Kinetic Model Parameter Estimation for Product Stability: *Non-uniform Finite Elements and Convexity Analysis*

Mark Daichendt, Lorenz Biegler
Carnegie Mellon University
Pittsburgh, PA 15217

Ben Weinstein, Robert Johnson, Steve Hodson
Procter & Gamble
West Chester, OH 45069

Fall EWO Meeting
21-22 September 2016

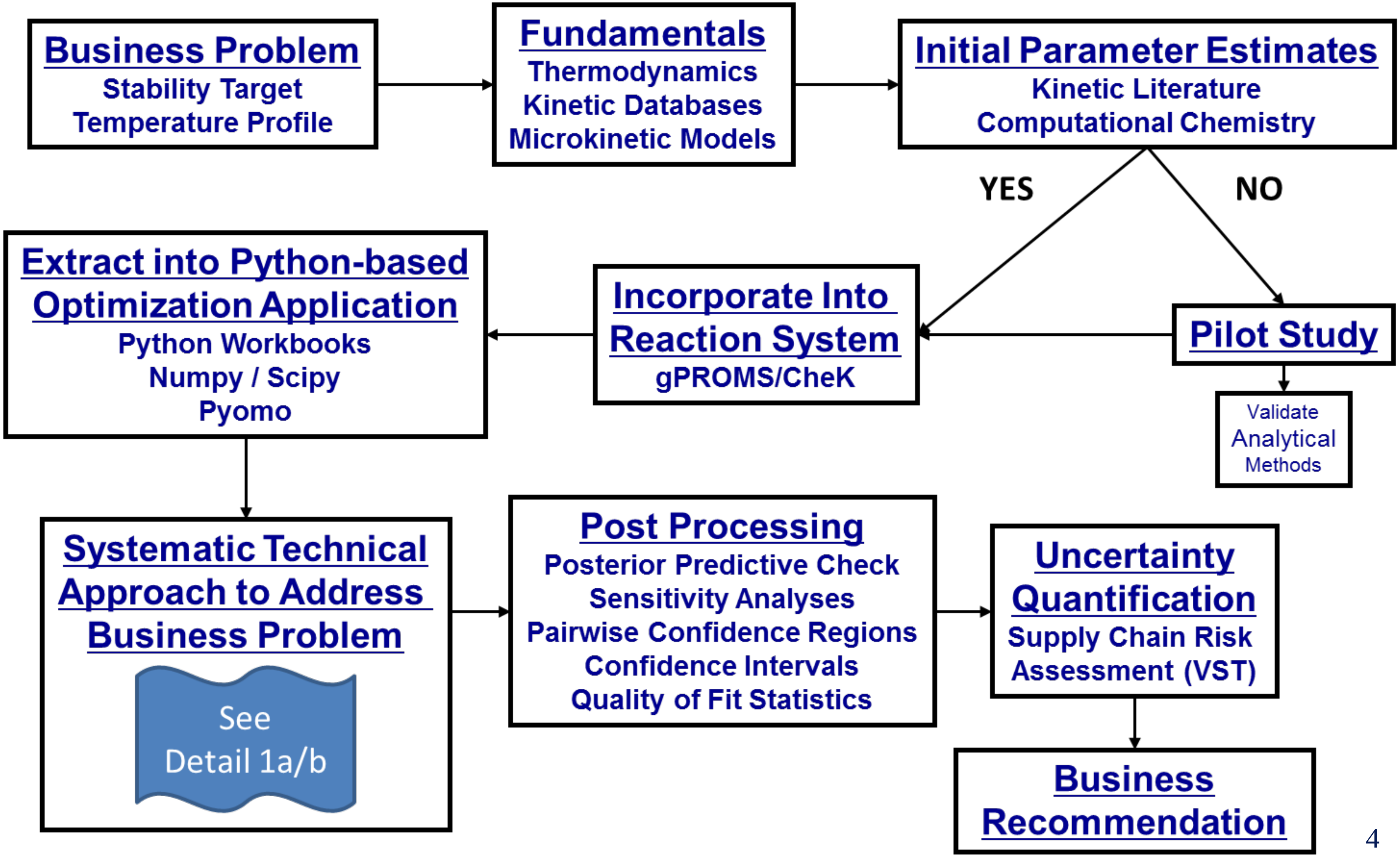
Motivation & Objective

- Consumer products may be subject to prolonged storage and transportation times
  - Ensuring Product Quality and Stability is a critical and time-consuming activity
  - Shelf-lives typically measured in years
- Need to understand impact of formula design on product performance and stability
  - Undesirable chemical reactions may lead to product degradation
  - Long time needed to collect data, which may have a low signal-to-noise ratio
- Develop a suite of modeling tools to support the development of new consumer products:
  - Simulate product chemistry
  - Design efficient experimental campaigns
  - Discriminate among alternative models and estimate model parameters
  - Understand the uncertainty associated with model predictions

