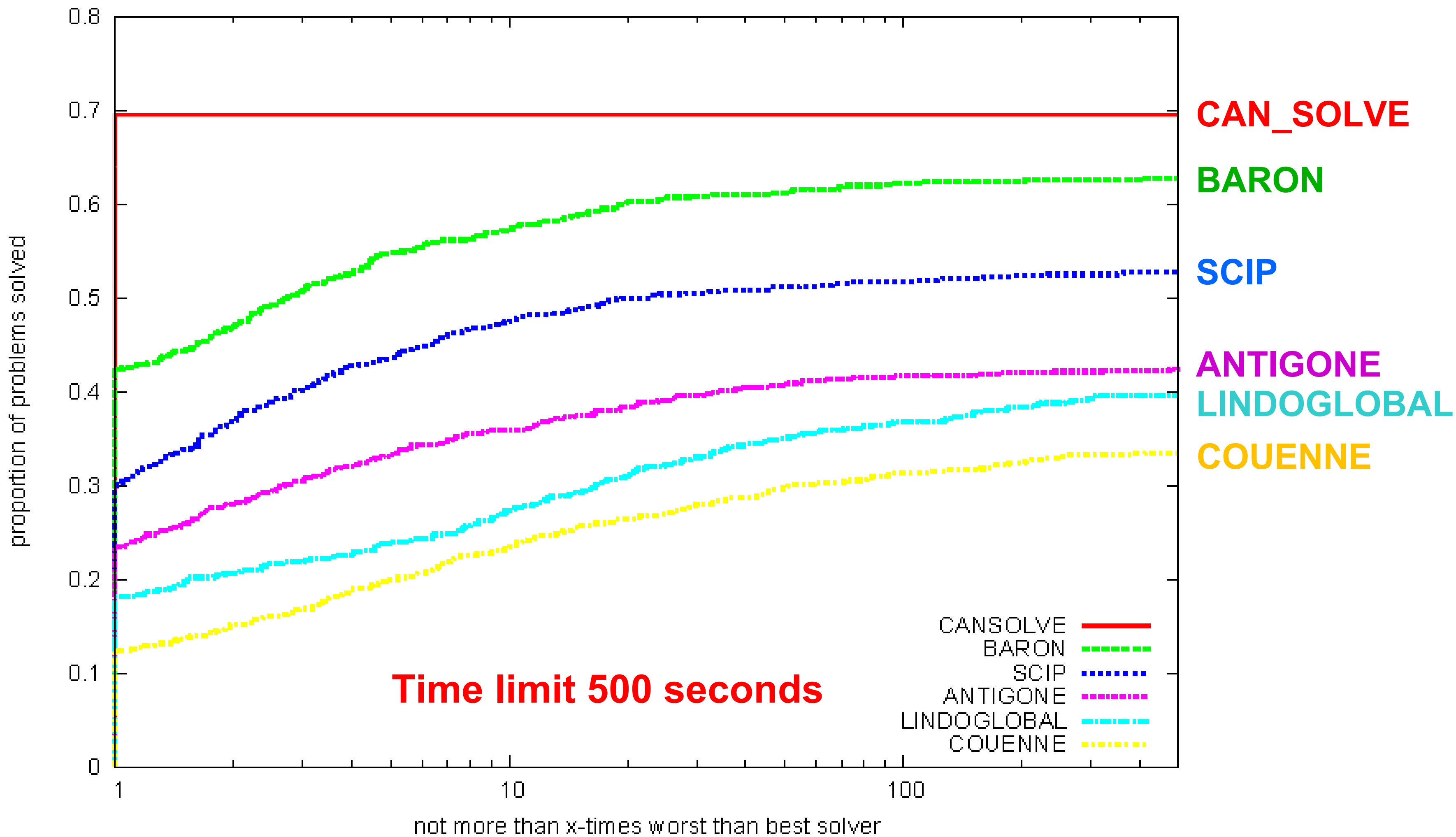
Classical optimization algorithms provide a local minimum “closest” to the starting point used

CONVEX ENVELOPES

Function	Domain		
$\sqrt{y}/x^2$	$x \in [-2, -1]$	$y \in [1, 4]$	
$y/(x_1 x_2)$	$x_1 \in [0.1, 1]$	$x_2 \in [1.5, 2]$	$y \in [0.5, 2]$
$y \exp(-x)$	$x \in [-1, 1]$	$y \in [1, 3]$	
$\log_{10} y/x^2$	$x \in [0.1, 2]$	$y \in [0.1, 10^2]$	
$y \exp(x_1 - x_2)$	$x_1 \in [0, 1]$	$x_2 \in [0, 1]$	$y \in [-1, 1]$
$x^2 \log_{10} y$	$x \in [-1, 2]$	$y \in [0.1, 10]$	
$y_1 y_2 / x$	$x \in [0.1, 1]$	$y_1 \in [0.1, 1]$	$y_2 \in [0.5, 1.5]$
$x^2 \sqrt{y_1 + y_2}$	$x \in [0.1, 0.5]$	$y_1 \in [0, 1]$	$y_2 \in [0.5, 1.5]$
$(2y_1 - y_2) \exp(-x)$	$x \in [-0.5, 1.0]$	$y_1 \in [0.6, 1.5]$	$y_2 \in [0.1, 1.0]$
$(y_1 + y_2)/x$	$x \in [1, 5]$	$y_1 \in [-2, 1]$	$y_2 \in [1, 3]$
$y_1 y_2 / x$	$x \in [0.1, 1]$	$y_1 \in [-1, 1]$	$y_2 \in [0.1, 1]$
$(\sqrt{y_1} - y_2) \exp(-x)$	$x \in [0, 1]$	$y_1 \in [0, 1]$	$y_2 \in [0.1, 2]$
$(y_1 y_2 - 2)/ \log x$	$x \in [10, 100]$	$y_1 \in [0, 1]$	$y_2 \in [1, 2]$

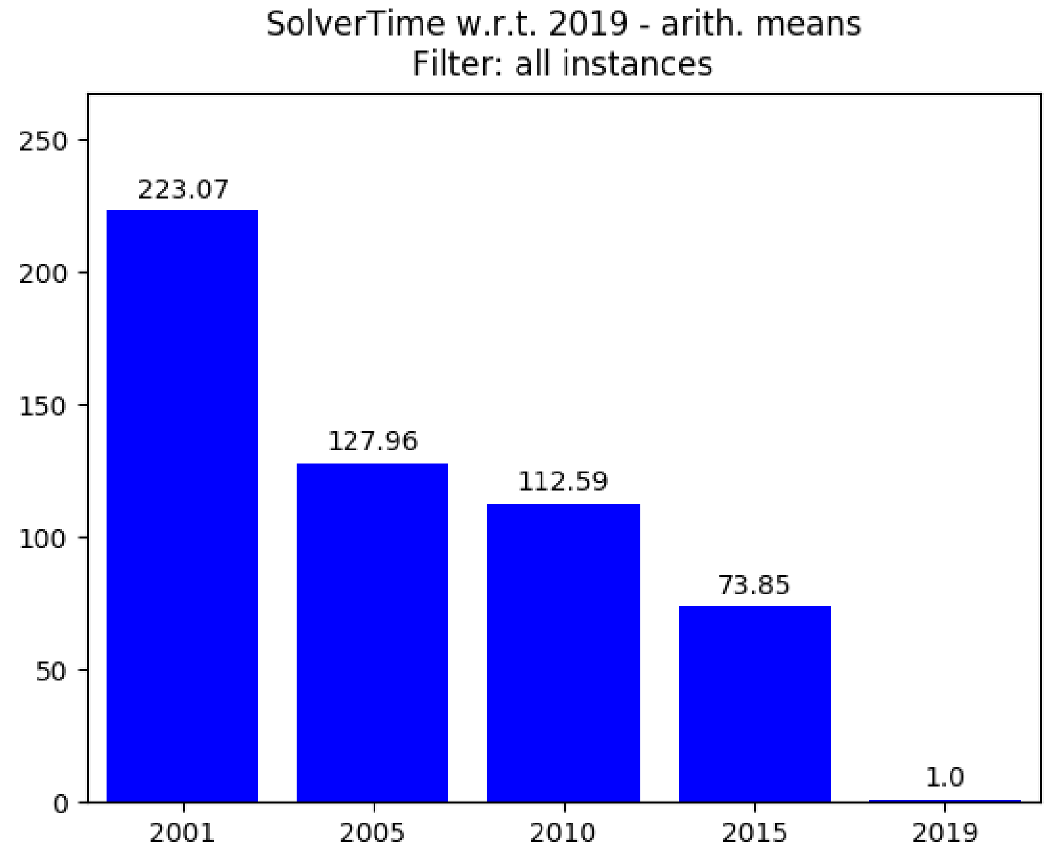
Khajavirad and Sahinidis, 2013, 2014

GLOBAL MINLP SOLVERS ON MINLPLIB2



Constraints: 1893 (1—164,321), Variables: 1027 (3—107,223), Discrete: 137 (1—31,824)

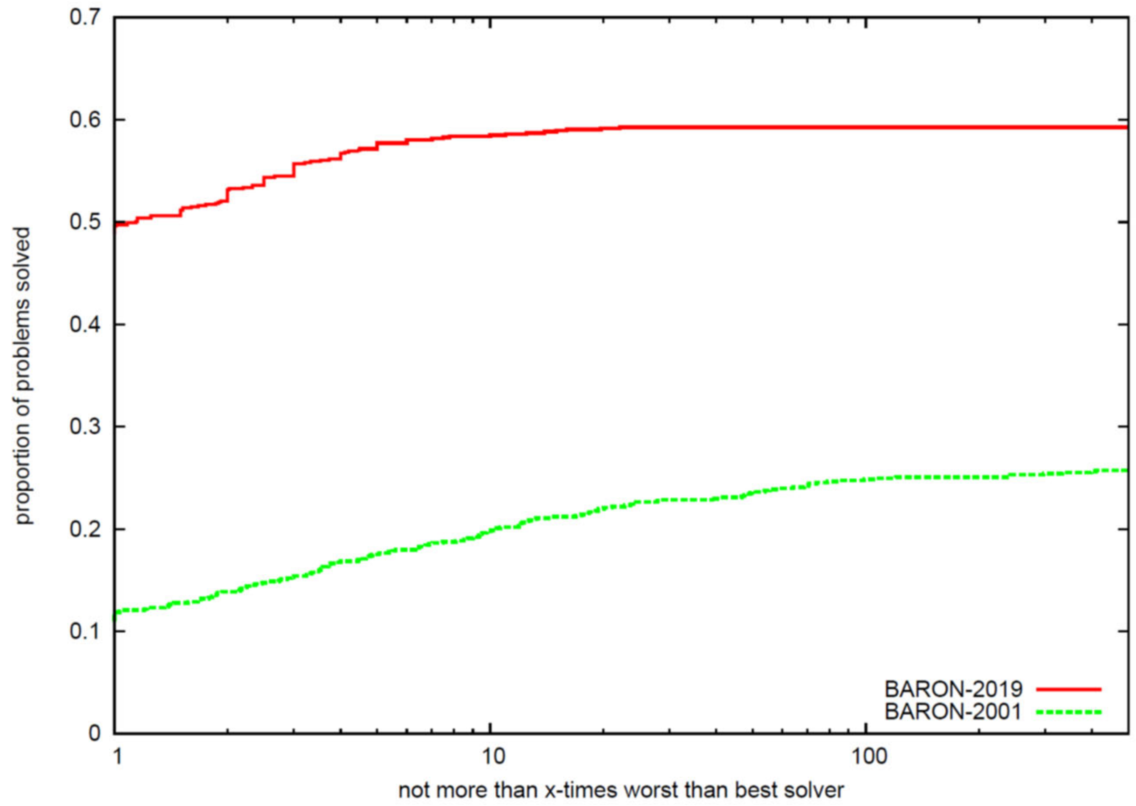
223 TIMES FASTER



- **Results on MINLPLIB2**
- **Constraints: 1893 (1—164,321), Variables: 1027 (3—107,223), Discrete: 137 (1—31,824)**

