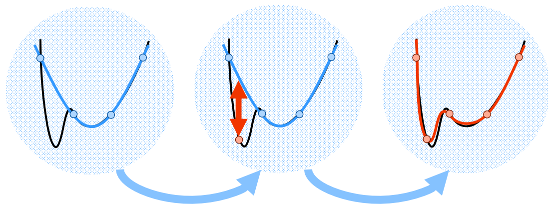
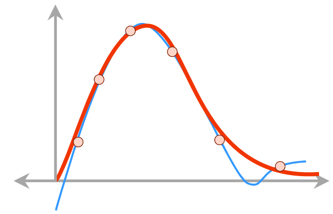




Machine learning, ALAMO, and constrained regression

Nick Sahinidis

Acknowledgments:

Alison Cozad, David Miller, Zach Wilson

Carnegie Mellon University

CAPD CENTER

NATIONAL ENERGY TECHNOLOGY LABORATORY

IDAES Institute for the Design of Advanced Energy Systems

The Optimization Firm

MACHINE LEARNING PROBLEM

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

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

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

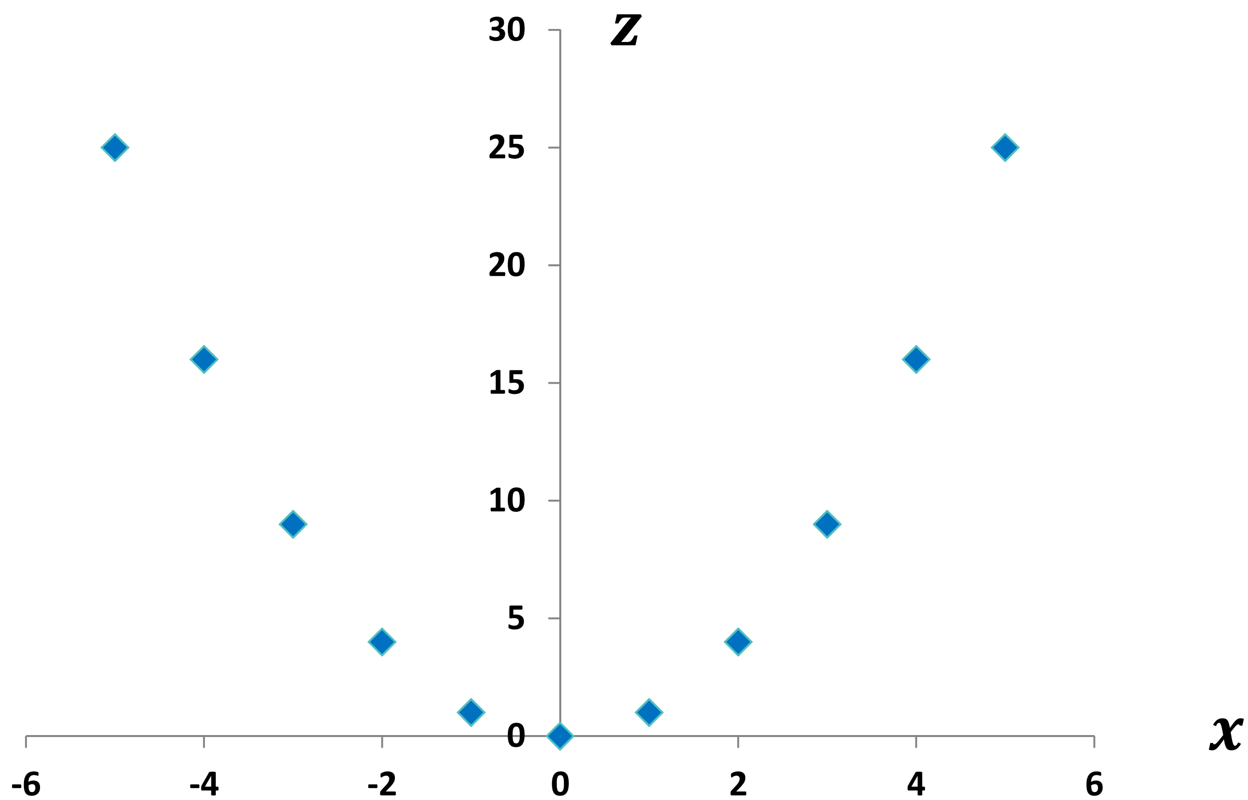
Independent variables:
Operating conditions, inlet flow properties, unit geometry

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

DESIRED MODEL ATTRIBUTES

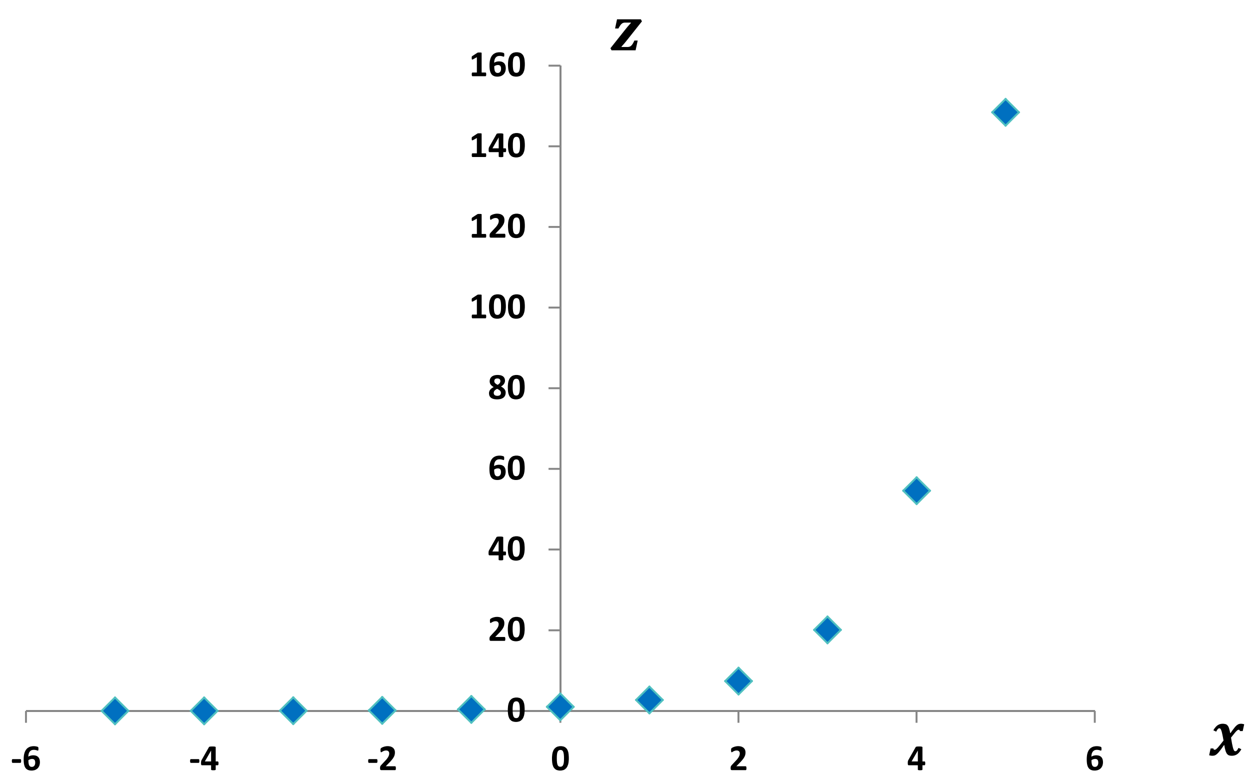
- Accurate
  - We want to reflect the true nature of the system
- Simple
  - Interpretable
  - Usable for algebraic optimization
- Generated from a minimal data set
  - Reduce experimental and simulation requirements