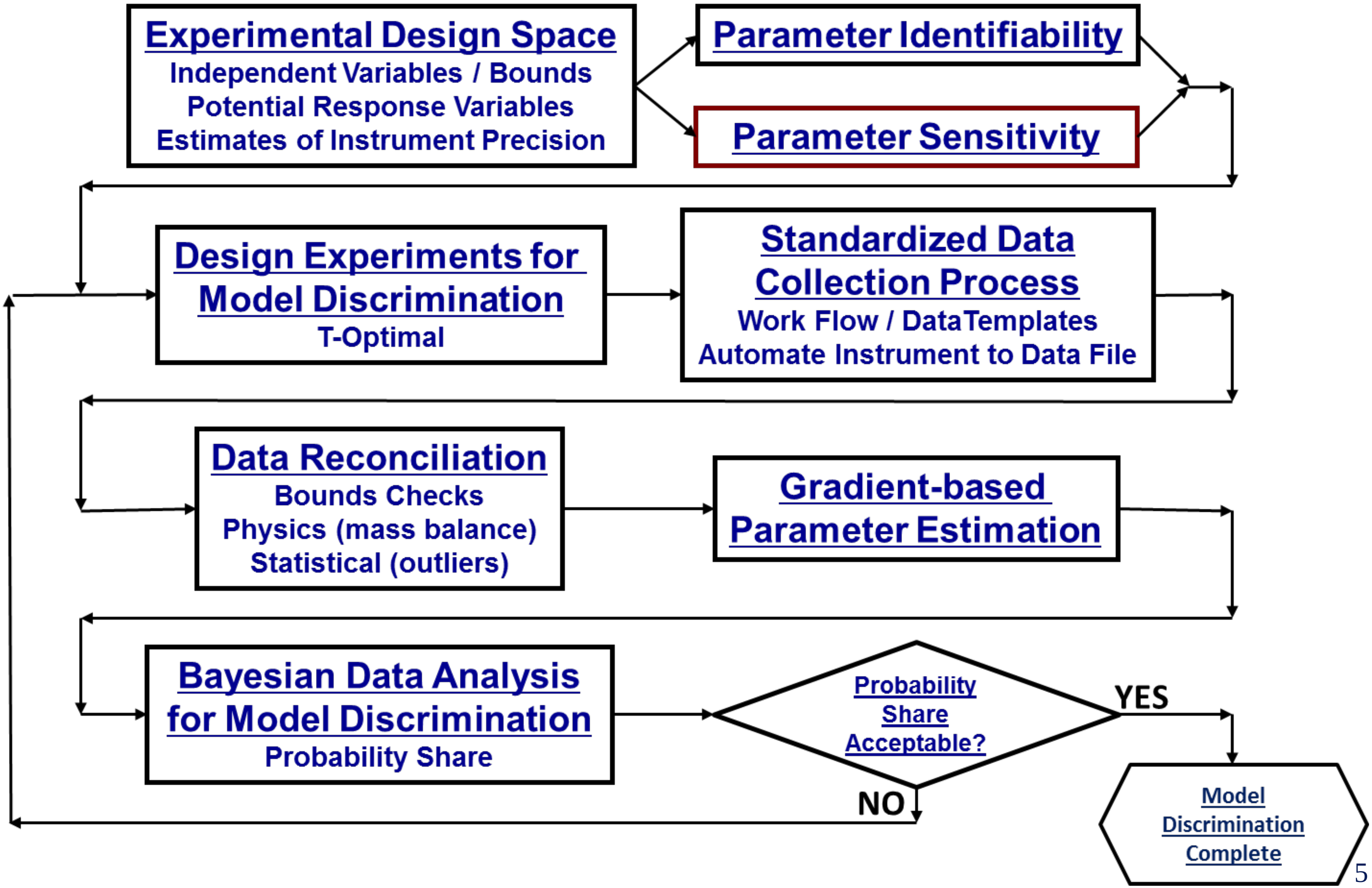
Framework

- CheK is an easy-to-use model library in gPROMS, developed and maintained by P&G to:
  - Encourage chemists to develop quantitative models of reaction networks
  - Allow for any combination of 8 kinetic models to form a complex reaction network
    - The reactor model is an isothermal batch reactor
    - For Product Stability applications, optimization for Parameter Estimation and Design of Experiments has been challenging in gPROMS
- The goal is to extend the current CheK capability by interfacing it with enhanced optimization tools and methods:
  - Integration using orthogonal collocation on finite elements
  - Simultaneous NLP solution using s-IPOPT
  - Open source modeling package PyOMO within Python platform
- The deliverable will be a flexible and robust application that can be maintained by P&G