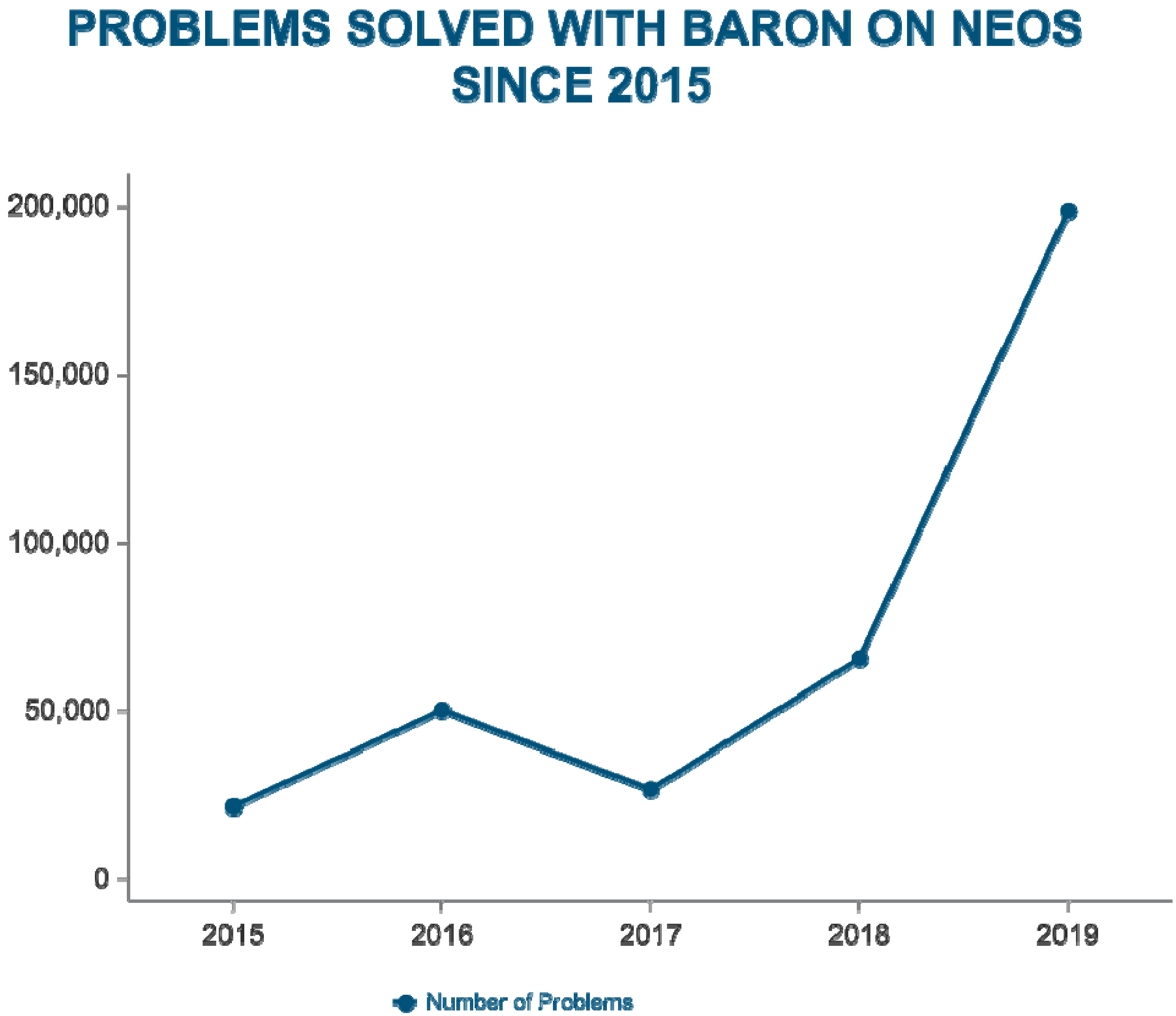
Time limit 500 seconds

SOLVES MORE 2.5X MORE PROBLEMS



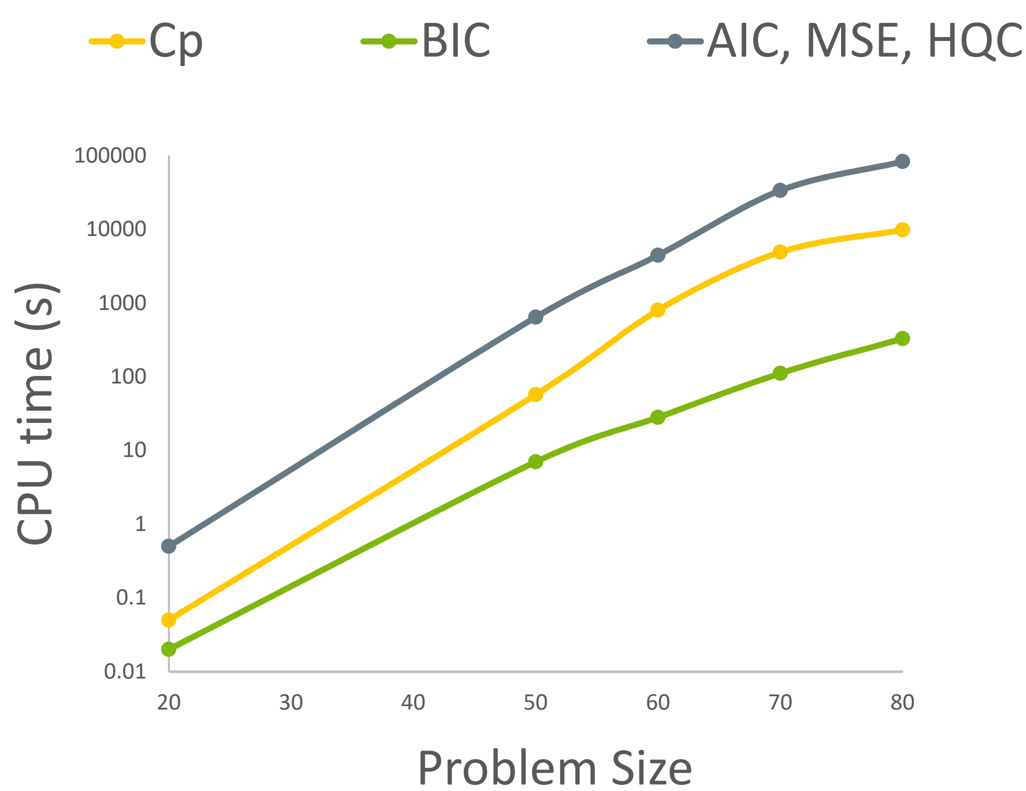
Comparisons based on solver ability to prove global optimality

FREE THROUGH THE NEOS SERVER



CPU TIME COMPARISON OF METRICS

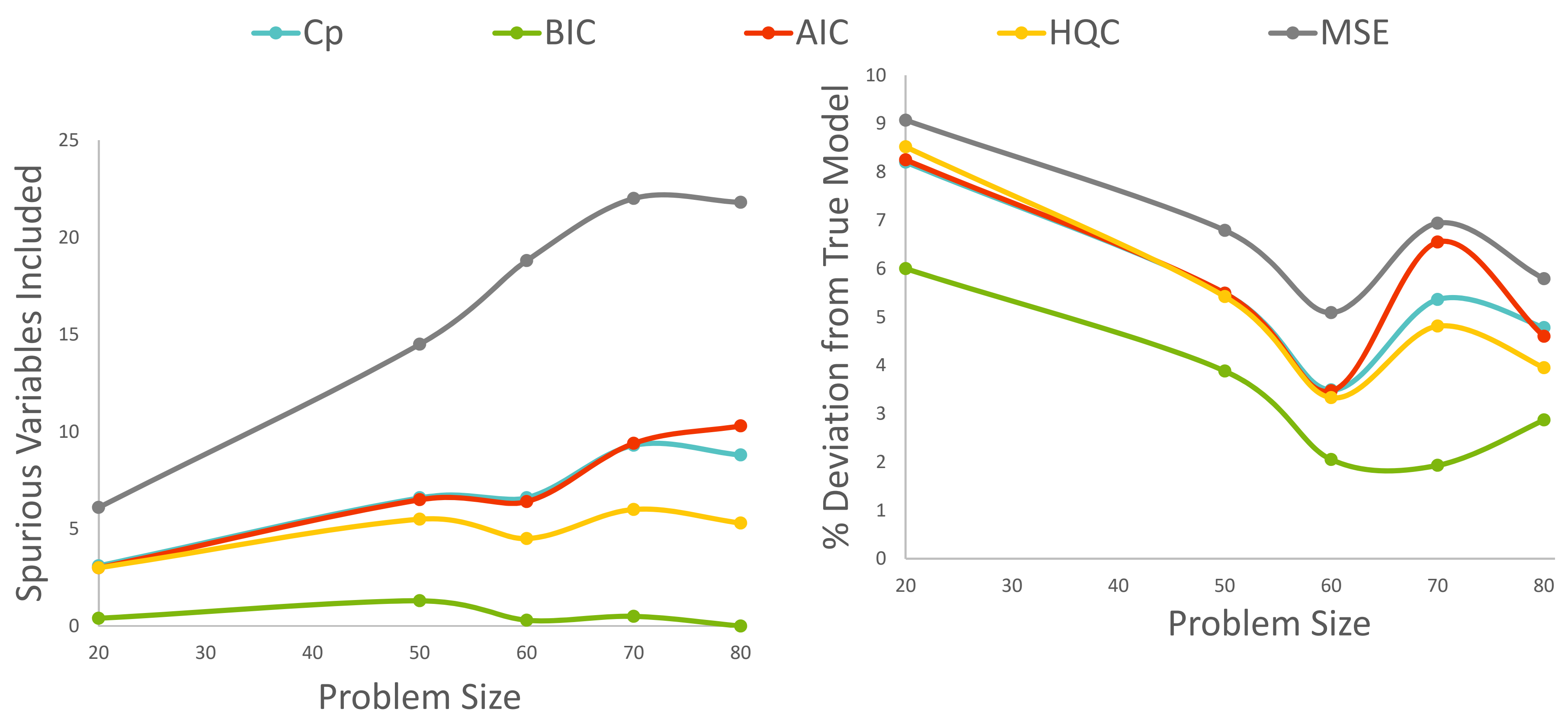
- Eight benchmarks from the UC-Irvine data set
- Seventy noisy data sets were generated with multicolinearity and increasing problem size (number of bases)



