



Optimization for Development of Reliable Fundamental Models

Kim McAuley

Department of Chemical Engineering,
Queen's University

Outline

- Modeling and models
 - Case study: Nylon 66 degradation
- Optimization during model building
 - Estimability analysis for ranking parameters
 - Deciding how many parameters to estimate
 - Sequential experimental design

Why Model Chemical Processes?

Objectives of chemical companies: \$\$\$

- Make products with targeted properties
- Devise improved operating strategies
- Bring new products to market quickly
- Develop process knowledge for design and trouble shooting

Why Model Chemical Processes?

Objectives of chemical companies:

\$\$\$

- Make products with targeted properties
- Devise improved operating strategies
- Bring new products to market quickly
- Develop process knowledge for design and trouble shooting

Models can help companies to:

- Design and test control schemes
- Optimize grade changes
- Design process improvements
- Plan experiments
- Test theories about what has gone wrong
- Capture, store and distribute knowledge

Fundamental Models	Empirical Models
Require detailed scientific knowledge	Require statistical skill, but less process knowledge and less time and effort
Can predict influence of changing equipment and operating conditions if scientific principles are known	May not be reliable when process changes occur that are not in training data
Can apply to many new situations	Usually only for system used to collect data
Useful for summarizing and disseminating scientific knowledge	Useful for exploring which variables are important

Steps in Fundamental Model Development

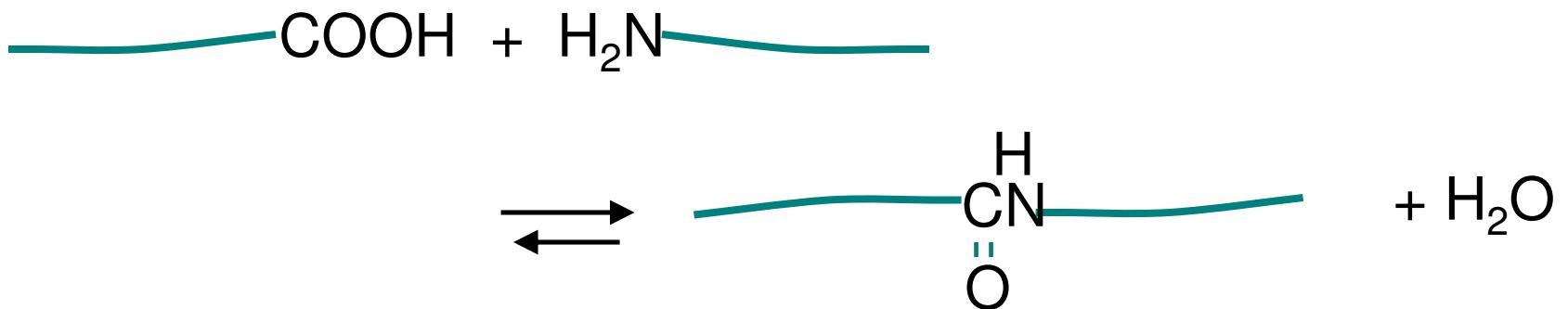
1. Derive equations based on knowledge and assumptions
2. Solve equations numerically using initial guesses for the parameters
3. Get values of some parameters from the literature
4. Use data from existing or new experiments to estimate remaining model parameters
5. Test model fit
6. Design and conduct more experiments?
7. Revise model structure based on new knowledge?
8. Validate model using additional experimental data
9. Use the model

Case Study

Model for Thermal Degradation of Nylon 66

Hadis Karimi
Mark Schaffer

Nylon Polymerization Basics



- Water is removed to obtain high molecular weight
- Water also influences degradation reactions