FITTING MODELS TO DATA



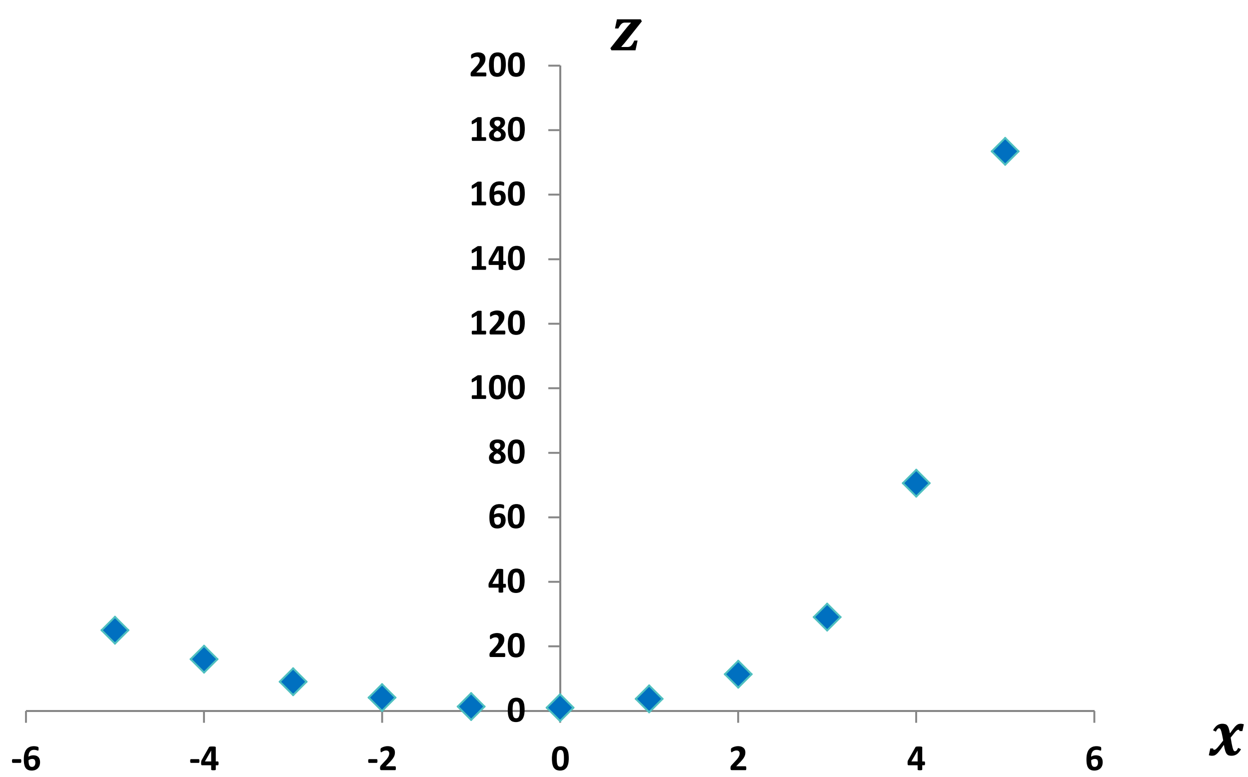
$$z = x^2$$

FITTING MODELS TO DATA



$$z = \exp(x)$$

FITTING MODELS TO DATA



$$z = x^2 + \exp(x)$$

EXAMPLE ALAMO INPUT FILE

```
ninputs 1
noutputs 1
xmin -5
xmax 5
xlabel x
zlabel z

ndata 11

BEGIN_DATA
-5      25
-4      16
-3      9
-2      4
-1      1
0       0
1       1
2       4
3       9
4       16
5       25
END_DATA

logfcns 1
expfcns 1
sinfcns 1
cosfcns 1
monomialpower 1 2 3
```

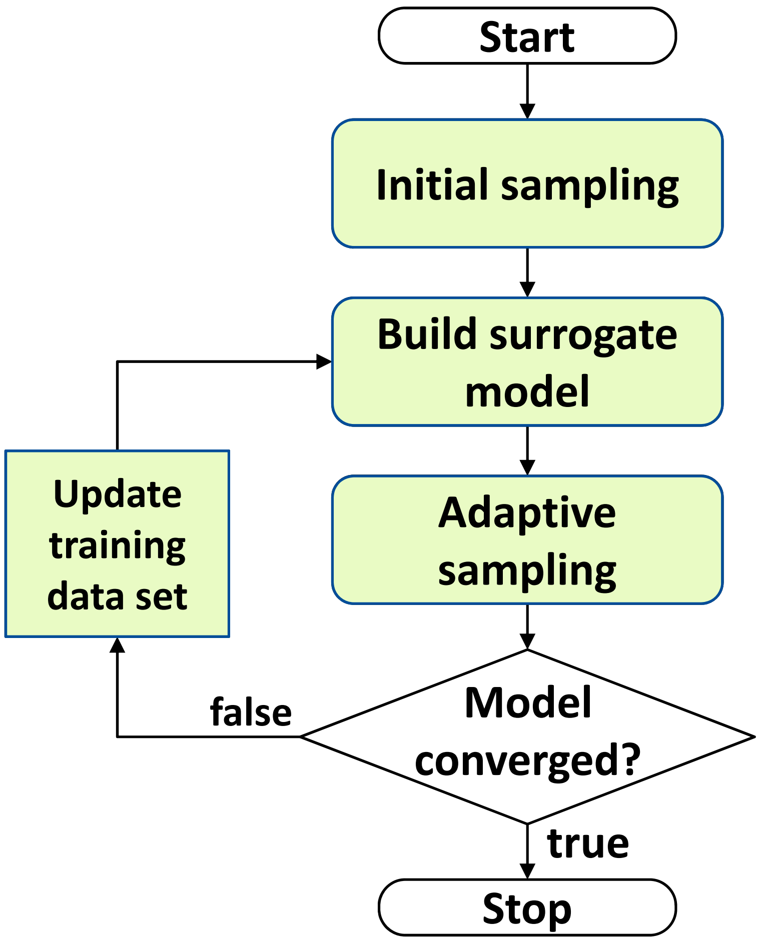
128 alternative models

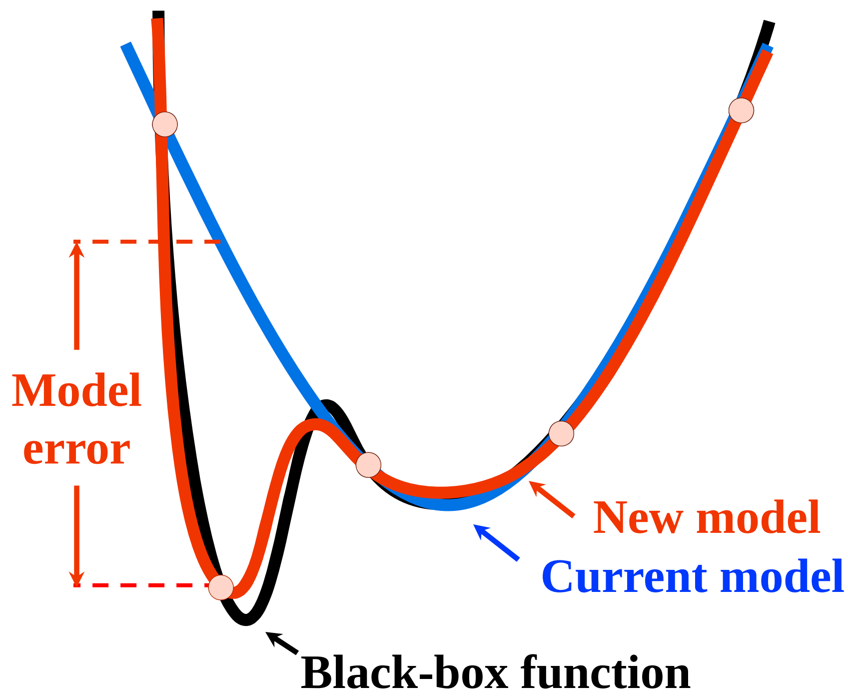
ALAMO OUTPUT

```
Step 1: Model building using BIC

Model building for variable z
----
BIC = -0.100E+31 with z = x**2.0

Calculating quality metrics on observed data set.

Quality metrics for output z
----------------------------
SSE OLR:               0.00
SSE:                   0.00
RMSE:                  0.00
R2:                    1.00
Model size:            1
BIC:                   -0.100E+31
Cp:                    -9.00
AICc:                  -0.100E+31
HQC:                   -0.100E+31
MSE:                   0.00
SSEp:                  0.00
RIC:                   3.89

Total execution time 0.30E-02 s
```