BIC solves more than two orders of magnitude faster than AIC, MSE and HQC

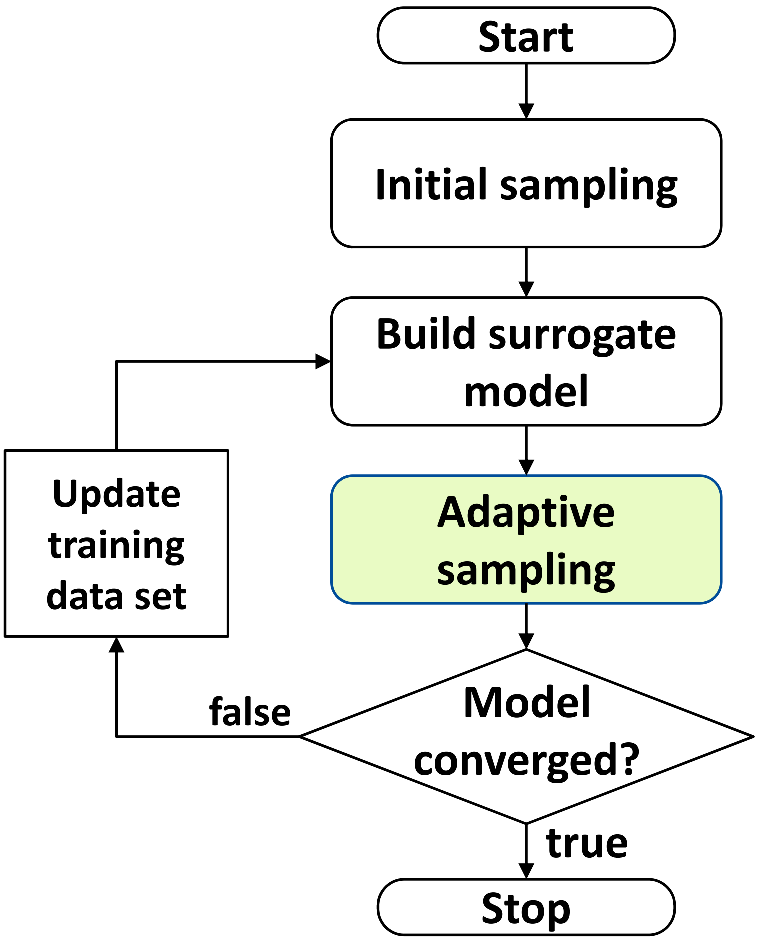
MODEL QUALITY COMPARISON

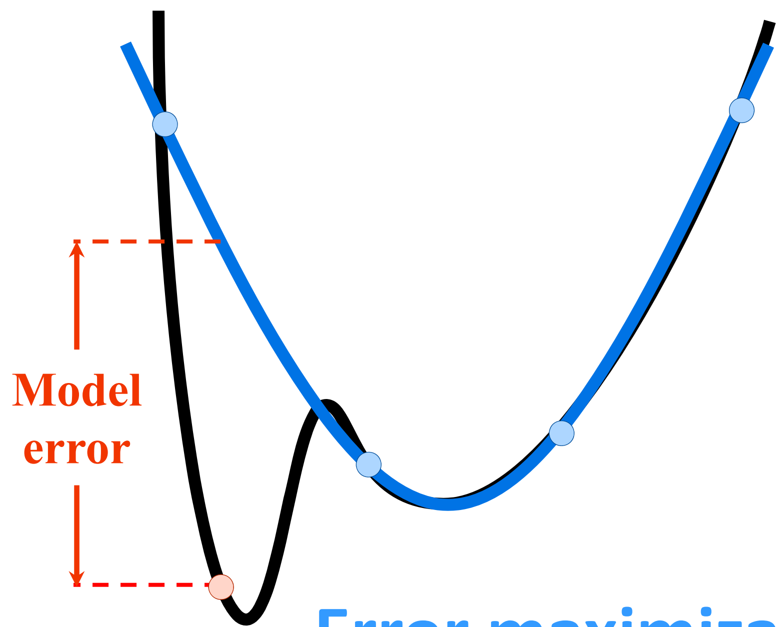
- **BIC leads to smaller, more accurate models**
  - Larger penalty for model complexity



ALAMO

Automated Learning of Algebraic MOdels





Error maximization sampling

ERROR MAXIMIZATION SAMPLING

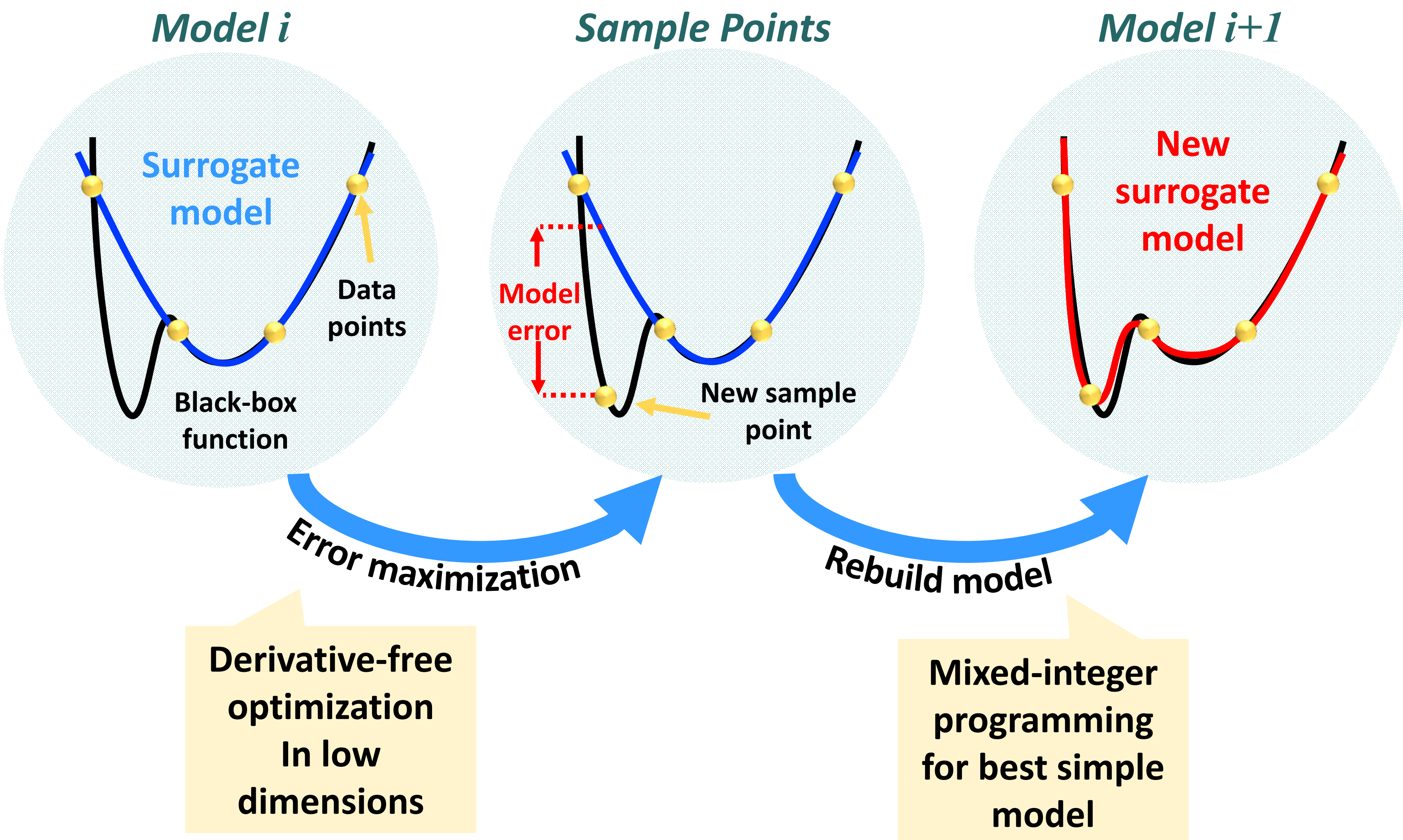
- **Search the problem space for areas of model inconsistency or model mismatch**
- **Find points that maximize the model error with respect to the independent variables**

Surrogate model

$$\max_{x} \left(z(x) - \hat{z}(x)\right)^2$$

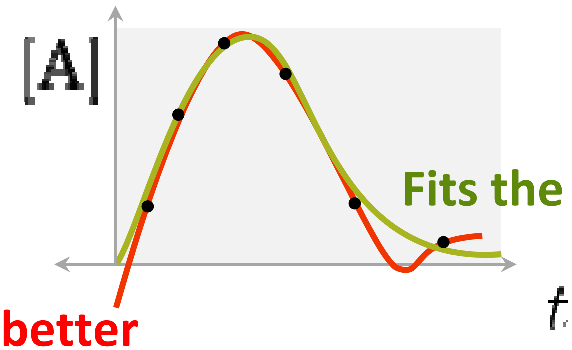
  - Optimized using derivative-free solver SNOBFIT (Huyer and Neumaier, 2008)
  - SNOBFIT has advantages over 20+ other derivative-free solvers (Rios and Sahinidis, 2013)