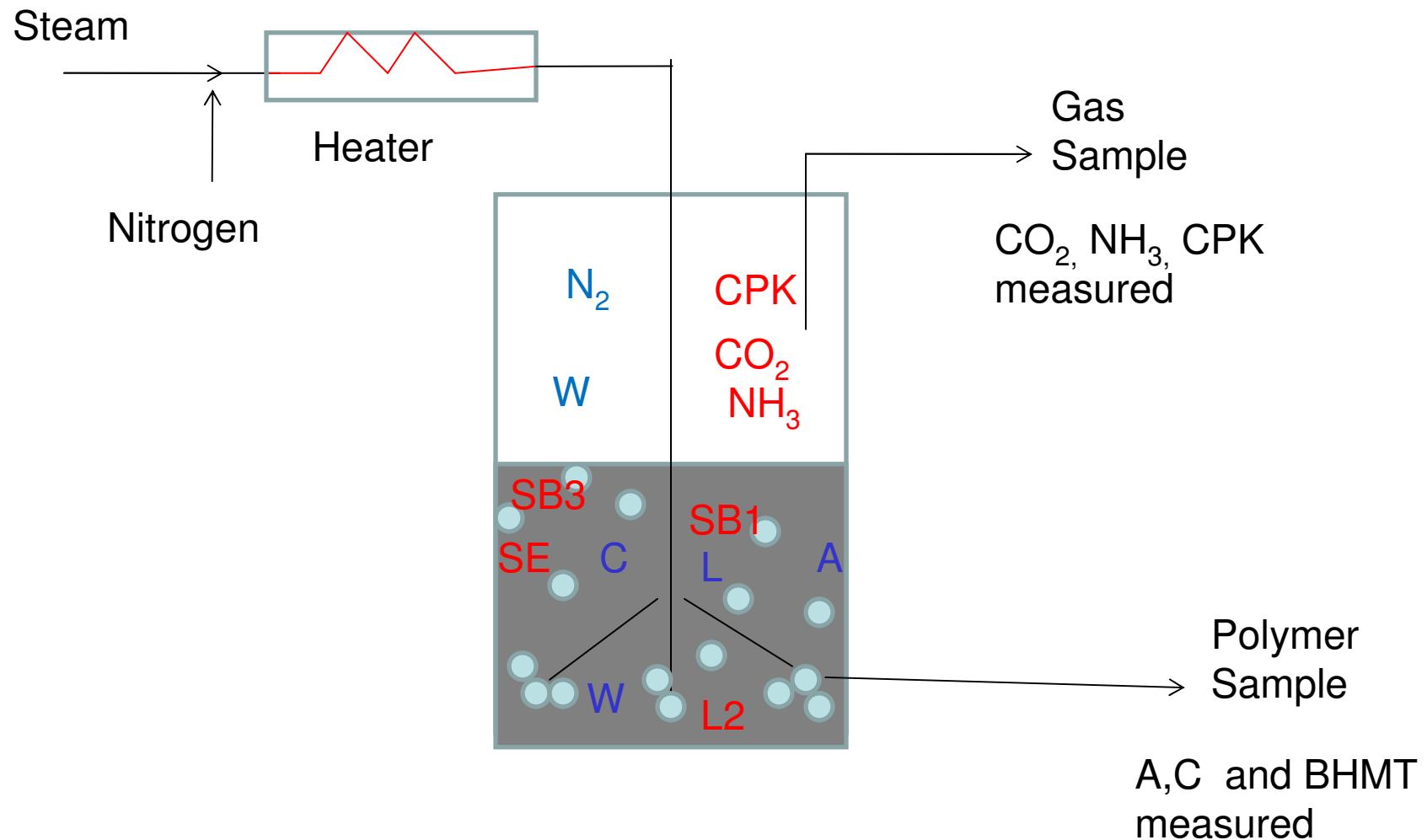
Nylon 66

- Nylon 66 is used to make fibres, film and molded objects



- Thermal degradation reactions result in:
 - changes in end groups of polymer molecules
 - evolution of gaseous degradation products
 - branching
 - Problems for dyeability and spinning into fibres

Thermal degradation experiments



Experimental conditions

Run	Temperature [°C]	P _w [kPa]	Duration [h]
1	281	92	9.1
2	275	57	10.6
3	285	93	6.6
4	292	59	3.1
5	286	23	2
6	290	101 → 23	4.5

Experiments cover operating range of industrial interest

14 Differential Equations

$$\frac{d[A]}{dt} = -R_7 - R_1$$

$$\frac{d[W]}{dt} = R_2 - R_6 + R_8 + R_1 + R_{11} - R_{m,W}$$

$$\frac{d[CPK]}{dt} = R_6 - R_{11} - R_{m,CPK}$$

.

.

$$\frac{d[P244end]}{dt} = R_{11}$$

Rate Expressions for 11 Reactions

$$R_1 = k_c \left([C][A] - \frac{[L][W]}{K_a} \right)$$

$$R_2 = k_2 [C]$$

$$R_3 = k_3 [L]$$

.

.

.

$$R_{11} = k_{11} [SB1][CPK]$$

26 Model Parameters

$$k_i = k_{i,0} \exp\left[-\frac{E_i}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right]$$

$$K_{\text{eq}} = k_{i,\text{eq}} \exp\left[\frac{\Delta H_i}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right]$$

$$R_{m,w} = k_{L,w} a ([w] - [w]_{\text{eq}})$$

5 Initial concentrations also unknown in each run

Initial Parameter Guesses

Schaffer

$k_{20}, E_2, k_{30}, E_3, k_{40}, E_4, k_{60}, K_{6\text{eq}}, \Delta H_6, k_{70}, E_7, k_{90}, k_{100}, E_9, (k_L a)_W, (k_L a)_{\text{CPK}}$

**Varziri
(nylon 612)**

$k_{\text{co}}, K_{\text{ao}}, E_c, a, \text{ and } \Delta H$

Zero

initial conditions for [SE], [A2] , [SB1], [SB2], [SB3]

?

$K_{2\text{eq}}, \Delta H_2, E_6, K_{6\text{eq}}, \Delta H_6, k_{110} \text{ and } E_{11}$

Estimating Parameters

$$J = \sum \left(\frac{[A]_m - [A]}{s_A} \right)^2 + \sum \left(\frac{[C]_m - [C]}{s_C} \right)^2 + \sum \left(\frac{[BHMT]_m - [BHMT]}{s_{BHMT}} \right)^2 \\ + \sum \left(\frac{N_{CO_2m} - N_{CO_2}}{s_{CO_2}} \right)^2 + \sum \left(\frac{N_{NH_3m} - N_{NH_3}}{s_{NH_3}} \right)^2 + \sum \left(\frac{N_{CPKm} - N_{CPK}}{s_{CPK}} \right)^2$$

Lsqnonlin optimizer in MATLAB™

- Default trust region reflective algorithm
- Numerical Jacobian
- Tolerance selection is important
- Sequential approach

Parameter Estimation

Parameter Estimation is Difficult:

- Too many parameters for available data
- Parameters with little influence
- Correlated effect of parameters
- Local minima

Selecting Parameters to Estimate

1. Rank parameters using estimability analysis
 - Requires scaled sensitivity matrix
 - Orthogonalization step accounts for correlations

2. Select number of parameters using mean-squared error criterion
 - Critical ratio calculated using objective function
 - Smallest value of corrected critical ratio indicates number of parameters to select

Estimability Analysis

- Select only the most important and uncertain parameters for estimation
- Leave other parameters at nominal values
- Estimability analysis ranks parameters from most estimable to least estimable based on:
 - influence on model predictions
 - correlation with other parameters
 - uncertainty in initial guesses

Estimability Analysis

The Approach

1. Construct sensitivity matrix containing derivatives of model predictions with respect to the parameters

$$Z = \begin{bmatrix} \frac{\partial y_1}{\partial \theta_1} \Big|_{t=t_1} & \dots & \frac{\partial y_1}{\partial \theta_P} \Big|_{t=t_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_R}{\partial \theta_1} \Big|_{t=t_1} & \dots & \frac{\partial y_R}{\partial \theta_P} \Big|_{t=t_1} \\ \frac{\partial y_1}{\partial \theta_1} \Big|_{t=t_2} & \dots & \frac{\partial y_1}{\partial \theta_P} \Big|_{t=t_2} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_R}{\partial \theta_1} \Big|_{t=t_N} & \dots & \frac{\partial y_R}{\partial \theta_P} \Big|_{t=t_N} \end{bmatrix}$$