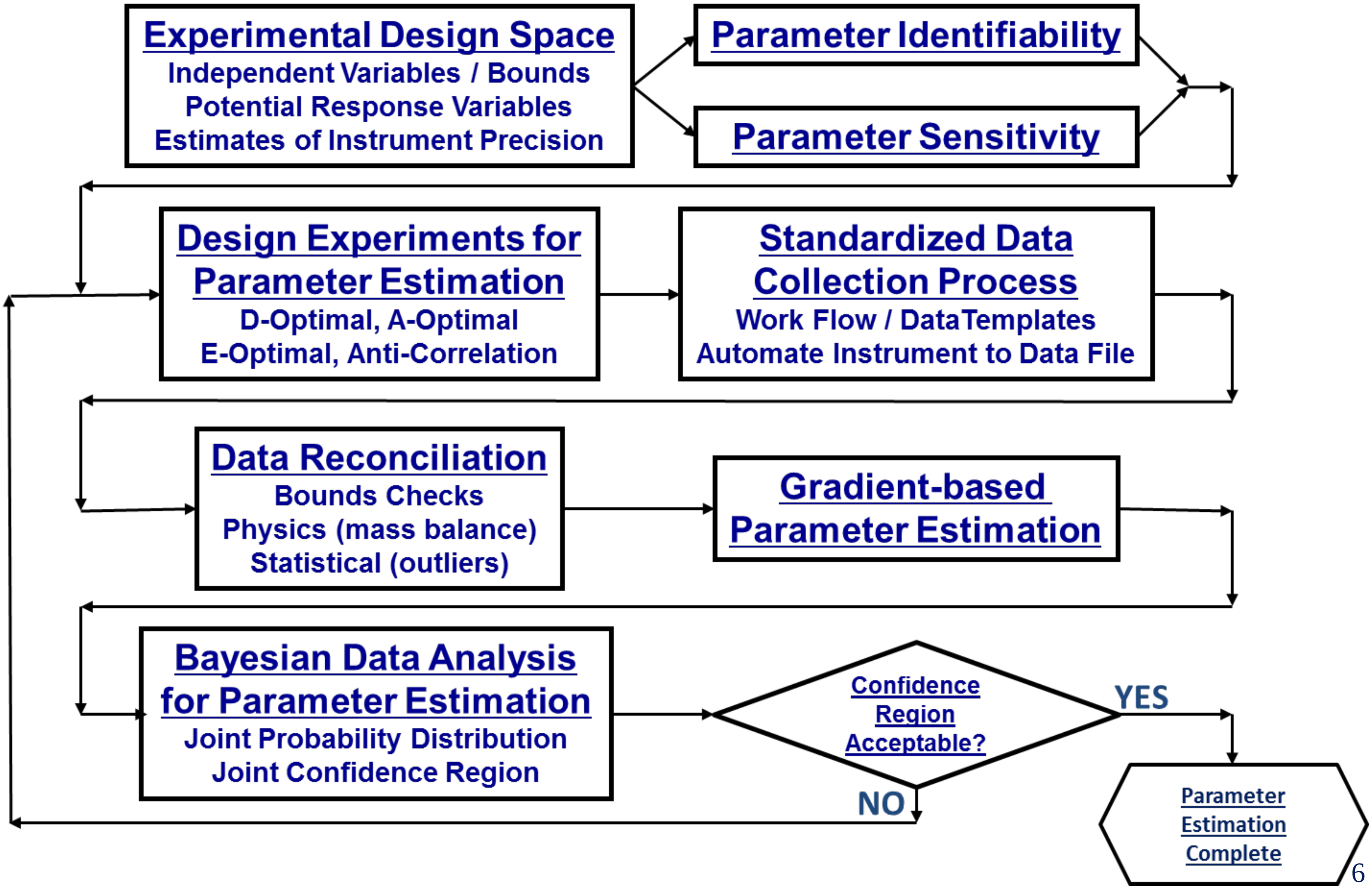
Full Product Life-Cycle Analysis



Detail 1a: Systematic Approach to Model Discrimination (if applicable)

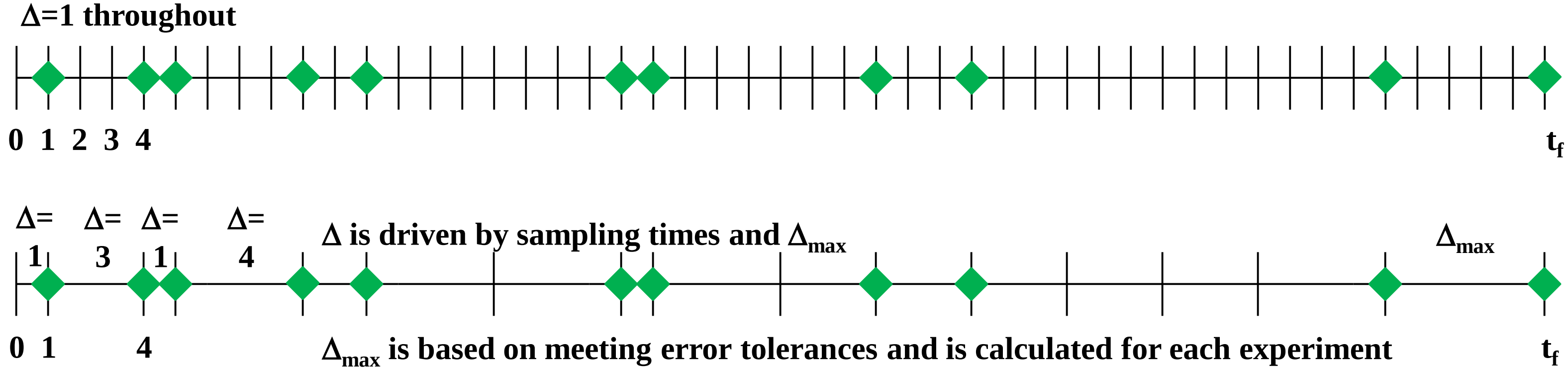


Detail 1b: Systematic Approach to Parameter Estimation



Implemented Non-uniform Finite Element Length

- Size implication of test problem
  - Experimental campaign length > 400 days
  - Fixed-interval duration = 1 day
  - Number of Experiments > 150
    - Potential number of sampling times > 60,000
    - Actual number of sampling times ~ 1000
- Length of Variable Finite Element Length:
  - Time between adjacent sampling times
    - Some are as close as 1 day
    - Others have spacing > 250 days
  - Maximum Allowable Global Error
    - Absolute
    - Relative



Carnegie Mellon

Residual-based and Global Error Estimate[1]

- Introduce a non-collocation point τ_{nc}

$$t_{nc} = t_{i-1} + h_i \tau_{nc} \qquad \tau_{nc} \in [0,1]$$

- Determine the residual-based error estimate $T_i(t)$ at τ_{nc} for the collocation polynomial of degree K

$$T_i(\tau_{nc}) = \frac{dz^K(\tau_{nc})}{d\tau} - h_i f\left(z^K(\tau_{nc}), y^K(\tau_{nc}), p\right)$$