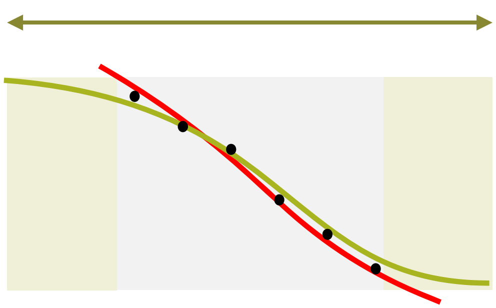
ALAMO METHODOLOGY



CONSTRAINED REGRESSION

$$0 < [A]_t < [A]^{max}$$





Safe extrapolation

CONSTRAINED REGRESSION

Standard regression

$$\min_{\beta_1,\beta_2} \sum_{i=1}^{6} \left(z_i - \underbrace{\hat{z}(x_i;\beta_1,\beta_2)}_{\text{Surrogate model}}\right)^2$$

Constrained regression

easy

$$\min_{\beta_1,\beta_2} \sum_{i=1}^{6} \left(z_i - \hat{z}(x_i;\beta_1,\beta_2)\right)^2$$
$$\text{s.t.} \quad \beta_1 \geq \beta_2$$

tough

$$\min_{\beta_1,\beta_2} \sum_{i=1}^{6} \left(z_i - \hat{z}(x_i;\beta_1,\beta_2)\right)^2$$
$$\text{s.t.} \quad \hat{z}(x_i;\beta_1,\beta_2) \geq 0 \quad \forall x$$

- **Challenging due to the semi-infinite nature of the regression constraints**
- **Use intuitive restrictions among predictor and response variables to infer nonintuitive relationships between regression parameters**

IMPLIED PARAMETER RESTRICTIONS

Find a model $\hat{z}$ such that $\hat{z}(x) \geq 0$ with a fixed model form:

$$\hat{z}(x) = \beta_1 x + \beta_2 x^3$$

Step 1: Formulate constraint in z- and x-space

$$\min_{\beta_1, \beta_2} \sum_{i=1}^{6} \left(z_i - [\beta_1 x + \beta_2 x^3]\right)^2$$

$$\text{s.t.} \quad \beta_1 x + \beta_2 x^3 \geq 0 \quad x \in [0, 1]$$

1 parametric constraint

4 β-constraints

Step 2: Identify a sufficient set of β-space constraints

$$\min_{\beta_1, \beta_2} \sum_{i=1}^{6} \left(z_i - [\beta_1 x + \beta_2 x^3]\right)^2$$

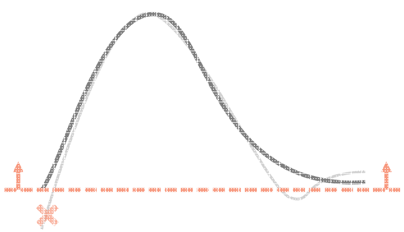
$$\text{s.t.} \quad \begin{cases} 0.240\,\beta_1 + 0.0138\,\beta_2 \geq 0 \\ 0.281\,\beta_1 + 0.0223\,\beta_2 \geq 0 \\ 0.120\,\beta_1 + 0.00173\,\beta_2 \geq 0 \\ 0.138\,\beta_1 + 0.00263\,\beta_2 \geq 0 \end{cases}$$

Global optimization problems solved with BARON

TYPES OF RESTRICTIONS

Response bounds

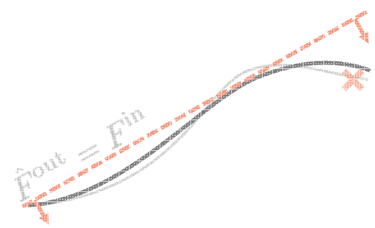
$$[\hat{A}]_t \geq 0$$



pressure, temperature, compositions

Individual responses

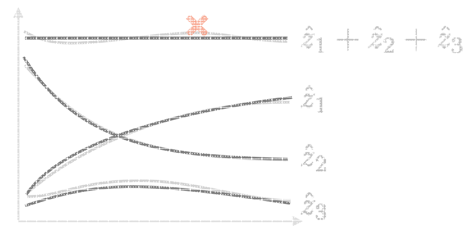
$$\hat{F}^{\text{out}}(x) \leq F^{\text{in}}$$



mass and energy balances, physical limitations

Multiple responses

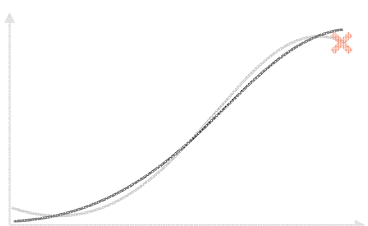
$$\hat{z}_1 + \hat{z}_2 + \hat{z}_3 = 1$$



mass balances, sum-to-one, state variables

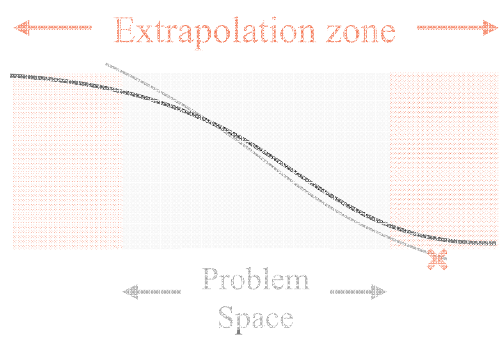
Response derivatives

$$\frac{dT}{dx} \geq 0$$



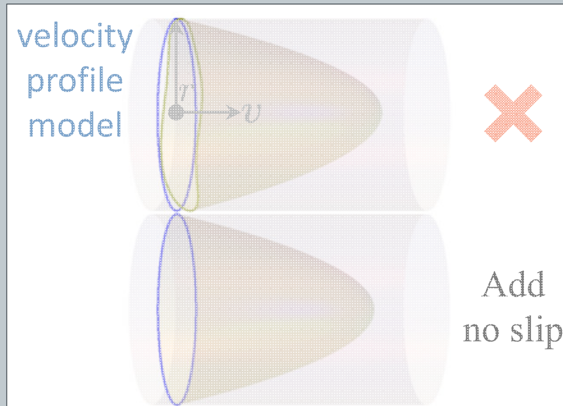
monotonicity, numerical properties, convexity

Alternative domains



safe extrapolation, boundary conditions

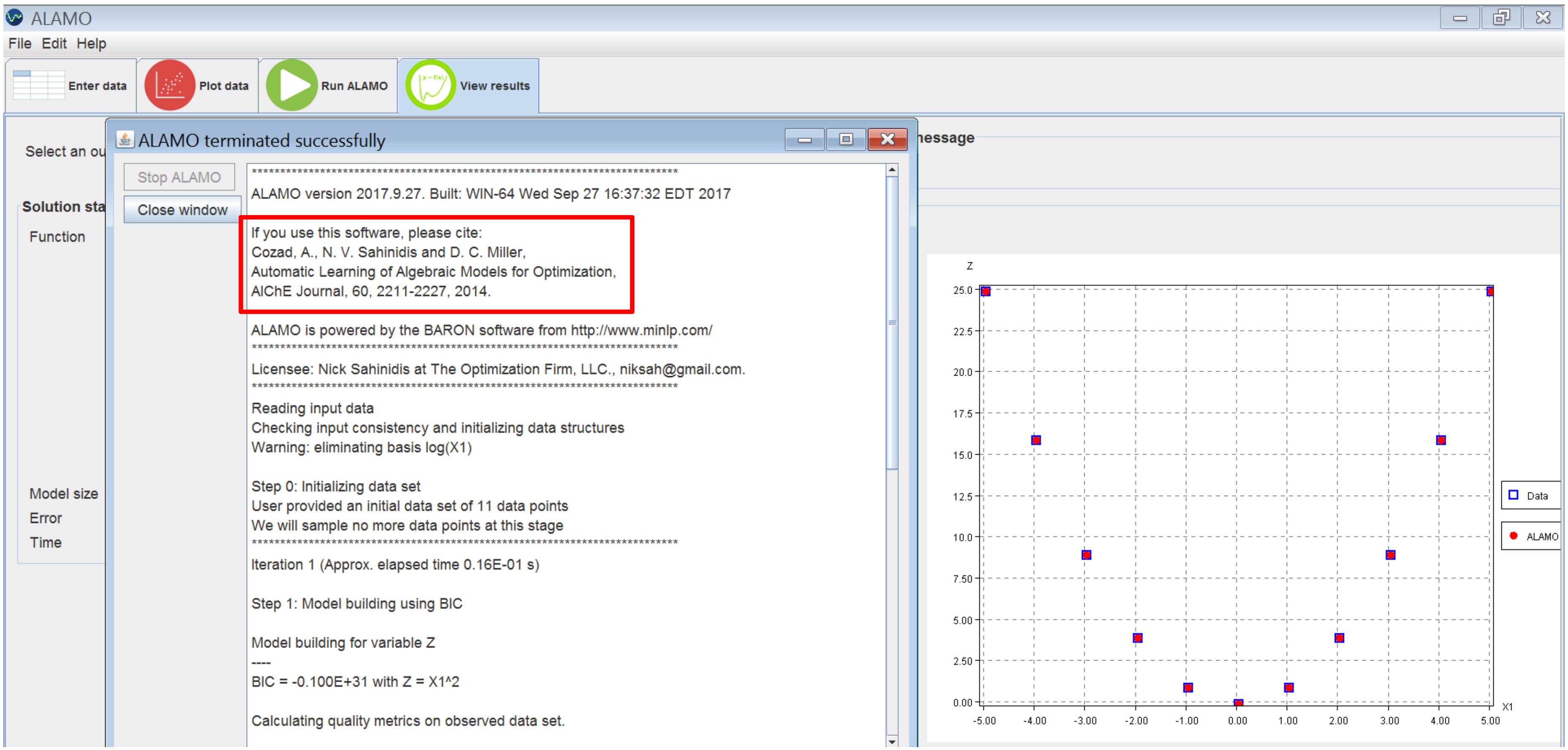
Boundary conditions



$$\hat{v}(R, \theta) = 0 \qquad \forall\theta$$

SOFTWARE AVAILABILITY

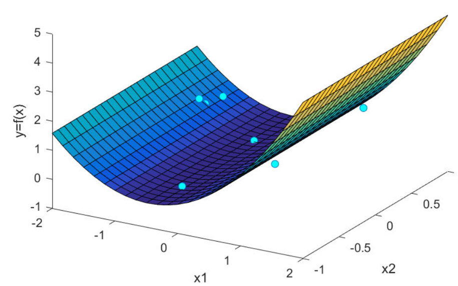
- Implemented by PSE in its gPROMS simulator
- Free from http://minlp.com