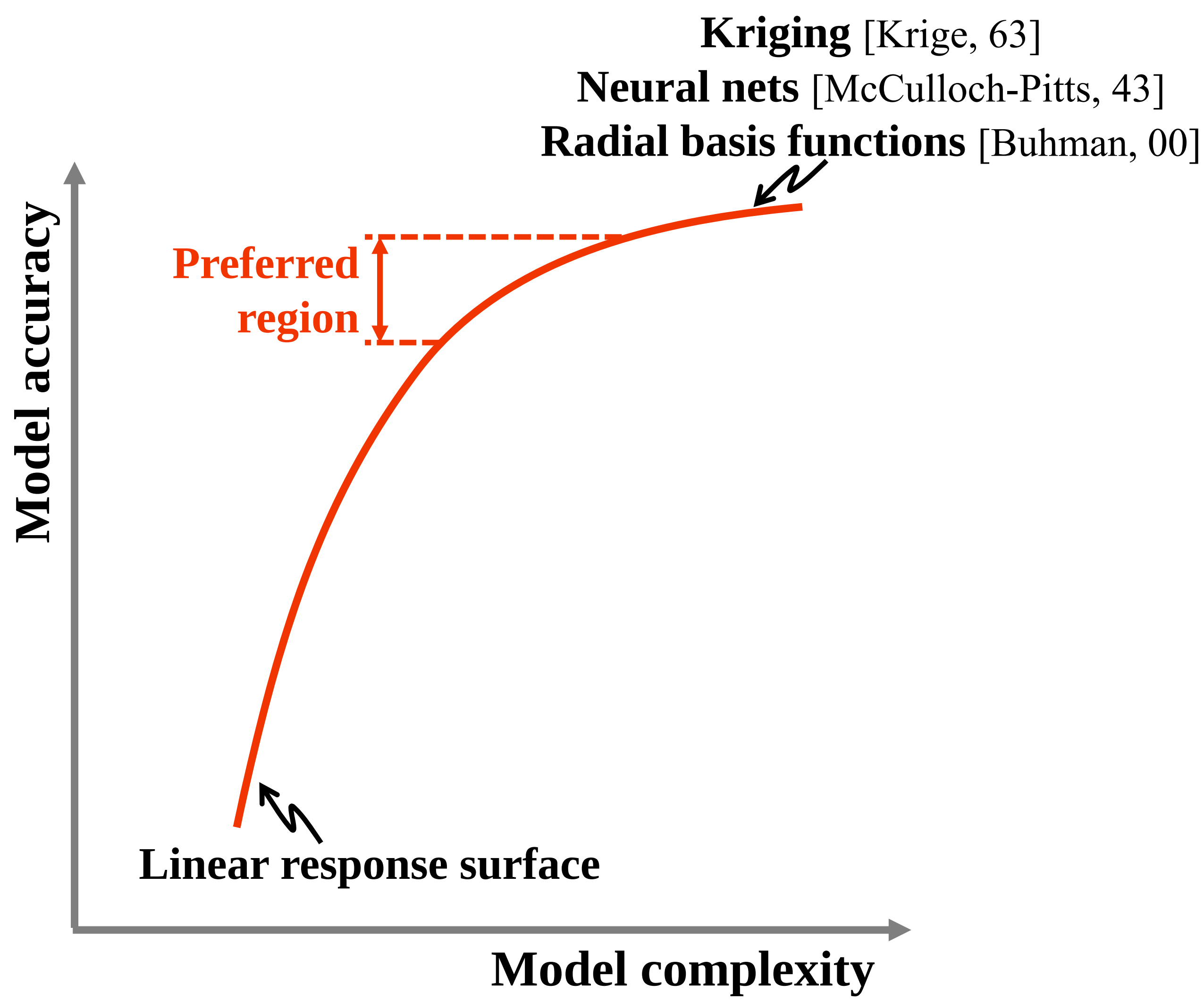
ALAMO

Automated Learning of Algebraic MOdels





MODEL COMPLEXITY TRADEOFF



MODEL IDENTIFICATION

- **Identify the functional form and complexity of the surrogate models $z = f(x)$**
- **Seek models that are combinations 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. **Radial basis functions for parametric regression**
  3. **User-specified basis functions for tailored regression**