2. Scale elements of Z to permit effective comparisons

$$\frac{\partial y_r}{\partial \theta_p} \Big|_{t_n} \frac{s_{\theta_p}}{s_{y_r}}$$

Estimability Analysis

3. Calculate magnitude of each column of Z .
Select parameter whose column has the largest magnitude as most estimable parameter.
4. Use selected column X_1 to calculate least-squares prediction of the full sensitivity matrix, Z , using
$$\hat{Z}_1 = X_1 \left(X_1^T X_1 \right)^{-1} X_1^T Z$$
5. Calculate residual matrix :

$$R_1 = Z - \hat{Z}_1$$

Estimability Analysis

6. Column of R_1 with the largest magnitude is the next most estimable parameter. Augment matrix X with new column
7. Obtain least-squares estimate of Z using augmented X
8. Continue until parameters ranked from most estimable to least estimable

Estimability Analysis

6. Column of R_1 with the largest magnitude is the next most estimable parameter. Augment matrix X with new column
7. Obtain least-squares estimate of Z using augmented X
8. Continue until parameters ranked from most estimable to least estimable

Problems

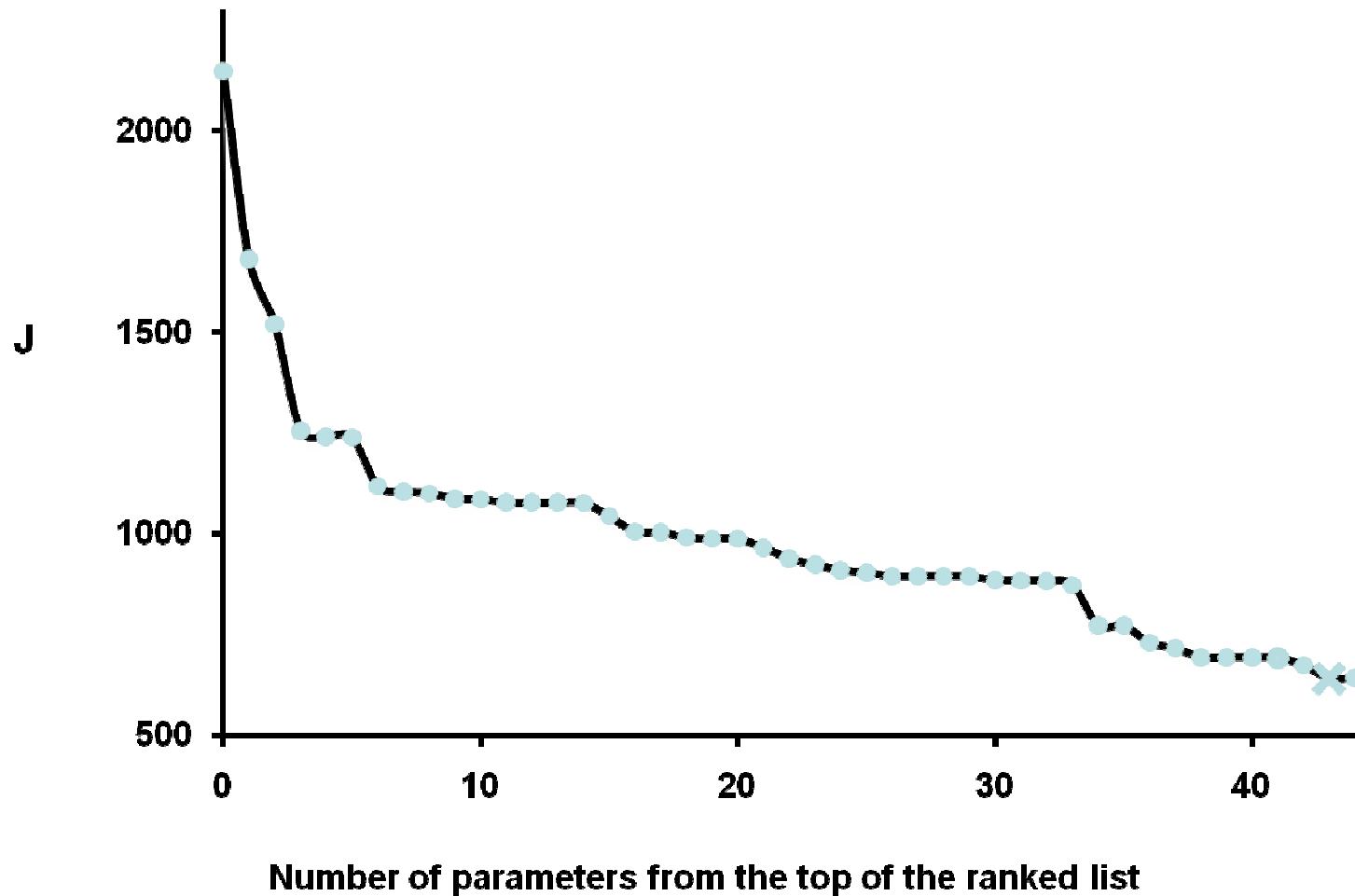
- Parameter ranking depends on assumptions about initial parameter guesses and scaling
- Selecting one parameter at a time can be suboptimal

Estimability Analysis Gives Ranked Parameter List for Nylon Model

Final Rank	Parameter Ranked List	Units	Initial Guess	Scaling Factor
1	k_{110}	$\text{Mg}\cdot\text{mol}^{-1} \text{ h}^{-1}$	1.00×10^{-2}	250.00×10^{-2}
2	k_{40}	h^{-1}	2	250
3	k_{70}	$\text{Mg}\cdot\text{mol}^{-1} \text{ h}^{-1}$	4.48×10^{-4}	250.00×10^{-4}
4	k_{20}	h^{-1}	1.58	500

How many should we estimate?