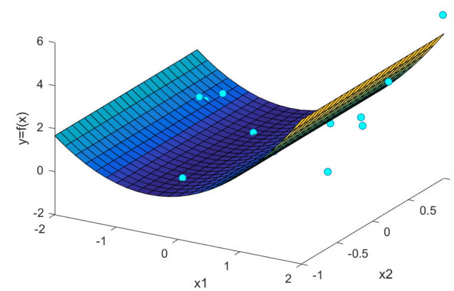
LEARNING THE SIX-HUMP CAMEL FUNCTION

$$f(x_1, x_2) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + x_2^2(4x_2^2 - 4) + \epsilon$$

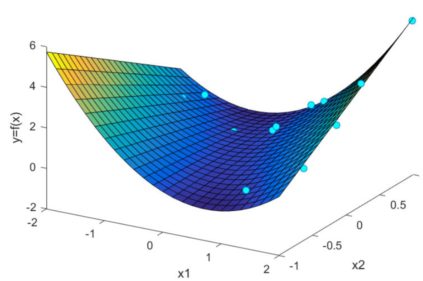
Iteration	N	R^2_{test}	$\lVert\beta\rVert_0$
$\mathbf{1}$	$\mathbf{10}$	$< \mathbf{0}$	$\mathbf{2}$
$\mathbf{2}$	$\mathbf{16}$	$< \mathbf{0}$	$\mathbf{2}$
$\mathbf{3}$	$\mathbf{19}$	$< \mathbf{0}$	$\mathbf{2}$
$\mathbf{4}$	$\mathbf{27}$	$\mathbf{0.98}$	$\mathbf{7}$



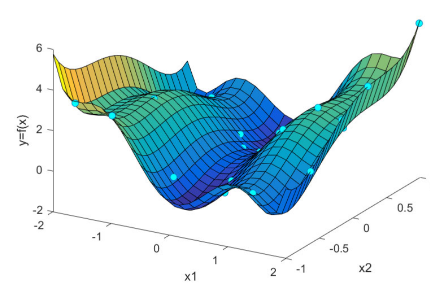
First iteration



Second iteration



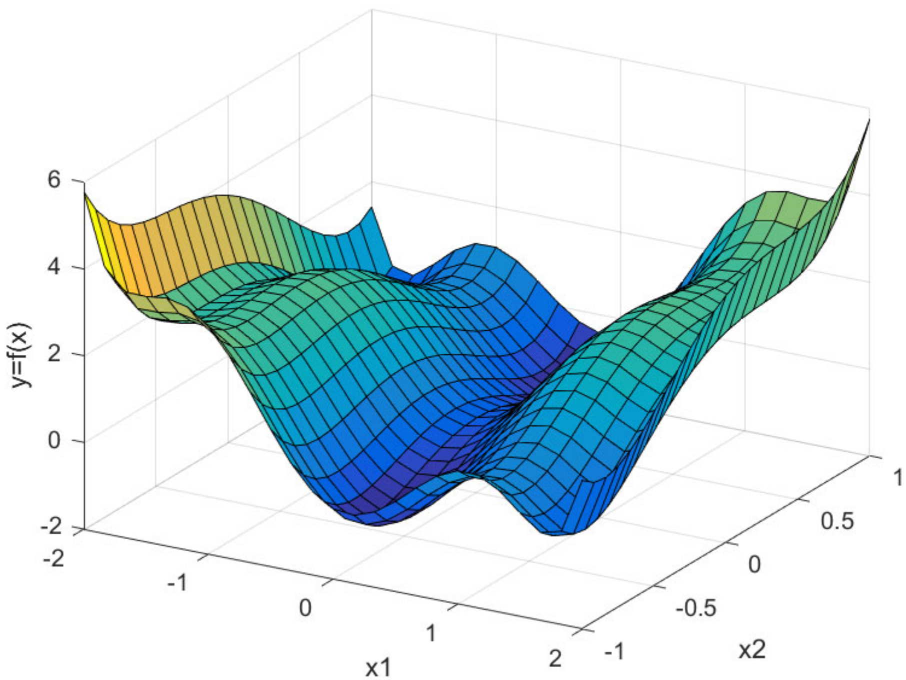
Third iteration



Final iteration

$$f = 4.56x_1^2 - 3.16x_2^2 - 2.41x_1^4 + 3.07x_2^4 + 0.38x_1^6 + 1.09x_1x_2 - 0.28$$

OPTIMIZATION WITH THE SURROGATE



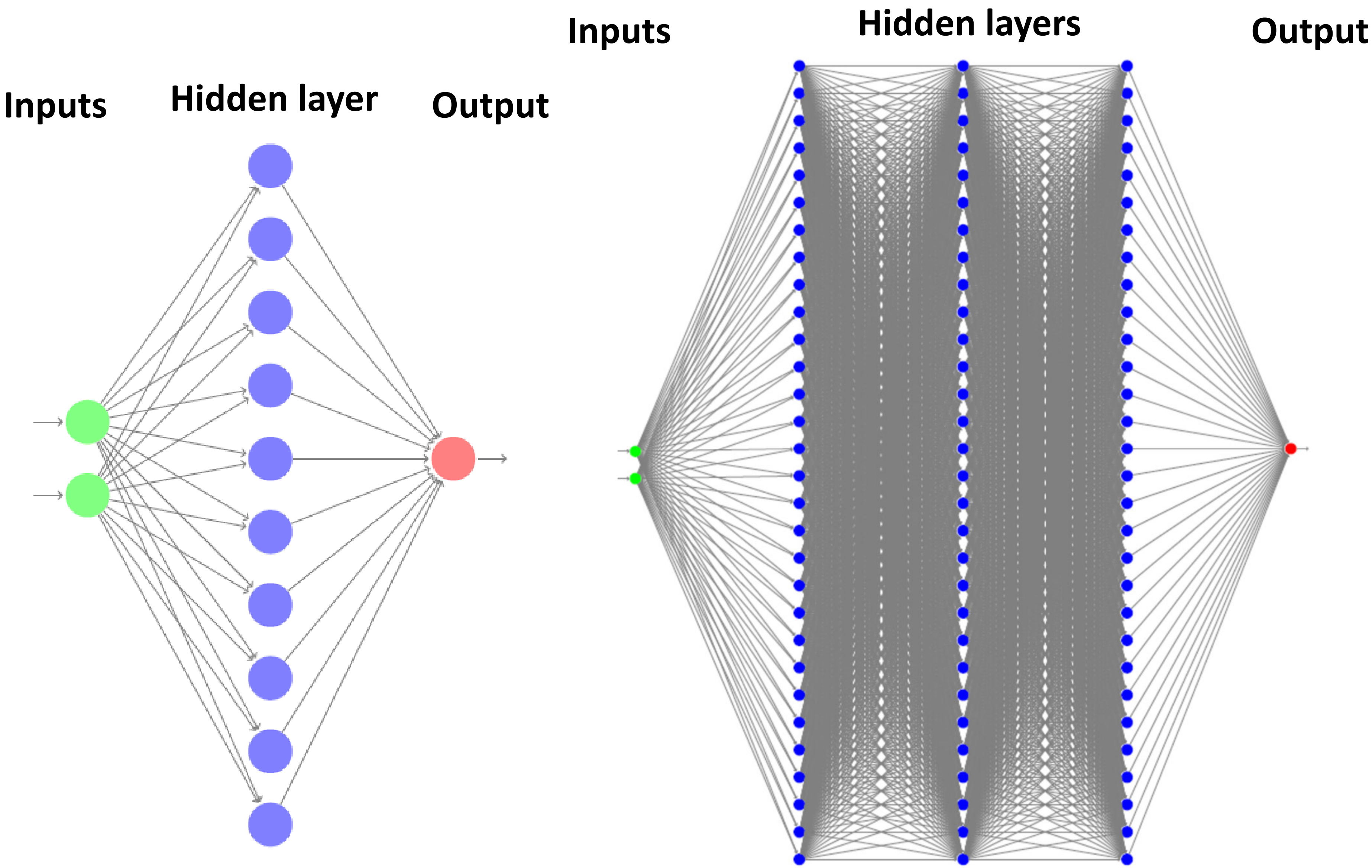
True minimum

$$\boldsymbol{f(0.0898, -0.7127) = -1.0316}$$

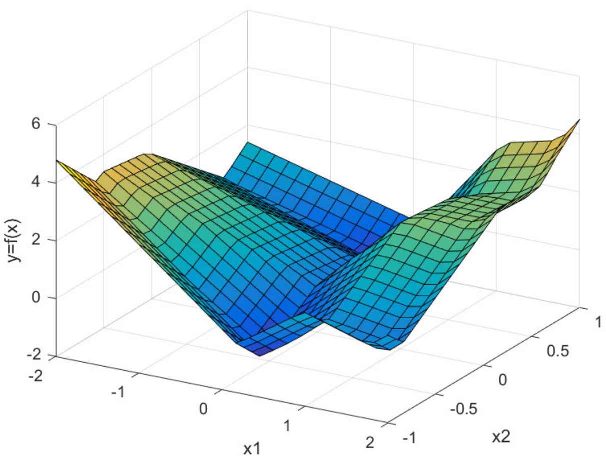
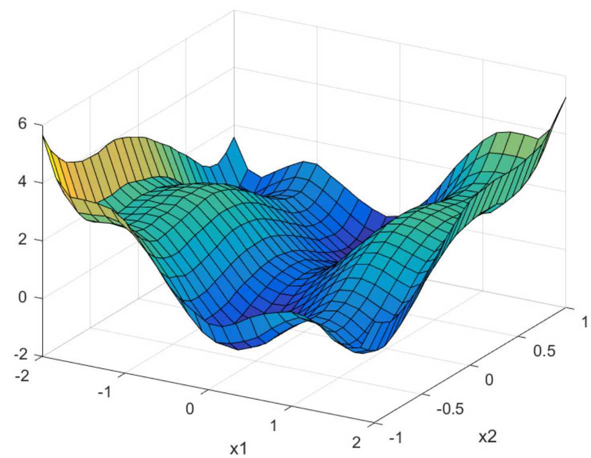
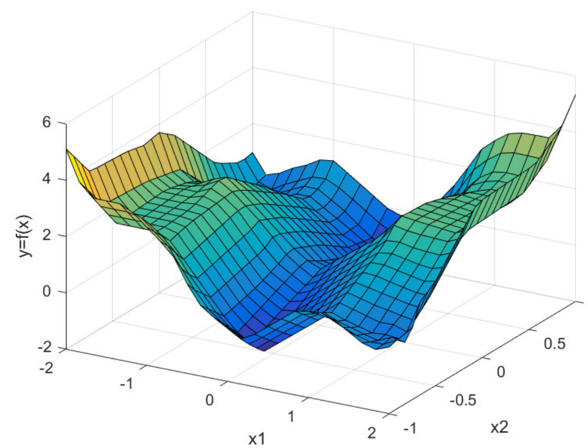
ALAMO surrogate minimum