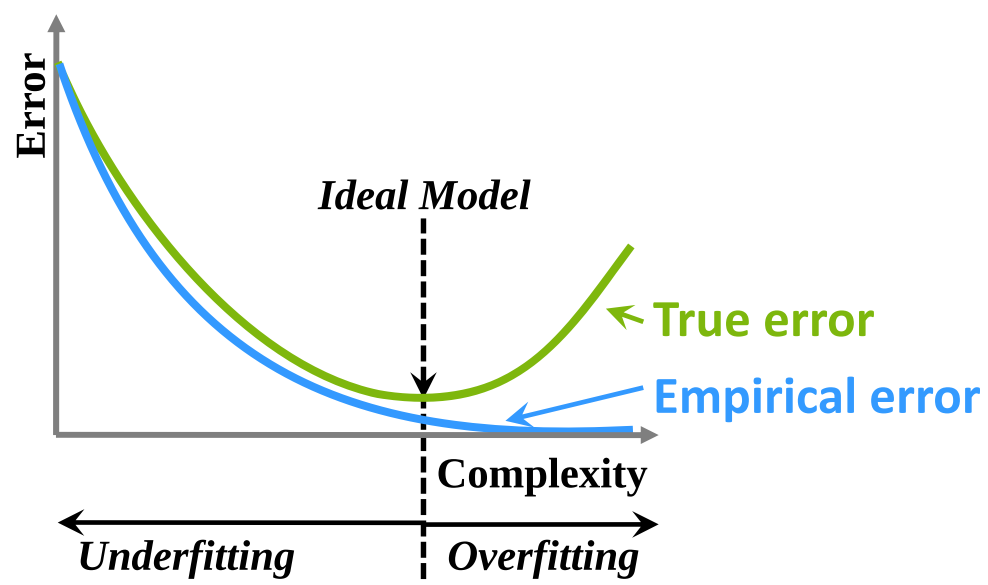
OVERFITTING AND TRUE ERROR

- **Step 1:** Define a large set of potential basis functions

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

- **Step 2:** Model reduction

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



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

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

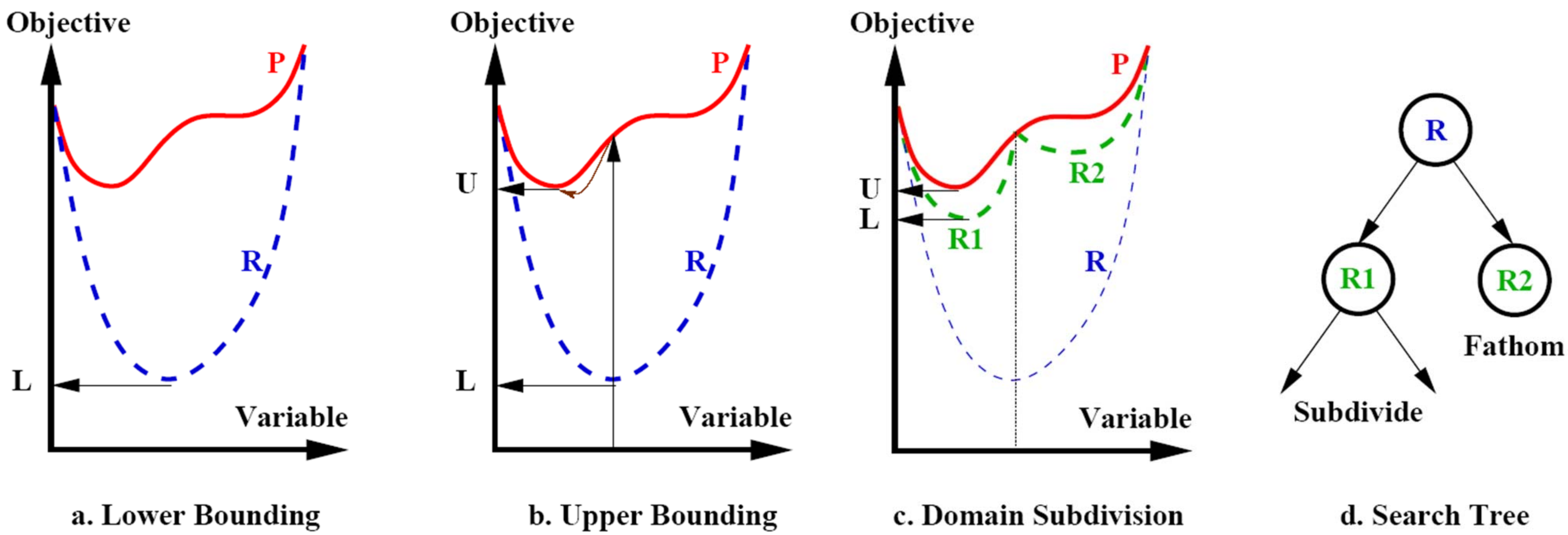
Bayes Information Criterion

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

Mean Squared Error

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

BRANCH-AND-BOUND

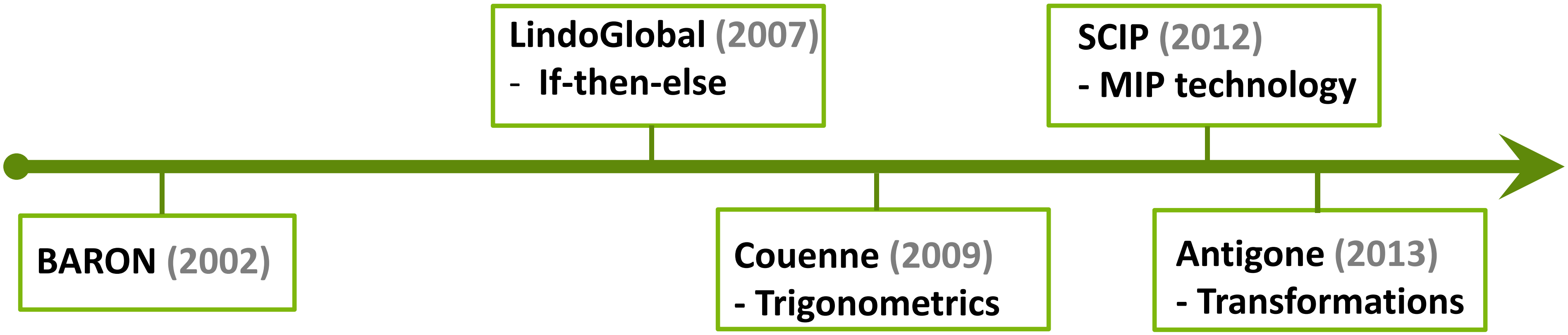


Falk and Soland, 1969; Soland 1971

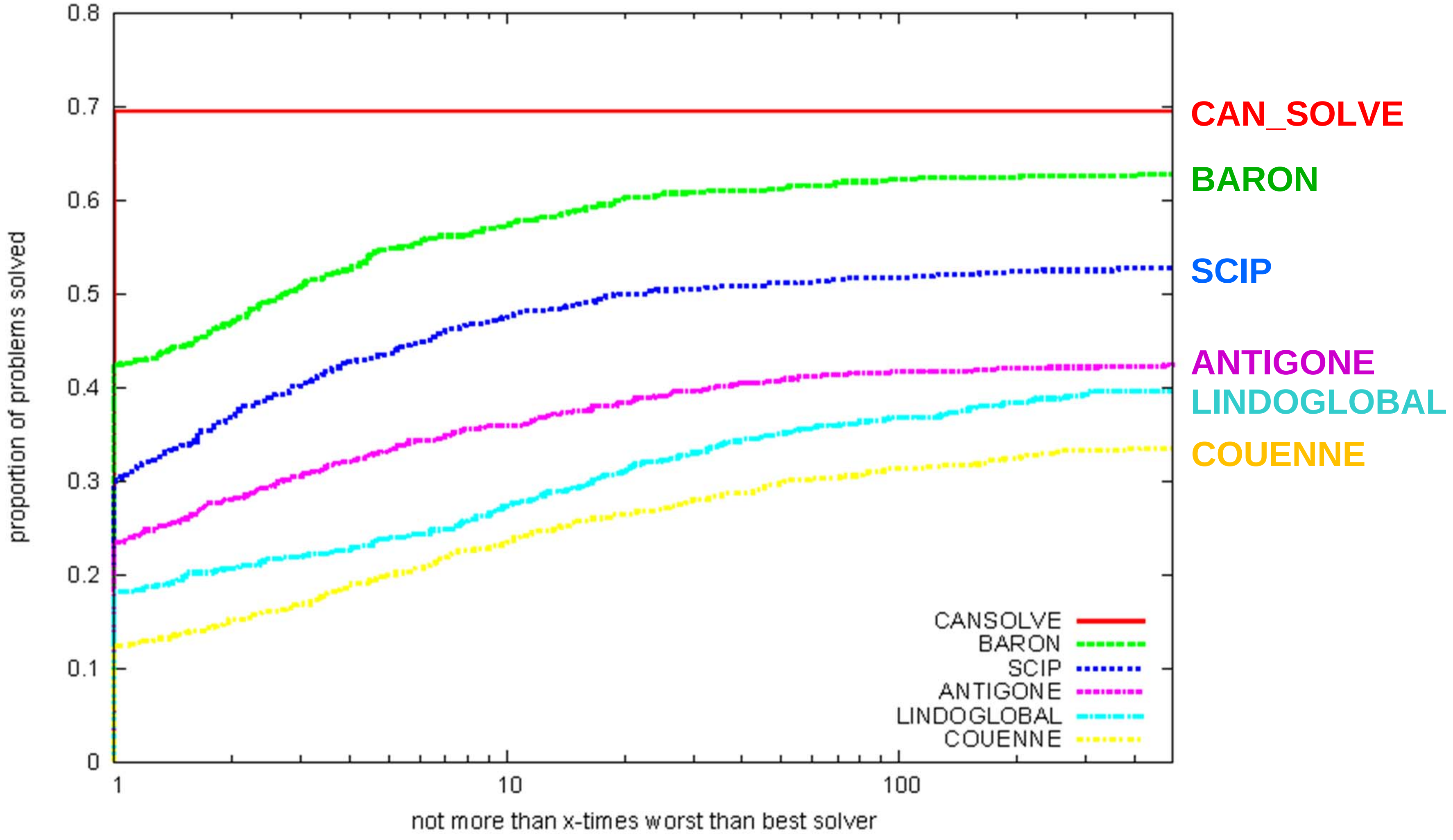
BRANCH-AND-REDUCE

- **Constraint propagation and duality-based bounds tightening**
  - Ryoo and Sahinidis, 1995, 1996
  - Tawarmalani and Sahinidis, 2004
- **Finite branching rules**
  - Shectman and Sahinidis, 1998
  - Ahmed, Tawarmalani and Sahinidis, 2004
- **Convexification**
  - Tawarmalani and Sahinidis, 2001, 2002, 2004, 2005
  - Khajavirad and Sahinidis, 2012, 2013, 2014, 2016
  - Zorn and Sahinidis, 2013, 2013, 2014
- **Implemented in BARON**
  - First deterministic global optimization solver for NLP and MINLP

GLOBAL MINLP SOLVERS IN GAMS



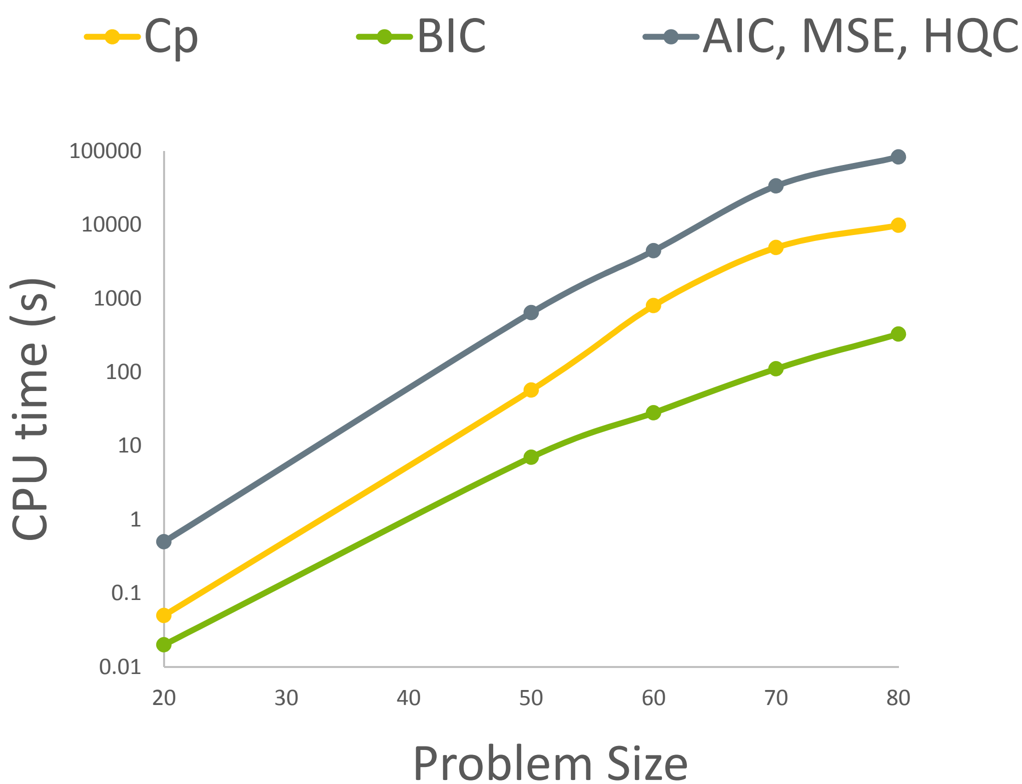
GLOBAL MINLP SOLVERS ON MINLPLIB2



Con: 1893 (1—164,321), Var: 1027 (3—107,223), Disc: 137 (1—31,824)