Objective Function



MSE Parameter Selection Criterion

Mean squared error

$$MSE(\hat{Y}) = E((\hat{Y} - Y_{true})^T (\hat{Y} - Y_{true}))$$

$$MSE(\hat{Y}) = \sum bias^2 + \sum variance$$

If $R_{Ck} = \frac{\sum bias_k^2}{\sum (variance_p - variance_k)} < 1$

estimating k parameters is preferable to estimating all p

Mean Squared Error Model Selection Criterion

- Use objective function values to estimate critical ratio:

$$r_{Ck} = \frac{(J_k - J_p)}{(p - k)}$$

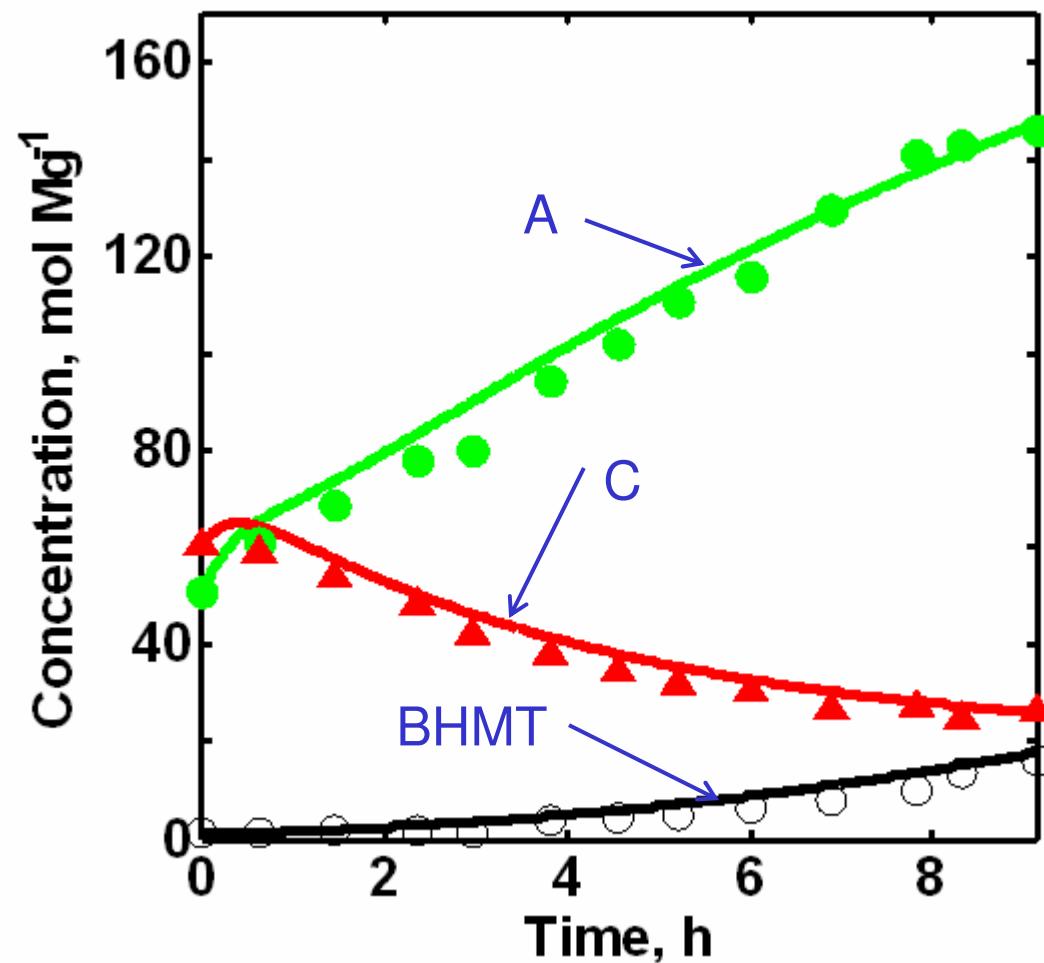
- $r_{Ck} < 1$ when model with k parameters estimated should give better predictions than model with all p parameters
- Use corrected critical ratio to select optimal number to estimate

$$r_{CC,k} = \frac{(p - k)}{N} (r_{CKub,k} - 1) \quad r_{CKub,k} = \max\left(r_{C,k} - 1, \frac{2}{p - k + 2} r_{C,k}\right)$$