$$\boldsymbol{f(0.0871, -0.7251) = -1.1248}$$

NEURAL NETWORK SURROGATES



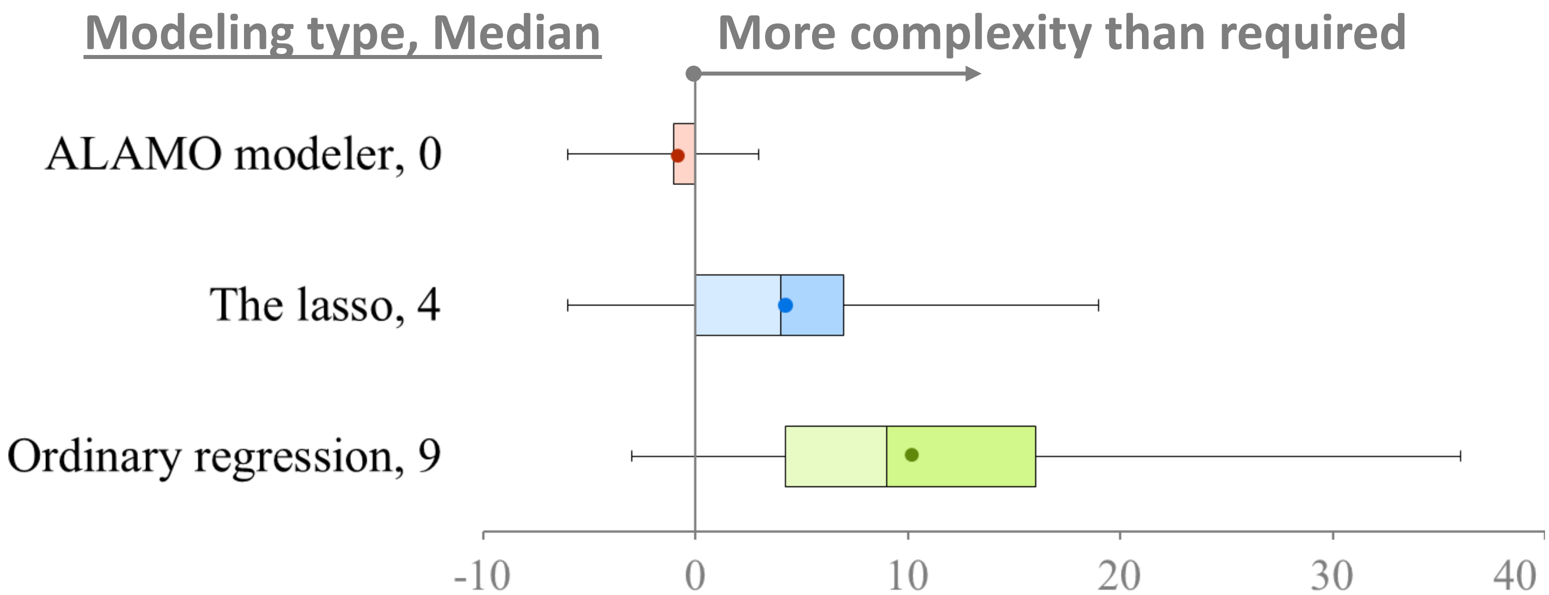
RELU NEURAL NETWORK SURROGATES

	1 hidden layer 10 nodes	1 hidden layer 200 nodes	3 hidden layer 30 nodes
			
f	-0.919	-1.22	-1.194
x_1	0.064	-0.134	-0.016
x_2	-0.596	0.683	0.639

True minima

$$f(0.0898, -0.7126) = -1.0316$$
$$f(-0.0898, 0.7126) = -1.0316$$

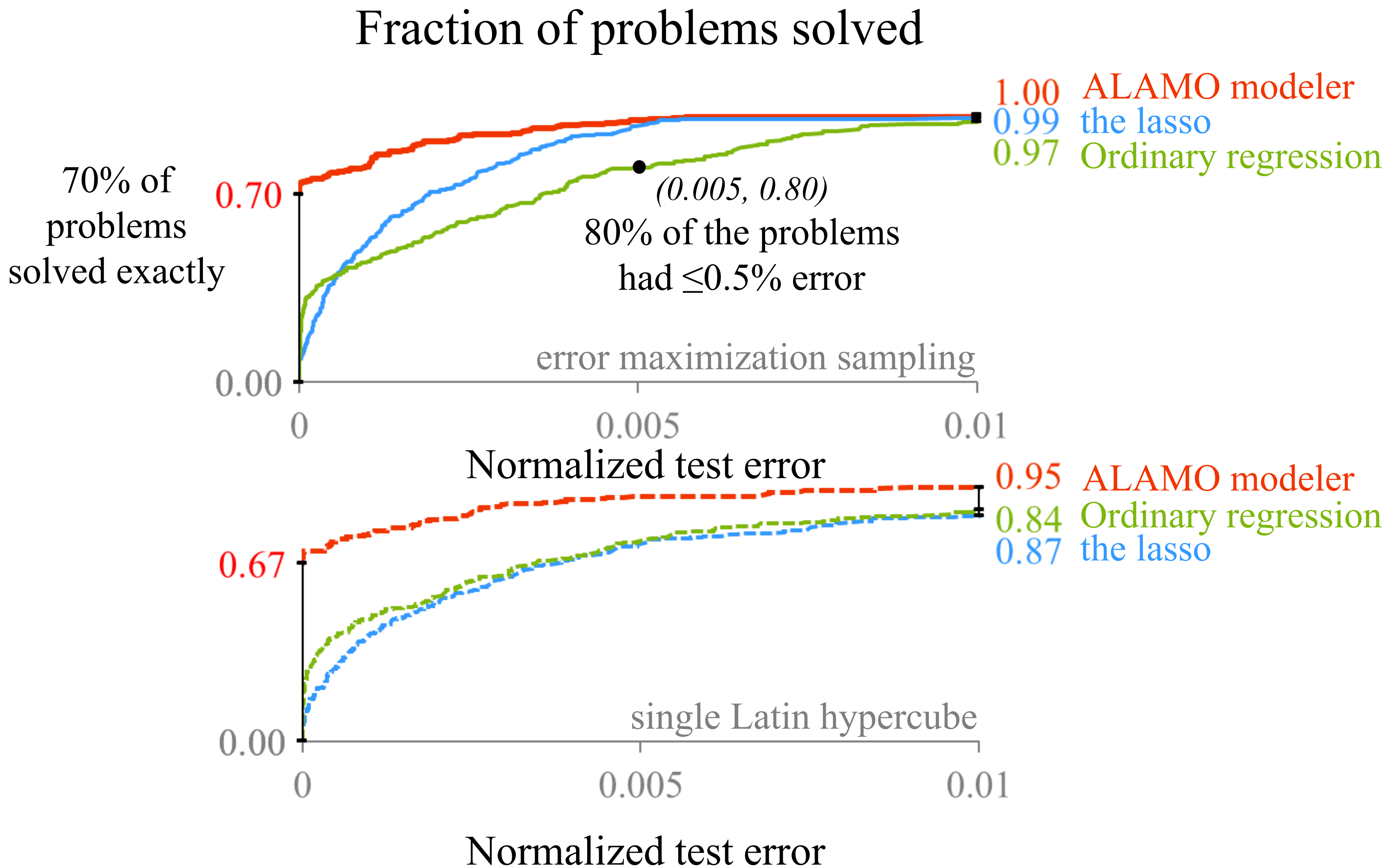
SIMPLE AND ACCURATE MODELS



[Number of terms in the surrogate model] – [Number of terms in the true function]

Results over a test set of 45 known functions treated as black boxes with bases that are available to all modeling methods

SAMPLING EFFICIENCY

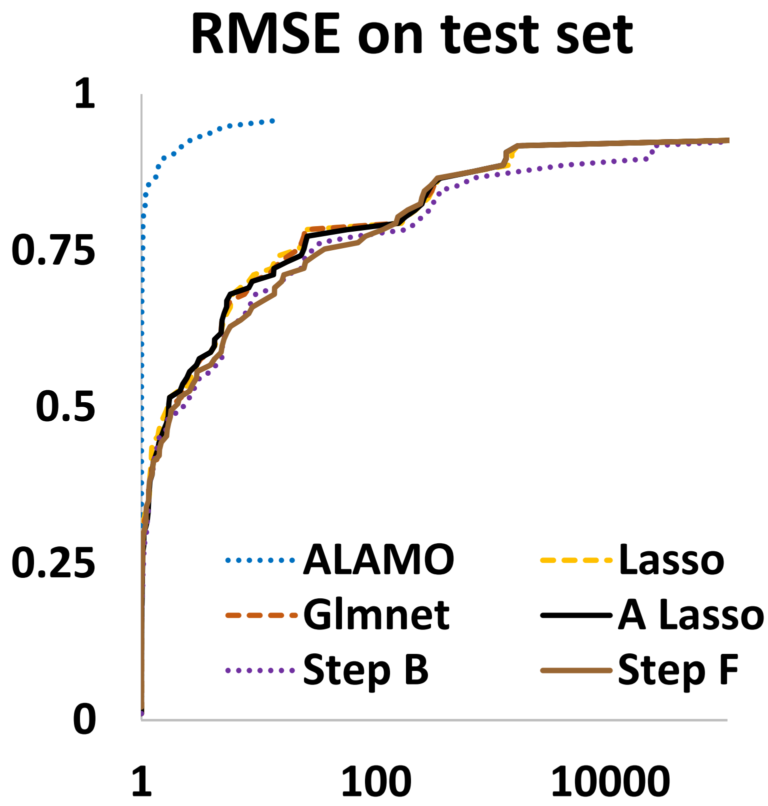


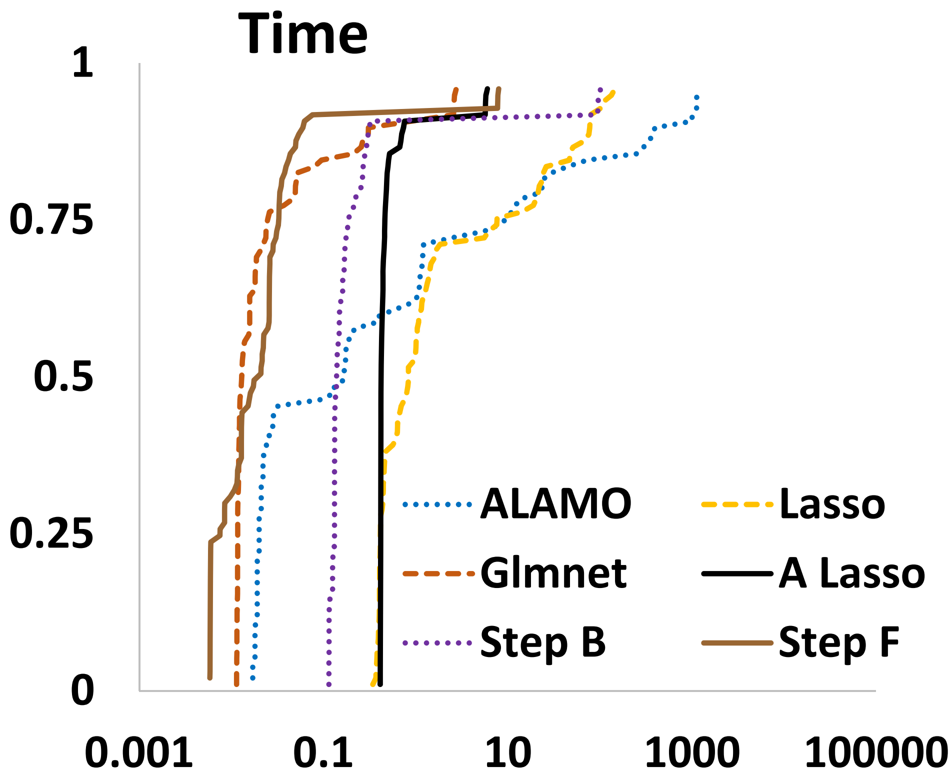
Results with 45 known functions with bases that are available to all modeling methods