CPU TIME COMPARISON OF METRICS

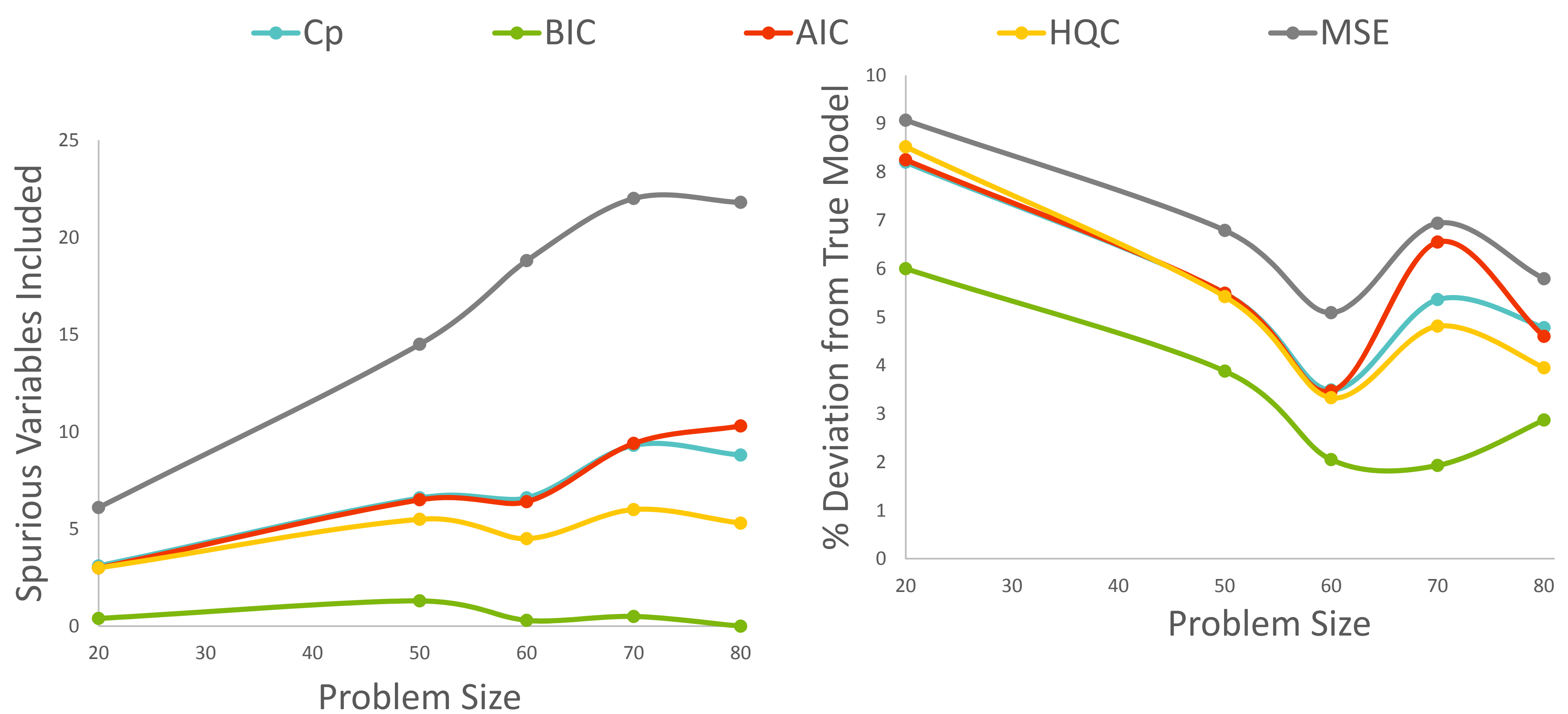
- **Eight benchmarks from the UCI and CMU data sets**
- **Seventy noisy data sets were generated with multicolinearity and increasing problem size (number of bases)**



- **BIC solves more than two orders of magnitude faster than AIC, MSE and HQC**
  - **Optimized directly via a single mixed-integer convex quadratic model**

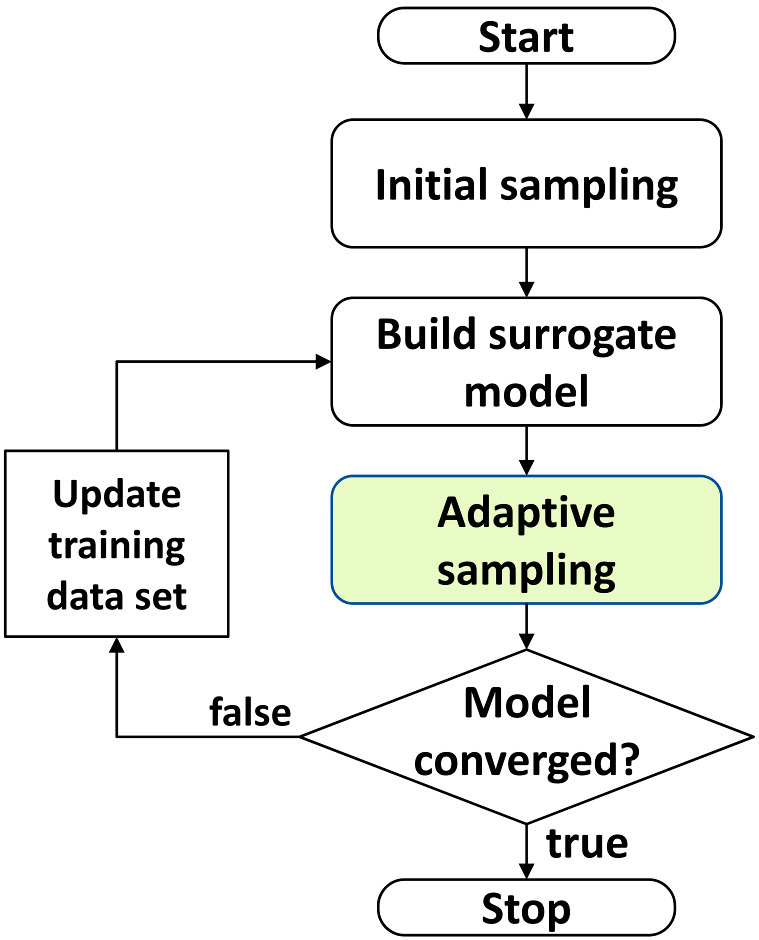
MODEL QUALITY COMPARISON

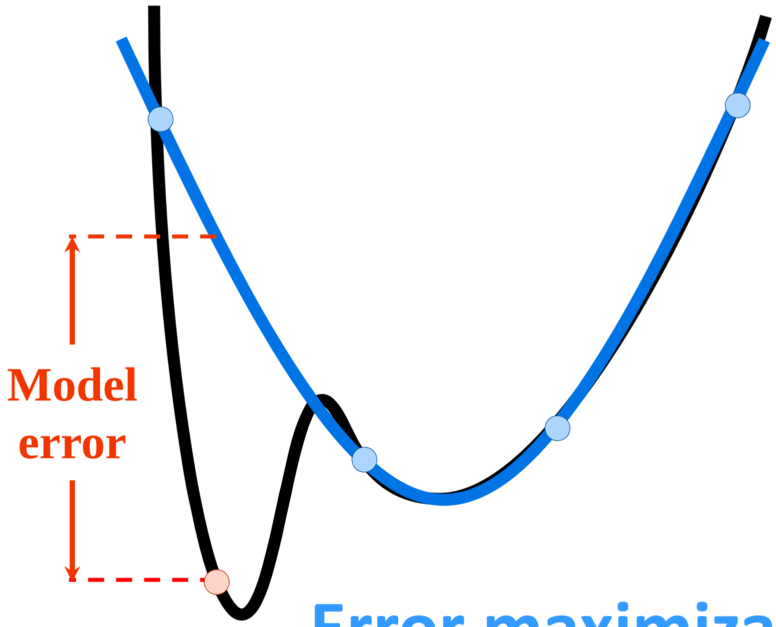
- **BIC leads to smaller, more accurate models**
  - **Larger penalty for model complexity**



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

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

  - Optimized using derivative-free solver SNOBFIT (Huyer and Neumaier, 2008)
  - SNOBFIT outperforms most derivative-free solvers (Rios and Sahinidis, 2013)