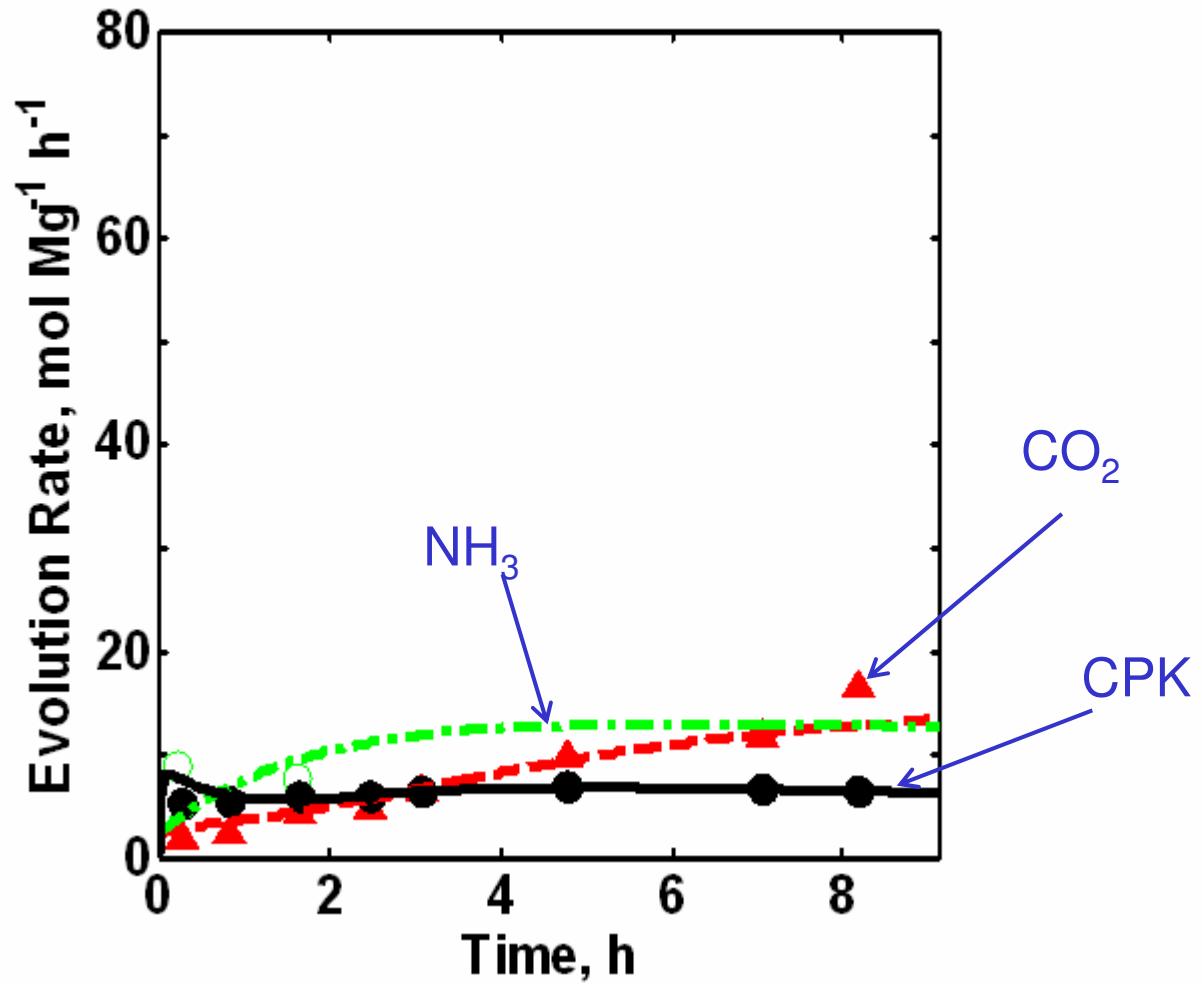
Optimal Number of Parameters in Nylon 66 Model

- 44 parameters could be estimated
- Model with top 43 parameters estimated corresponds to the smallest value of r_{cck} and should provide predictions with lowest mean squared error
- In other studies we have estimated far fewer parameters
 - E.g., expanded polystyrene model
Only 4 parameters out of 40 estimated

Nylon Polymer Properties (Run #3)



Gas Evolution Rates (Run #3)



Summary for Nylon 66 Modeling

- Kinetic model accounts for thermal degradation of molten nylon 66
- Estimability analysis and MSE criterion determined that 43 out of 56 model parameters should be estimated to provide the best predictions
- Model describes the experimental polymer property data and gas evolution well
- Model will be useful for selecting temperature and moisture level during final stages of industrial nylon 66 production

Summary for Nylon 66 Modeling

- Kinetic model accounts for thermal degradation of molten nylon 66
- Estimability analysis and MSE criterion determined that 43 out of 56 model parameters should be estimated to provide the best predictions
- Model describes the experimental polymer property data and gas evolution well
- Model will be useful for selecting temperature and moisture level during final stages of industrial nylon 66 production
- Model will be used to test spline-based advanced AML estimation techniques in future
 - Objective function accounts for model imperfections and dynamic disturbances
 - Many nuisance parameters due to B-spline coefficients
 - IPOPT and AMPL for parameter estimation

Other Uses of Estimability Analysis

- Tuning models for commercial processes
 - Start with model and parameters from lab-scale studies
 - Obtain new data at commercial scale
 - Use estimability analysis to decide which parameters to adjust so model matches commercial operation
- Sequential experimental design
 - Do estimability analysis by adding rows to Z for different proposed experiments
 - Determine additional experiments that will yield the smallest joint confidence regions for the parameters or the smallest variance for model predictions
 - $Z^T Z$ is a Fisher Information Matrix

Other Uses of Estimability Analysis

- Tuning models for commercial processes
 - Start with model and parameters from lab-scale studies
 - Obtain new data at commercial scale
 - Use estimability analysis to decide which parameters to adjust so model matches commercial operation
- Sequential experimental design
 - Do estimability analysis by adding rows to Z for different proposed experiments
 - Determine additional experiments that will yield the smallest joint confidence regions for the parameters or the smallest variance for model predictions
 - $Z^T Z$ is a Fisher Information Matrix
 - Future work: Select experiments to minimize MSE of predictions in desired operating range

Acknowledgments

- Collaborators
 - E. K. Marchildon
 - P.J. McLellan, D.A. Bacon, T.J. Harris
- Graduate Students and Postdocs
 - H. Karimi, M. Schaffer
 - B. Shaw, B. Kou, Z. Yao, S. Varziri, D. Thompson, S. Wu, K. McLean, J. Woloszyn
- Final Support from
 - MITACS, NSERC
 - DuPont, BASF, INEOS, Shell
 - Cybernetica, Hatch, Matrikon, SAS

Summary

- Chemical engineers develop fundamental models to optimize industrial processes
- Models are important for process design and selecting operating conditions
 - Physical understanding \Rightarrow many parameters
 - Parameter estimation is a difficult challenge
- New techniques for
 - Selecting which parameters to estimate
 - Designing additional experiments