- The global error is related to the residual-based error

$$\|e_i(t)\| \leq \bar{C} \|T_i(\tau_{nc})\|$$

- $\bar{C}$ is given by:

$$\bar{C} = \frac{1}{A} \int_0^{\tau_{nc}} \prod_{j=1}^{K} (s - \tau_j) ds$$

$$A = \prod_{j=1}^{K} (\tau_{nc} - \tau_j)$$

[1]Biegler, L.T., *Nonlinear Programming: Concepts, Algorithms, and Applications to Chemical Processes*, MOS-SIAM, Philadelphia (2010) p. 297

---- 7445 PARAMETER h_init h days (25 FE max) ---- 7458 PARAMETER time_init t days

ife	EX1_10C	EX1_30C	EX1_50C
1	1.000	1.000	1.000
2	25.000	18.000	18.000
3	8.000	7.000	7.000
4	6.000	8.000	8.000
5	7.000	6.000	6.000
6	11.000	7.000	7.000
7	32.000	11.000	11.000
8	64.000	32.000	32.000
9	27.000	63.000	10.000
10	216.646	28.000	53.000
11	12.354	159.975	28.000
12		69.025	86.879
13			102.827
14			39.294

ife	EX1_10C	EX1_30C	EX1_50C
2	1.000	1.000	1.000
3	26.000	19.000	19.000
4	34.000	26.000	26.000
5	40.000	34.000	34.000
6	47.000	40.000	40.000
7	58.000	47.000	47.000
8	90.000	58.000	58.000
9	154.000	90.000	90.000
10	181.000	153.000	100.000
11	397.646	181.000	153.000
12	410.000	340.975	181.000
13		410.000	267.879
14			370.706
15			410.000

ife	EX9_10C	EX9_40C	EX9_54C
1	1.000	1.000	1.000
2	9.000	9.000	18.313
3	7.000	7.000	8.687
4	11.000	11.000	34.626
5	32.000	32.000	46.222
6	64.000	10.000	59.529
7	14.000	54.000	74.238
8	12.000	14.000	86.977
9	207.028	13.000	80.409
10	52.972	107.717	
11		138.491	
12		12.792	

ife	EX9_10C	EX9_40C	EX9_54C
2	1.000	1.000	1.000
3	10.000	10.000	19.313
4	17.000	17.000	28.000
5	28.000	28.000	62.626
6	60.000	60.000	108.848
7	124.000	70.000	168.376
8	138.000	124.000	242.614
9	150.000	138.000	329.591
10	357.028	151.000	410.000
11	410.000	258.717	
12		397.208	
13		410.000	

Bayesian Estimation from Multiresponse Data With Missing Observations[2]

- Example problem has two blocks of experiments with "missing observations"
  - Two component concentrations are measured
    - Two "blocks" of experiments
      - Different sensors for same component measurement are treated as different responses
      - Within a block there are sampling times with:
        - Both component concentrations measured
        - Only one or the other component concentration measured
      - Within one of the blocks there are experiments where only one component concentration is measured
    - Depending which experiments are selected, we could have one, two, three, or four responses
    - Automatically determined

u	Y_{u1}	Y_{u2}	Y_{u3}	Y_{um}
1	+	+		
2	+			
3		+		
4	+	+		
5			+	+
6			+	
7			+	
n			+	

[2]Stewart, W.E. and Sørensen, J.P., "Bayesian Estimation of Common Parameters from Multiresponse Data with Missing Observations", *Technometrics*, **23** (1981) p. 131-141

Simultaneous Estimation of Model and Variance-Covariance Parameters (2x2 case)

- Generalized sum-squared error objective function

$$S(\boldsymbol{\psi}) = S(\boldsymbol{\theta}, \boldsymbol{\sigma}) = (m+1)\ln|\boldsymbol{\sigma}| + \sum_{u=1}^{n} \ln|\boldsymbol{\sigma}_u| + \sum_{u=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} \sigma_u^{ij}\left[Y_{ui} - f_{ui}(\boldsymbol{\theta})\right]\left[Y_{uj} - f_{uj}(\boldsymbol{\theta})\right]$$

- Relationship among $\boldsymbol{\sigma}$, $\boldsymbol{\sigma}_u$, and σ_u^{ij} for 2x2 case. This is generically calculated using Cholesky factorization for any *m*.