COMPUTATIONAL RESULTS

- **Goal – Compare methods on three target metrics**

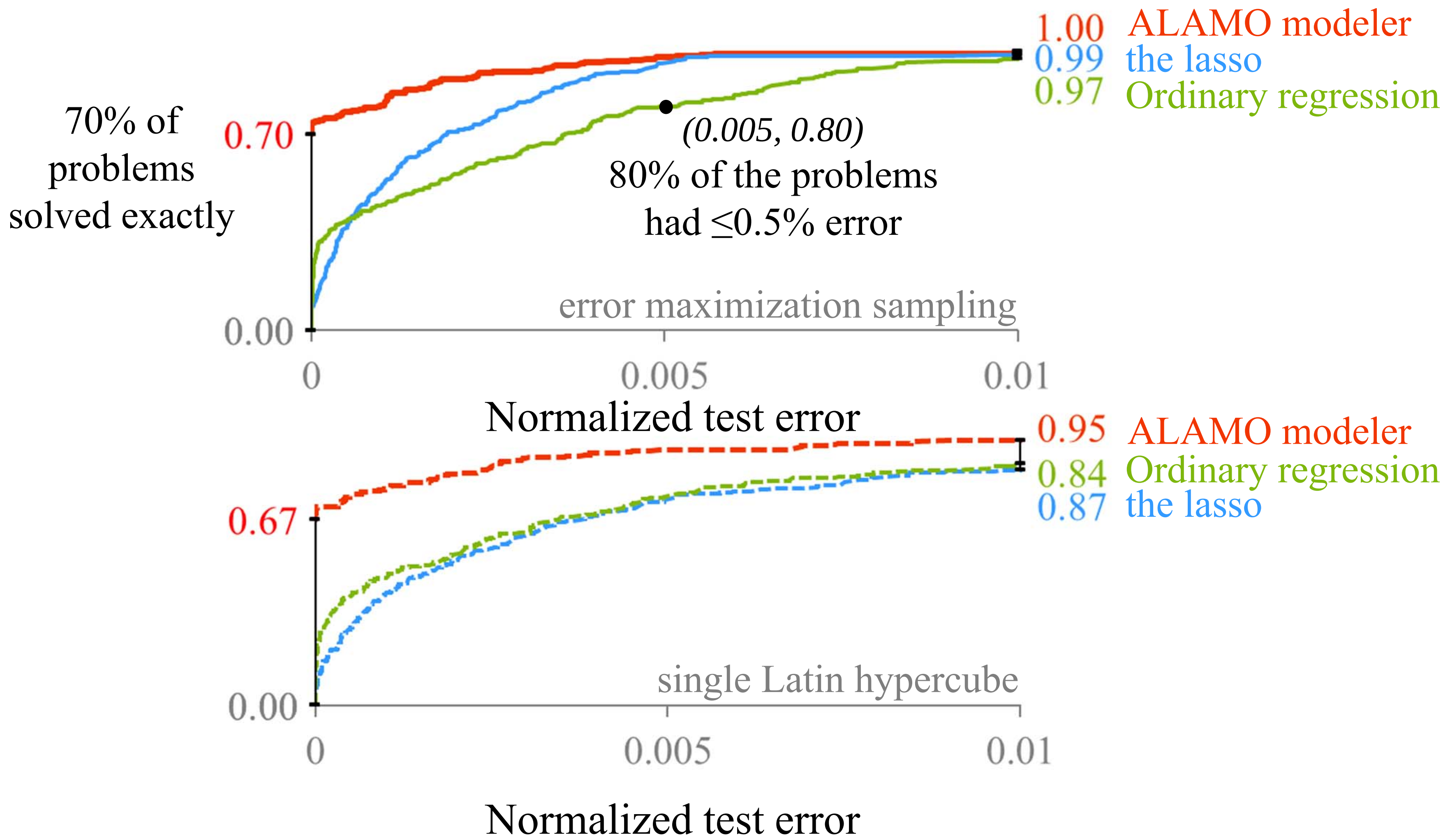
1 Model accuracy	2 Data efficiency	3 Model simplicity

- **Modeling methods compared**
  - **ALAMO modeler – Proposed best subset methodology**
  - **The LASSO – The lasso regularization**
  - **Ordinary regression – Ordinary least-squares regression**

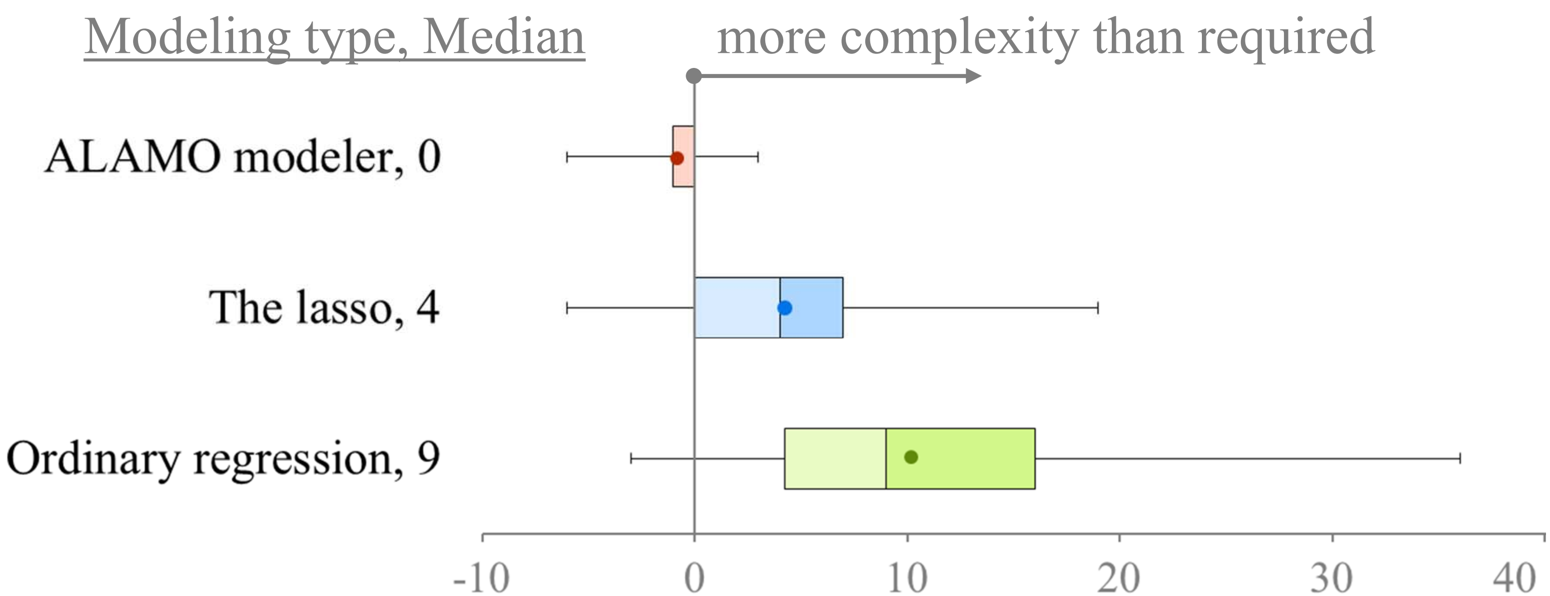
- **Sampling methods compared (over the same data set size)**
  - **ALAMO sampler – Proposed error maximization technique**
  - **Single LH – Single Latin hypercube (no feedback)**

1 Model accuracy | **2 Data efficiency** | 3 Model simplicity

Fraction of problems solved



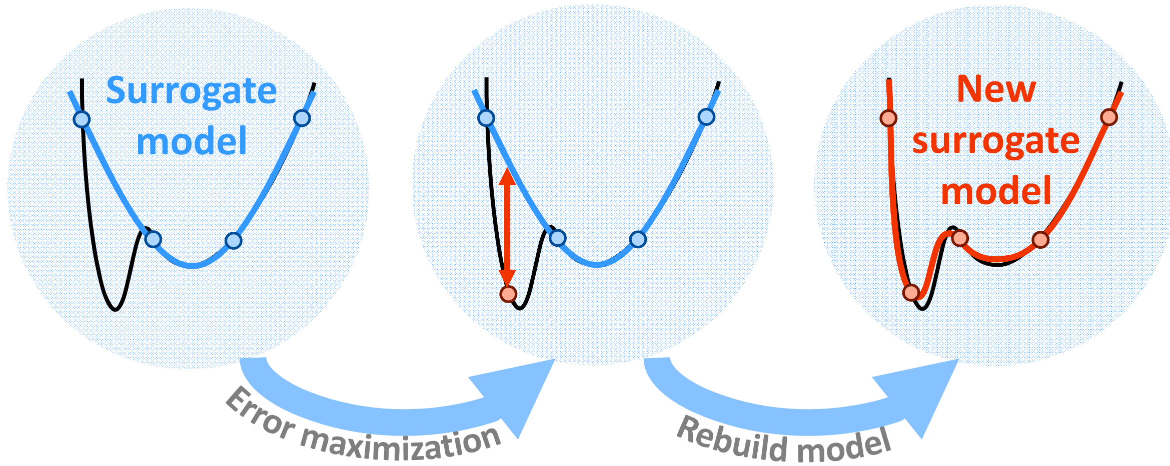
1 Model accuracy | 2 Data efficiency | **3 Model simplicity**



$$\left[\text{Number of terms in the surrogate model}\right] - \left[\text{Number of terms in the true function}\right]$$