Unimportant Parameters

Final Rank	Parameter Ranked List	Unit	Initial Guess	Scaling Factor
44	$[SB2]_{0run,1}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	25.6
45	$[SB2]_{0run,2}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	27
46	$[SB2]_{0run,4}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	31.5
47	$[SB2]_{0run,5}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	20.75
48	$[SB2]_{0run,6}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	31.2
49	$[A2]_{0run,4}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	31.5
50	$[A2]_{0run,6}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	31.2
51	ΔH_6	$\text{kJ}\cdot\text{mol}^{-1}$	41.85	837
52	$[SB3]_{0run,1}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	25.6
53	$[SB3]_{0run,2}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	27
54	$[SB3]_{0run,4}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	31.5
55	$[SB3]_{0run,5}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	20.75
56	$[SB3]_{0run,6}$	$\text{mol}\cdot\text{Mg}^{-1}$	0	31.2

67°C Data Only – Objective Function and Critical Ratio

Ranked List:

K_1

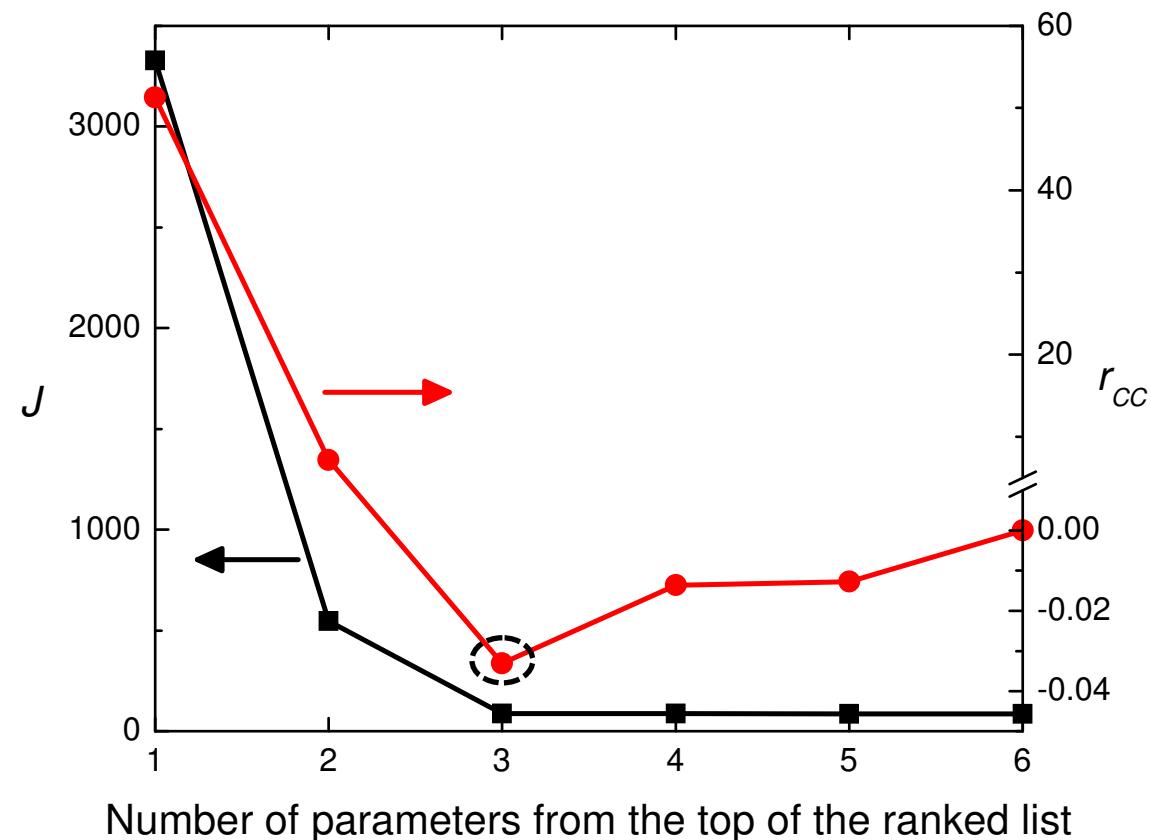
k_{20}

k_{10}

k_{-10}

K_2

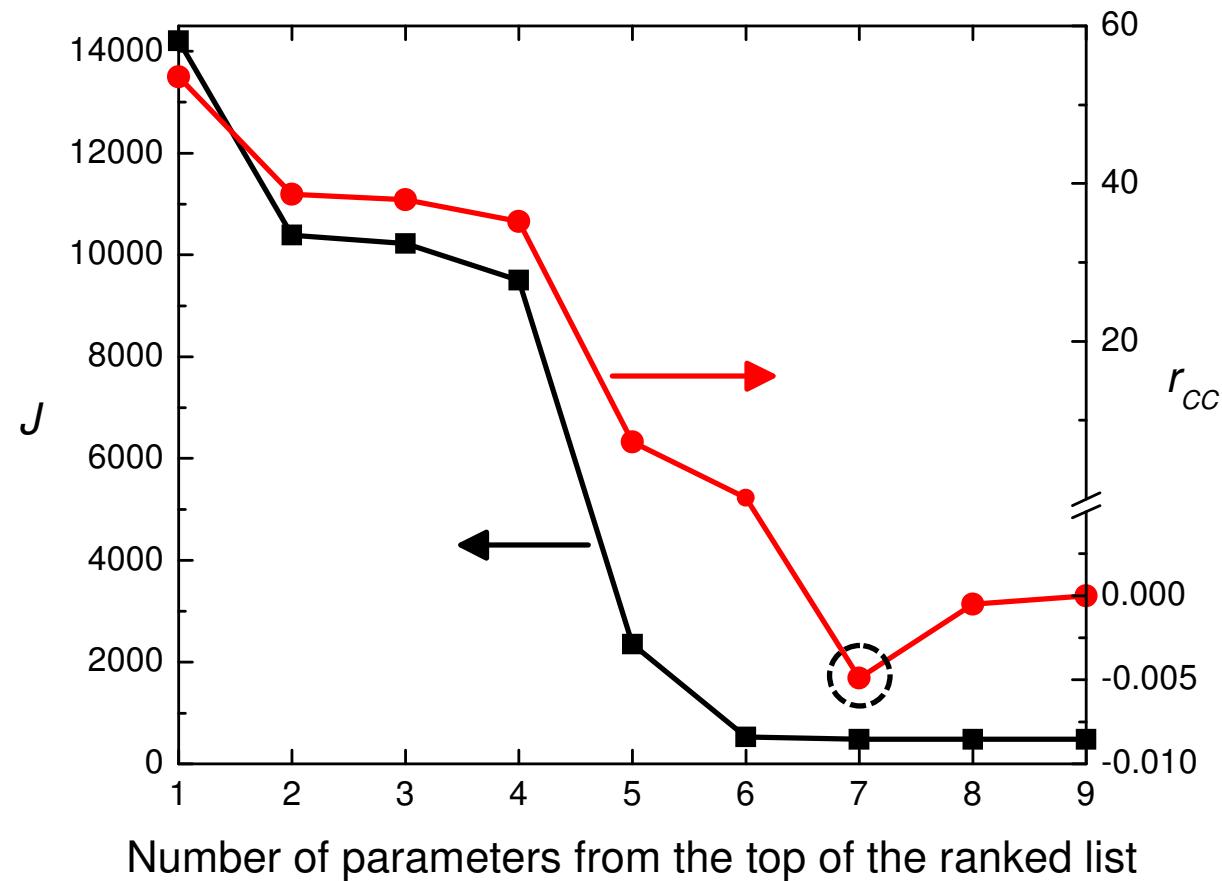
K_3



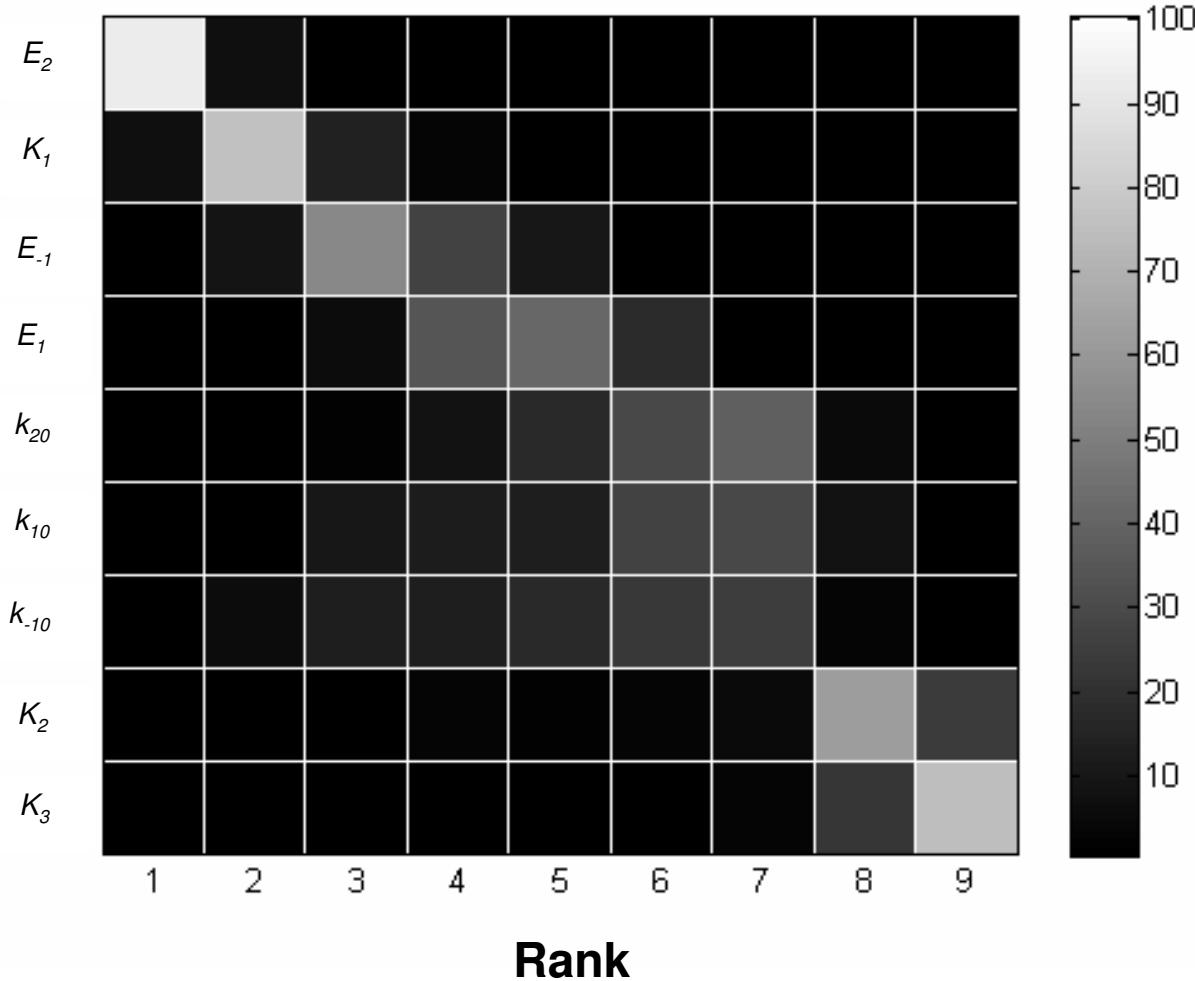
All Data - Objective Function and Critical Ratio

Ranked List:

E_2
 K_1
 E_{-1}
 E_1
 k_{20}
 k_{10}
 k_{-10}
 K_2
 K_3

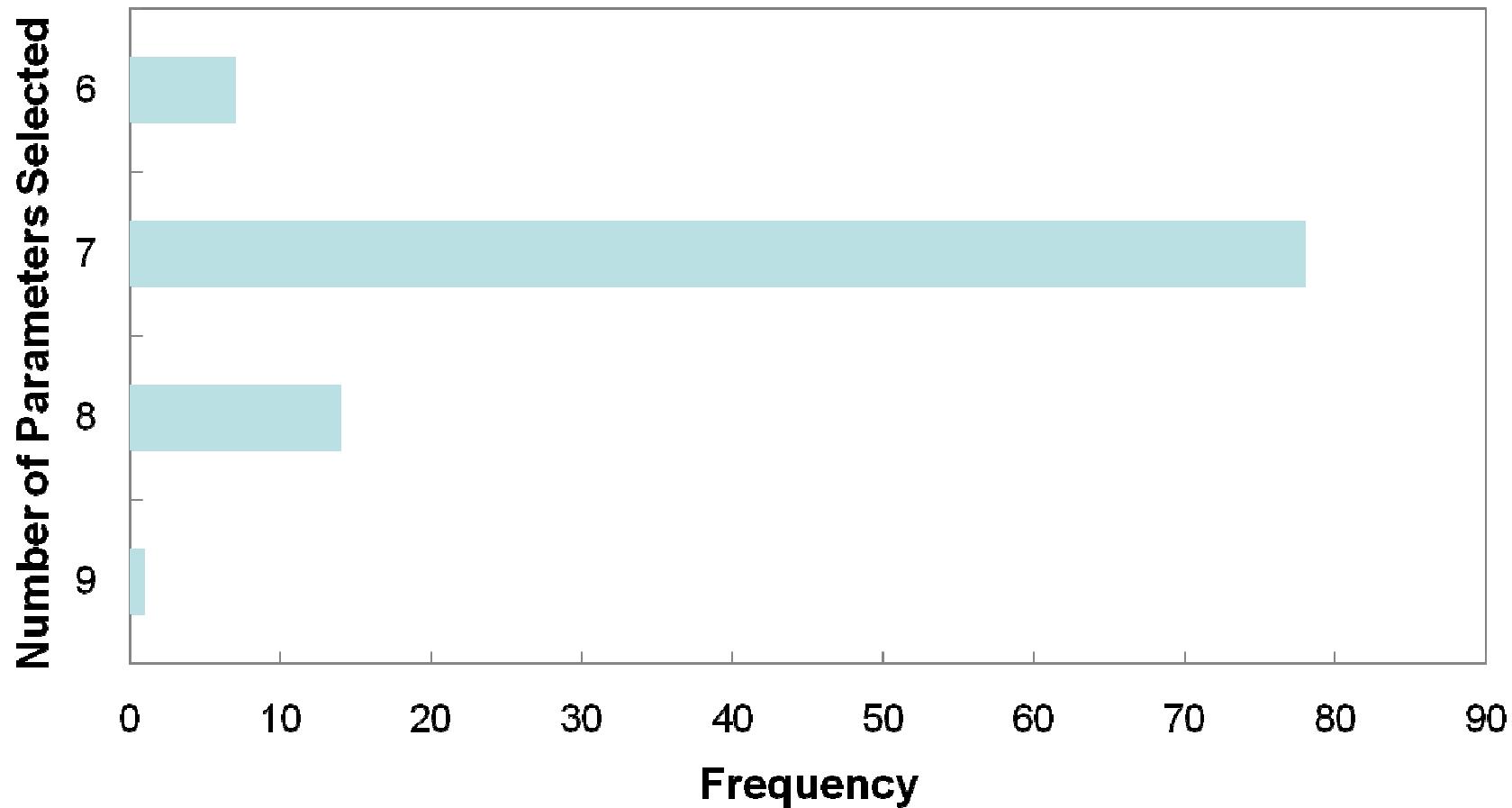


Frequency of Parameter Rankings Using All Data



Stortelder W., Ph.D. thesis at Centrum Wiskunde and Informatica, pp. 132-142 (1997).

Frequency of Subset Selection Using All Data



The same 7 parameters are selected most of the time

AML Estimation Technique for Stochastic DEs

$$\frac{dx}{dt} = f(x, u, \theta) + \eta(t), \quad x(0) = x_0$$
$$y_i = x_i + \varepsilon_i \quad (i = 1, \dots, n) \quad \varepsilon_i \sim N(0, \sigma_m^2)$$

- Two noise sources
 - Measurement noise
 - Stochastic process disturbances that can account for
 - Uncertainties in u
 - Unknown or unmeasured inputs
 - Structural imperfections

We can also estimate unknown or poorly-known initial conditions and incorporate nonstationary disturbances

AML Parameter Estimation Technique for Stochastic DEs

$$\frac{dx_{\sim}}{dt} = f(x_{\sim}, u, \theta) + \eta(t), \quad x_{\sim}(0) = x_0$$

$$y_i = x_{\sim i} + \varepsilon_i \quad (i = 1, \dots, n)$$

- Our approach:
 - Assume that the solution to the differential equations can be represented using B-splines or other basis functions:

$$x(t) \equiv x_{\sim}(t) = \sum_{i=1}^b \varphi_i(t) \beta_i$$

$\frac{dx_{\sim}(t)}{dt}$ is used to convert DEs into algebraic equations

Approximate Maximum Likelihood Estimation

- Assume the solution of the dynamic system can be well approximated by B-splines with unknown coefficients β
- We want to estimate the fundamental model parameters θ and the unknown spline coefficients β
 - Select $\hat{\theta}$ and $\hat{\beta}$ to minimize

$$J = \sum_{i=1}^n (y_i - x_{i\sim})^2 + \lambda \int \left(\frac{dx_{\sim}(t)}{dt} - f(x_{\sim}(t), \mathbf{u}(t), \boldsymbol{\theta}) \right)^2 dt$$

Apparatus