COMPARISONS ON BENCHMARKS

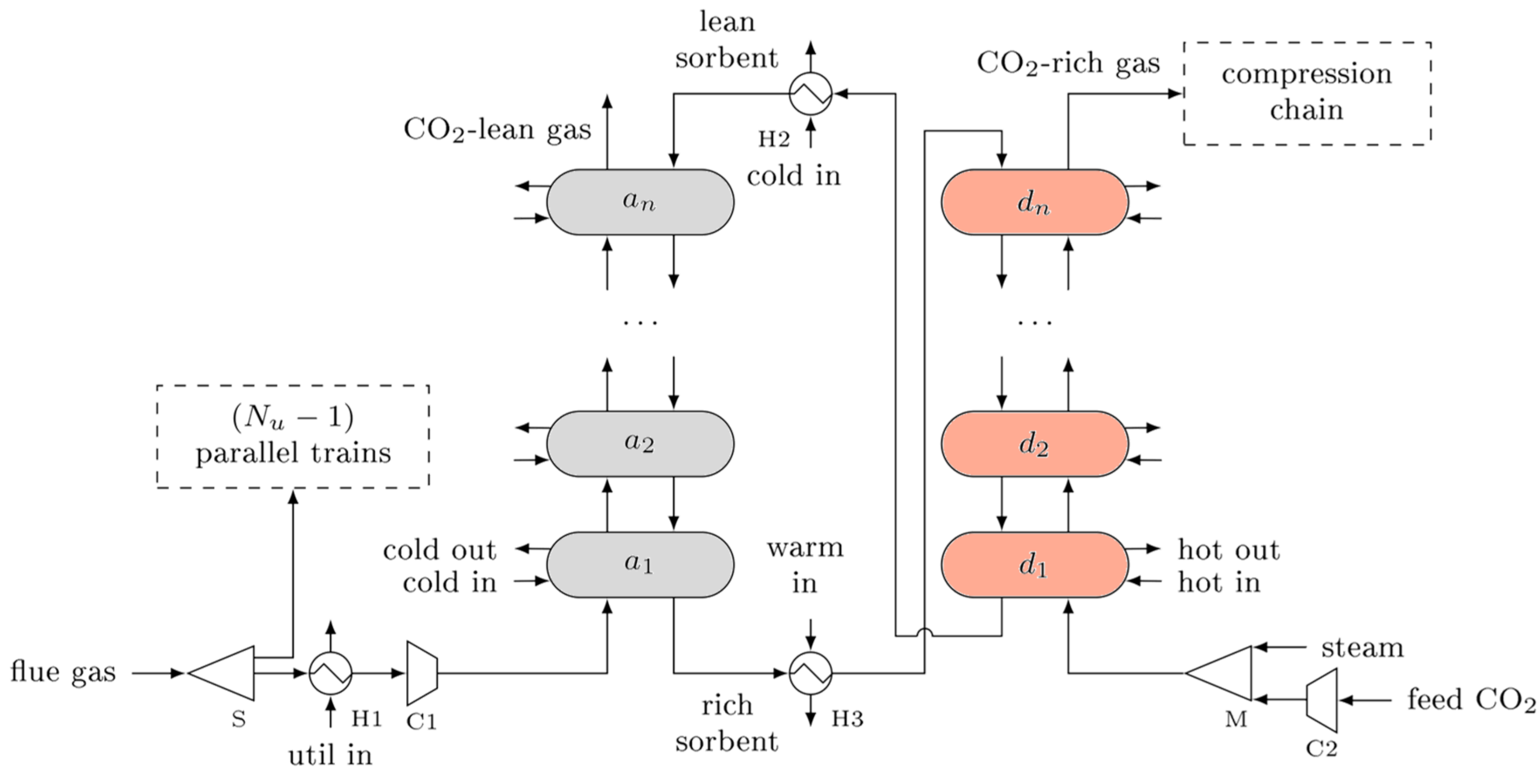
- **98 problems**
  - 30 from the UC-Irvine ML repository
  - 37 from the NIST standard regression database
  - 31 from the Virtual Library of Simulated Experiments
- **Number of inputs: 1—105 (average 11)**
- **Number of features: 6—735 (average 90)**
- **Number of measurements: 6—32561 (average 2144)**
- **Algorithms compared**
  - Lasso (Matlab)
  - Glmnet (lasso in R)
  - A lasso (adaptive lasso option in Glmnet)
  - Step F/B (Matlab)
  - ALAMO with BIC solved to optimality

TIME AND QUALITY COMPARISONS





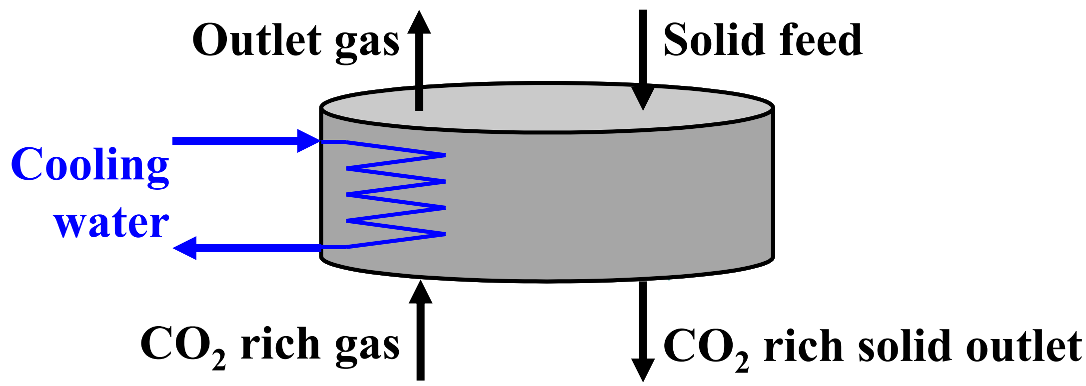
CARBON CAPTURE SYSTEM DESIGN



- **Discrete decisions:** How many units? Parallel trains? What technology used for each reactor?
- **Continuous decisions:** Unit geometries
- **Operating conditions:** Vessel temperature and pressure, flow rates, compositions

BUBBLING FLUIDIZED BED

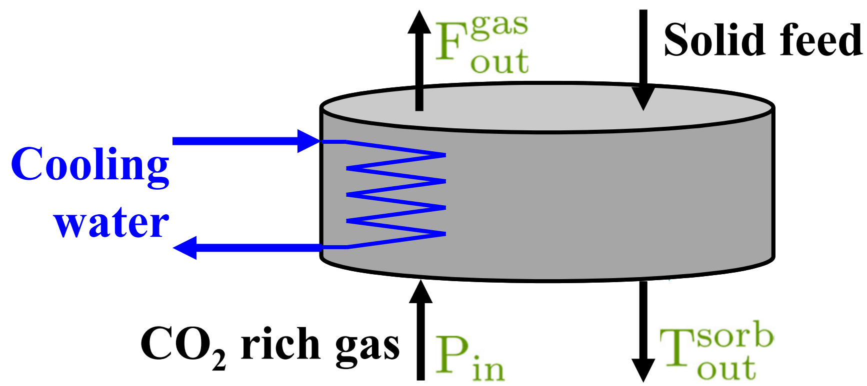
Bubbling fluidized bed adsorber diagram



- **Model inputs (16 total)**
  - Geometry (3)
  - Operating conditions (5)
  - Gas mole fractions (2)
  - Solid compositions (2)
  - Flow rates (4)
- **Model outputs (14 total)**
  - Geometry required (2)
  - Operating condition required (1)
  - Gas mole fractions (3)
  - Solid compositions (3)
  - Flow rates (2)
  - Outlet temperatures (3)