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

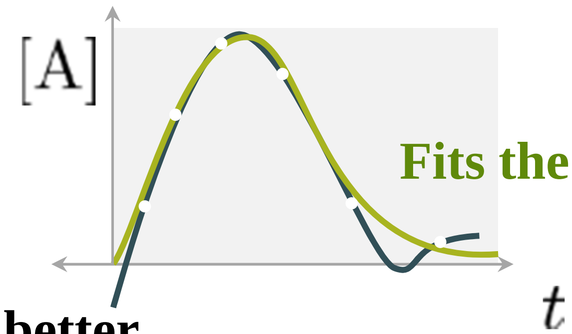
KEY INGREDIENT: OPTIMIZATION

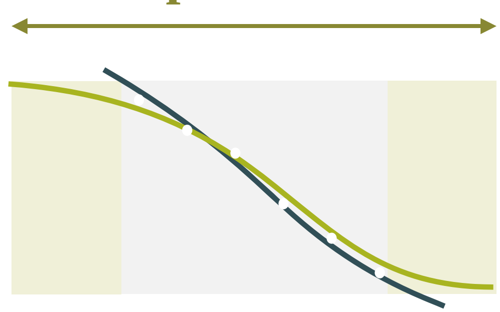


- **Surrogate model identification**
  - Simple, accurate model identification
  - Integer optimization
- **Error maximization sampling**
  - More information found per simulated data point
  - Derivative-free optimization

CONSTRAINED REGRESSION

$$0 \leq [\mathrm{A}]_t \leq [\mathrm{A}]^{\max}$$





Safe extrapolation

CONSTRAINED REGRESSION

Standard regression

$$\min_{\beta_1,\beta_2} \sum_{i=1}^{4} \left(z_i - \hat{z}(x_i;\beta_1,\beta_2)\right)^2$$

Surrogate model

easy

tough

Constrained regression

$$\min_{\beta_1,\beta_2} \sum_{i=1}^{4} \left(z_i - \hat{z}(x_i;\beta_1,\beta_2)\right)^2$$
$$\text{s.t.} \quad \beta_1 \geq \beta_2$$

$$\min_{\beta_1,\beta_2} \sum_{i=1}^{4} \left(z_i - \hat{z}(x_i;\beta_1,\beta_2)\right)^2$$
$$\text{s.t.} \quad \hat{z}(x_i;\beta_1,\beta_2) \geq 0 \quad \forall x$$

- **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

$$\min_{\beta_1, \beta_2} \quad \sum_{i=1}^{4} \left(z_i - [\beta_1 \, x + \beta_2 \, x^3]\right)^2$$

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

1 parametric constraint

4 β-constraints

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

$$\min_{\beta_1, \beta_2} \quad \sum_{i=1}^{4} \left(z_i - [\beta_1 \, x + \beta_2 \, x^3]\right)^2$$

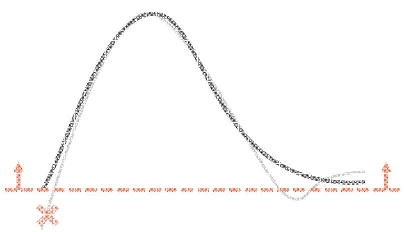
$$\text{s.t.} \quad \begin{cases} 0.240 \, \beta_1 + 0.0138 \, \beta_2 \geq 0 \\ 0.281 \, \beta_1 + 0.0223 \, \beta_2 \geq 0 \\ 0.120 \, \beta_1 + 0.00173 \, \beta_2 \geq 0 \\ 0.138 \, \beta_1 + 0.00263 \, \beta_2 \geq 0 \end{cases}$$

Global optimization problems solved with BARON

TYPES OF RESTRICTIONS

Response bounds

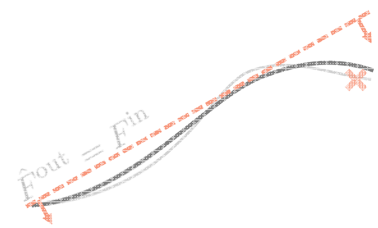
$$[\hat{A}]_t \geq 0$$



pressure, temperature, compositions

Individual responses

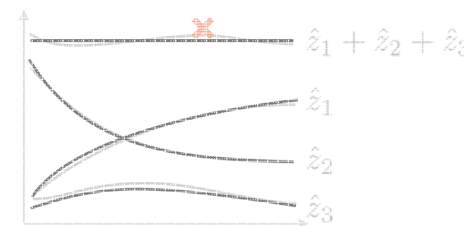
$$\hat{F}^{\text{out}}(x) \leq F^{\text{in}}$$



mass and energy balances, physical limitations

Multiple responses

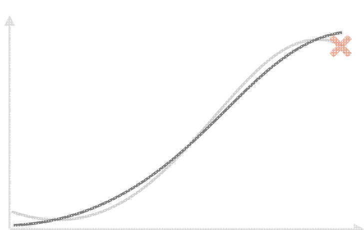
$$\hat{z}_1 + \hat{z}_2 + \hat{z}_3 = 1$$



mass balances, sum-to-one, state variables

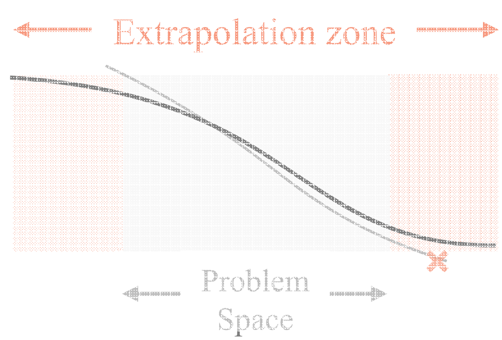
Response derivatives

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



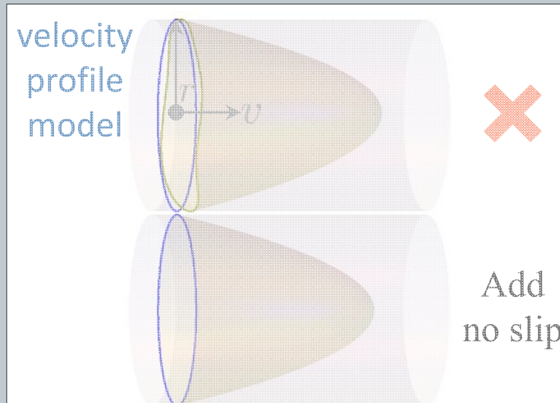
monotonicity, numerical properties, convexity

Alternative domains



safe extrapolation, boundary conditions

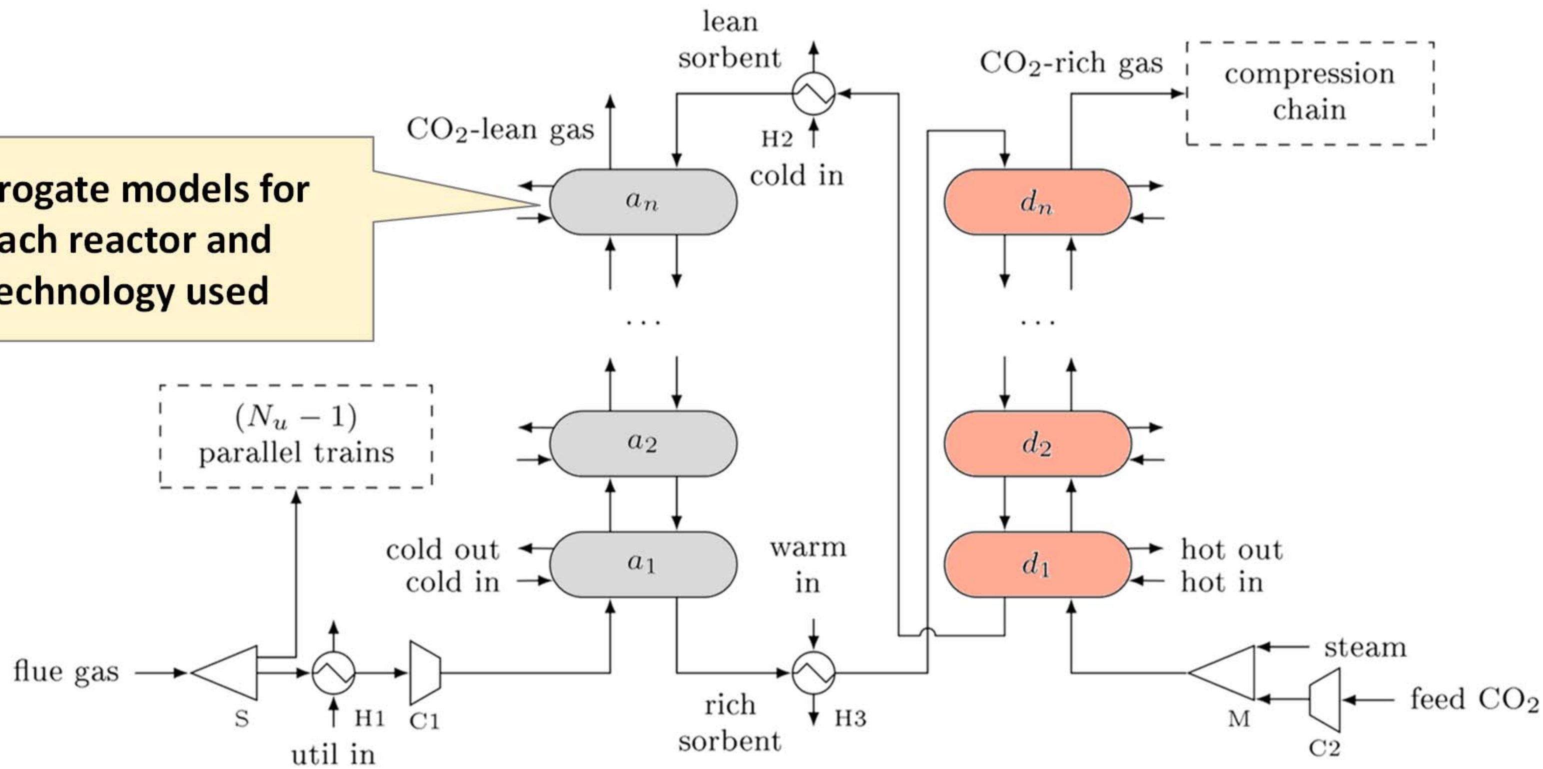
Boundary conditions



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

CARBON CAPTURE SYSTEM DESIGN

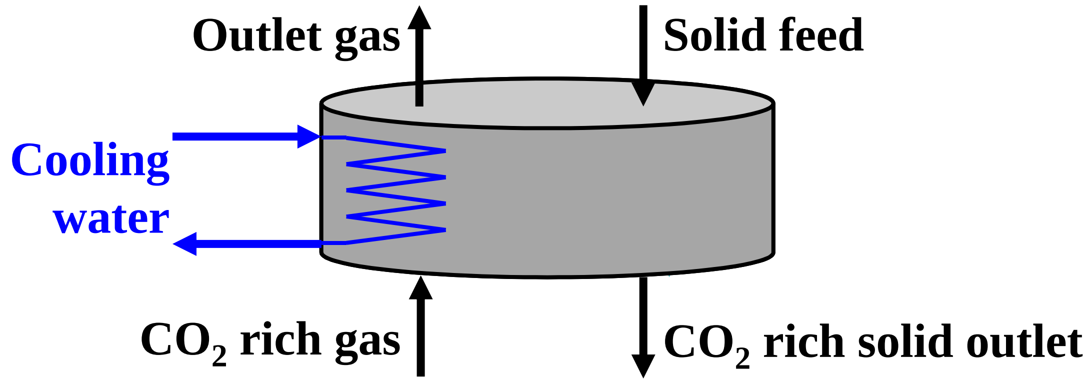
Surrogate models for each reactor and technology used



- **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

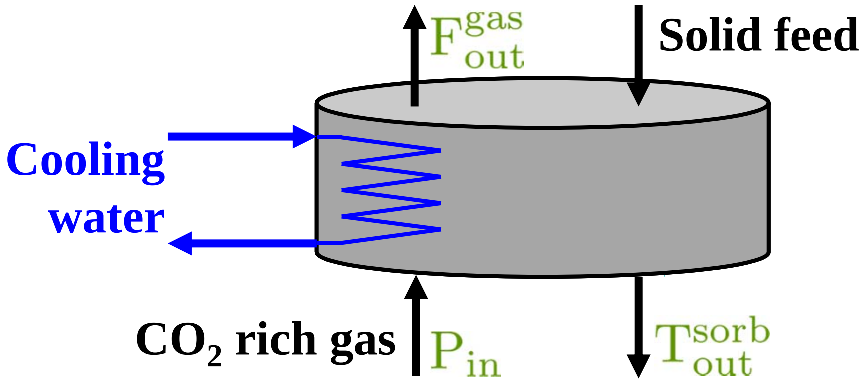
Bubbling fluidized bed adsorber diagram



- **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



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

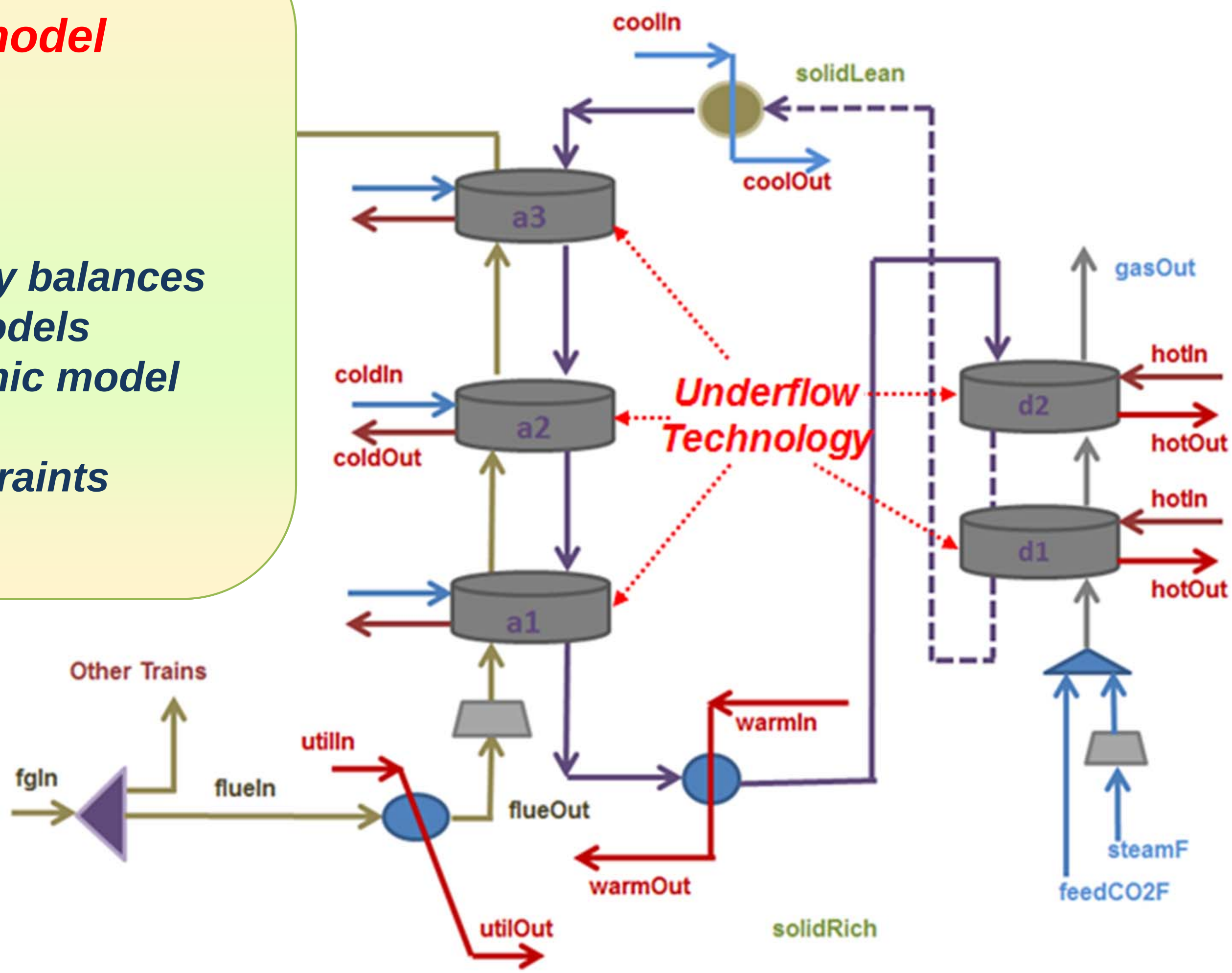
$$
T^{sorb}_{out} = 1.0\,\mathrm{T^{gas}_{in}} - \frac{\left(1.77\cdot 10^{-10}\right)\,\mathrm{NX}^2}{\gamma^2} - \frac{3.46}{\mathrm{NX\,T^{gas}_{in}\,T^{sorb}_{in}}} + \frac{1.17\cdot 10^4}{\mathrm{F^{sorb}\,NX\,xH2O^{ads}_{in}}}
$$

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

SUPERSTRUCTURE 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

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
  - ✓ Convexity/monotonicity of response variables
  - ✓ Variable groups

- **Built on top of state-of-the-art optimization solvers**

- **ALAMO site: archimedes.cheme.cmu.edu/?q=alamo**