Model created by Andrew Lee at the National Energy Technology Laboratory

EXAMPLE MODELS - ADSORBER

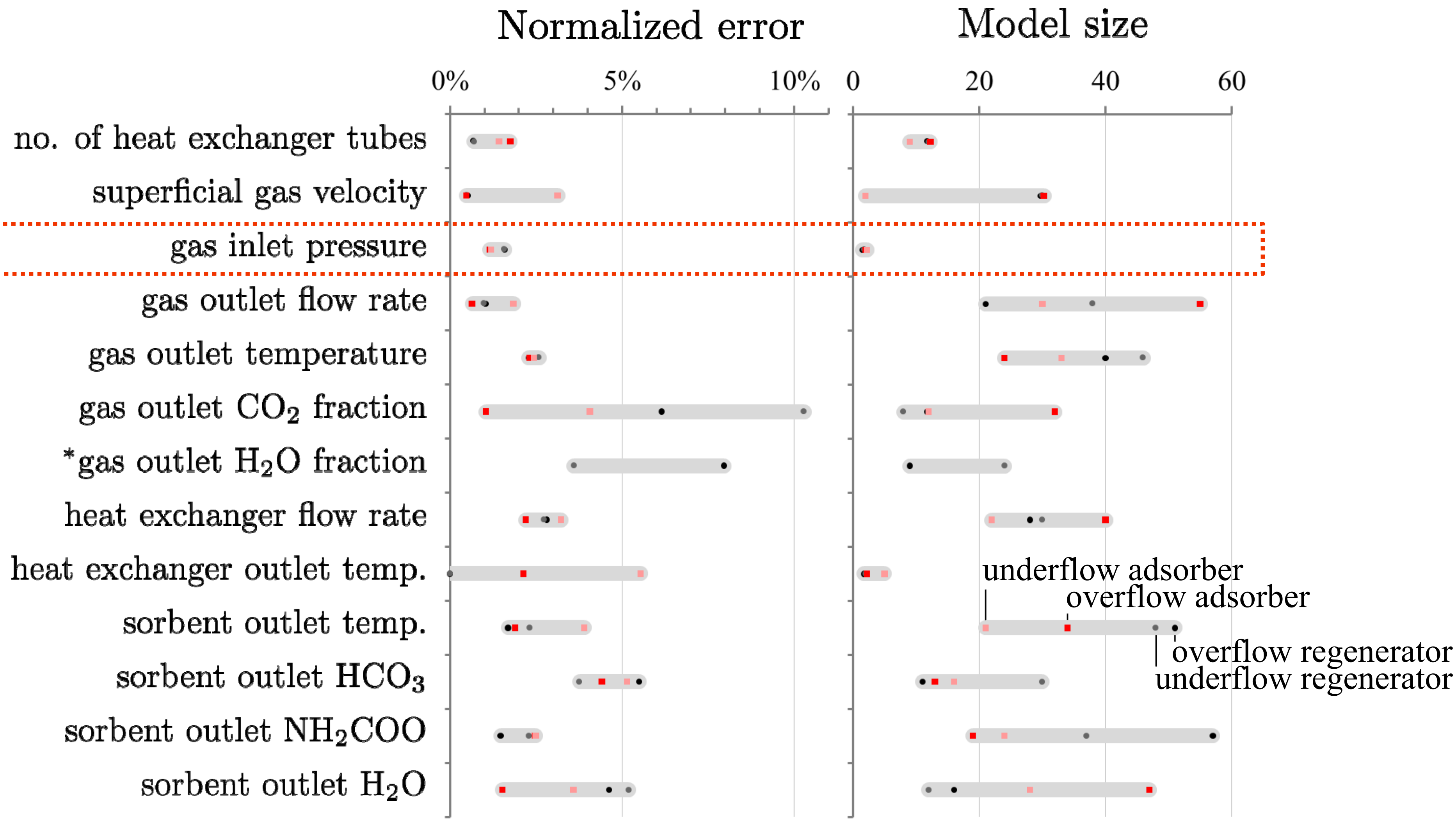


$$\mathrm{P_{in}} = 1.0\,\mathrm{P_{out}} + 0.0231\,L_b - 0.0187\,\ln(0.167\,L_b) - 0.00626\,\ln(0.667\,\mathrm{v_{gi}}) - \frac{51.1\,\mathrm{xHCO3^{ads}_{in}}}{\mathrm{F^{gas}_{in}}}$$

$$T^{sorb}_{out} = 1.0\,\mathrm{T^{gas}_{in}} - \frac{\left(1.77\cdot 10^{-10}\right)\,\mathrm{NX}^2}{\gamma^2} - \frac{3.46}{\mathrm{NX\,T^{gas}_{in}\,T^{sorb}_{in}}} + \frac{1.17\cdot 10^4}{\mathrm{F^{sorb}\,NX\,xH2O^{ads}_{in}}}$$

$$F^{gas}_{out} = 0.797\,\mathrm{F^{gas}_{in}} - \frac{9.75\,\mathrm{T^{sorb}_{in}}}{\gamma} - 0.77\,\mathrm{F^{gas}_{in}\,xCO2^{gas}_{in}} + 0.00465\,\mathrm{F^{gas}_{in}\,T^{sorb}_{in}} - 0.0181\,\mathrm{F^{gas}_{in}\,T^{sorb}_{in}\,xH2O^{gas}_{in}}$$

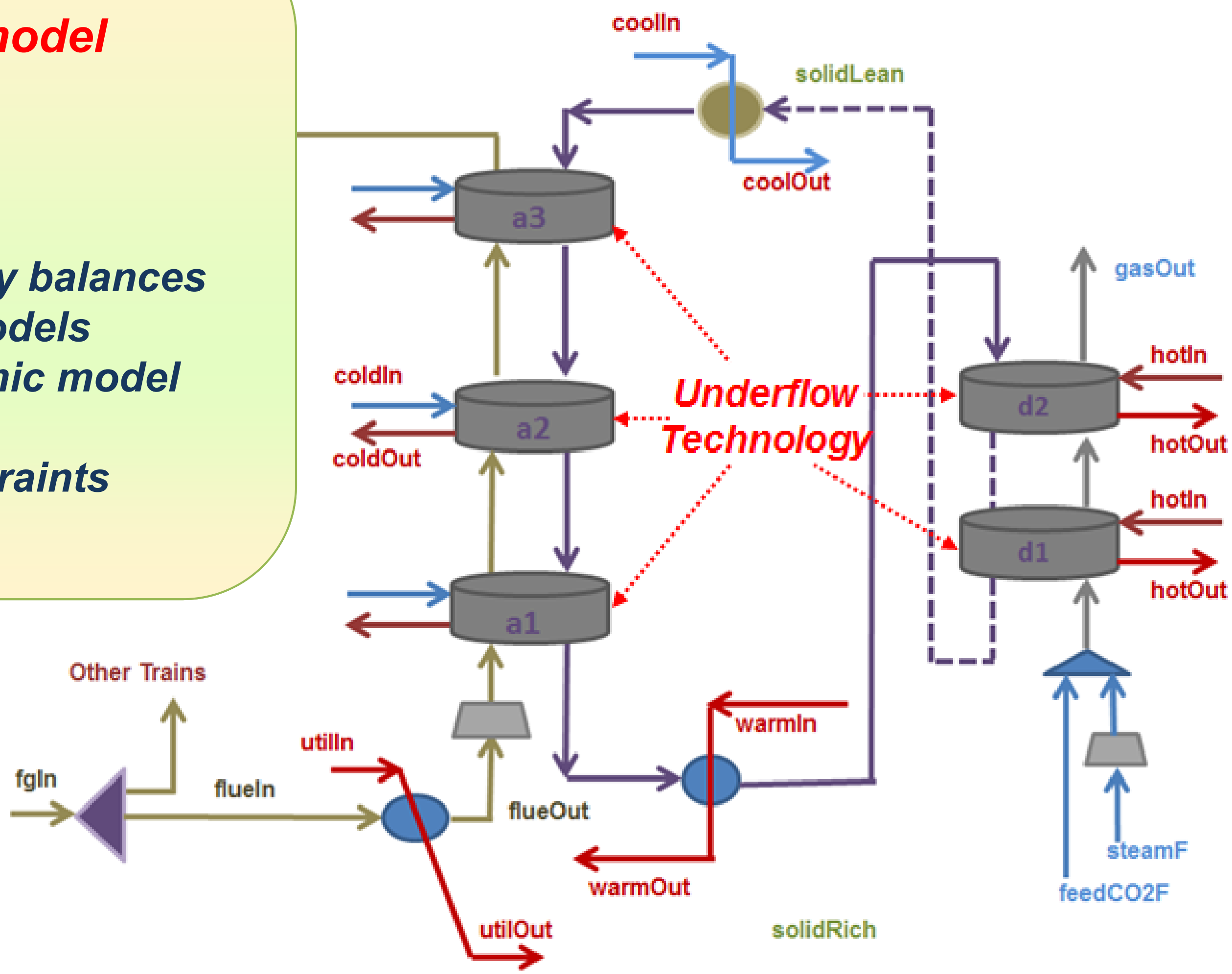
SURROGATE MODEL RESULTS



SYSTEM OPTIMIZATION

Mixed-integer nonlinear programming model

- ***Economic model***
- ***Process model***
- ***Material balances***
- ***Hydrodynamic/Energy balances***
- ***Reactor surrogate models***
- ***Link between economic model and process model***
- ***Binary variable constraints***
- ***Bounds for variables***



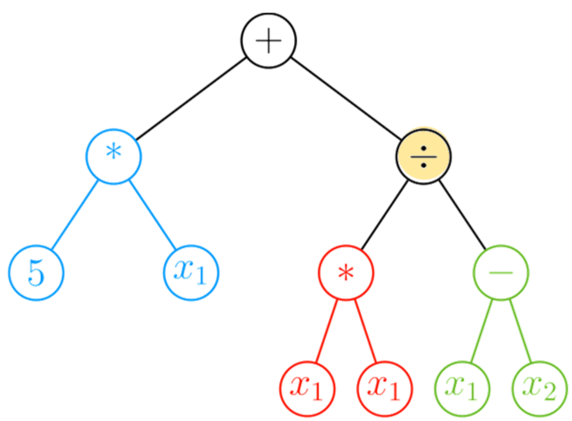
MINLP solved with BARON

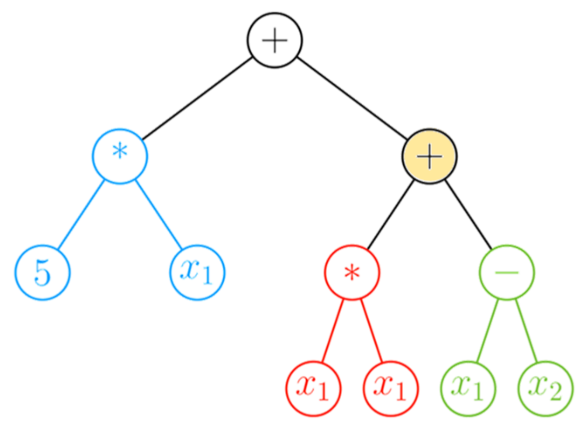
EXTENSIONS

- **Other metrics—L0L2 regularization, cross validation**
- **Thermodynamics** (HELMET; idaes.org distribution)
  - Equations of state
- **Kinetics** (RIPE; idaes.org distribution)
  - Simultaneous mechanism and parameter estimation
- **Symbolic regression**

$$5\,x_1 + \frac{(x_1)^2}{x_1 - x_2}$$

$$5\,x_1 + (x_1)^2 + x_1 - x_2$$





CONCLUSIONS

- **ALAMO provides algebraic models that are**
  - ✓ Accurate
  - ✓ Simple
  - ✓ Generated from a minimal number of data points
- **ALAMO's constrained regression facility allows modeling of**
  - ✓ Bounds on response variables
  - ✓ Variable groups
  - ✓ Forthcoming: constraints on gradient of response variables
- **Built on top of state-of-the-art optimization solvers**
- **Extends the applicability of algebraic optimization to simulation- and experiment-based optimization**