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \qquad \sigma_{11}, \sigma_{22} > 0 \qquad \sigma_{12} = \sigma_{21}$$

$$\boldsymbol{\sigma}_{u^{+0}} = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{\sigma}_{u^{++}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \qquad \boldsymbol{\sigma}_{u^{+0}} = \begin{bmatrix} 1 & 0 \\ 0 & \sigma_{22} \end{bmatrix}$$

$$\boldsymbol{\sigma}_{u^{+0}}^{-1} = \begin{bmatrix} 1/\sigma_{11} & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{\sigma}_{u^{++}}^{-1} = \frac{1}{|\boldsymbol{\sigma}|}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix} \qquad \boldsymbol{\sigma}_{u^{0+}}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1/\sigma_{22} \end{bmatrix}$$

Estimability: Convexity of Reduced Hessian

- Confirm that reduced Hessian is positive definite to
  - Assure convergence to a local solution
  - Determine confidence limits (parameter is estimable)
- Generalized sum-squared error introduces additional nonconvexities (compared to $\boldsymbol{\sigma}$ known and fixed *a priori*)
  - Proposed sequential optimization
    - $\boldsymbol{\sigma}$ for fixed $\boldsymbol{\theta}$
    - $\boldsymbol{\theta}$ for fixed $\boldsymbol{\sigma}$
  - Simultaneously optimize $\boldsymbol{\sigma}$ and $\boldsymbol{\theta}$ when nearly converged
- Combine the two ideas:
  - Optimize over the full set of $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$
  - Analyze the reduced Hessian to determine if positive definite (PD)
    - If not, determine the largest subset of parameters for which reduced Hessian is PD
      - If more than one combination, chose smallest $\lambda_{max}^{subset}/\lambda_{min}^{subset}$
      - Continue optimization over subset $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$, repeating if necessary

Algorithm

$\psi_\ell^k = \theta_\ell^k \cup \sigma_\ell^k$

$k = 0$

$\ell = 0$

Initial Model Parameters θ_0^0

Initialize Variance Covariance Parameters σ_0^0 for fixed θ_0^0
$\min_{\sigma} S(\theta, \sigma) = S(\psi)$

k = 1

Optimize Model Parameters θ_ℓ^k for fixed σ_ℓ^{k-1}
$\min_{\theta} S(\theta, \sigma)$

Optimize Variance Covariance Parameters σ_ℓ^k for fixed θ_ℓ^k
$\min_{\sigma} S(\theta, \sigma)$

$\|\psi_\ell^k - \psi_\ell^{k-1}\| / |\psi_\ell^k| \leq \varepsilon^\psi$

NO — **k = k+1**

YES

Simultaneously optimize θ_ℓ^{k+1} and σ_ℓ^{k+1}
$\min_{\psi} S(\psi)$

Is Reduced Hessian Positive Definite?

NO — $\ell = \ell + 1$

Find largest subset(s) of $\theta_\ell^0 \in \theta_{\ell-1}^{k+1}$ and $\sigma_\ell^0 \in \sigma_{\ell-1}^{k+1}$ For PD reduced Hessian

If multiple subsets exist
$\psi_\ell^0 = \psi_{\ell'}^0 \Big| \min_{\ell'}\left(\lambda_{\ell'}^{max} / \lambda_{\ell'}^{min}\right)$

k = 1

YES

Stop
Determine Confidence Intervals

Note, for results presentation on next slide, the Sequential-Simultaneous approach is shown for the 1st iteration *for purposes of illustration*.

For this example problem, directly applying the simultaneous approach gave an initial reduced Hessian with fewer negative eigenvalues, so the simultaneous approach was applied to generate the reduced Hessians.

At this stage, however, we can't generalize whether Seq-Sim or direct Sim will be better all problems.

Summary

- Implemented
  - Non-uniform finite elements
    - Balances error tolerances and matching sampling times
  - Parameter estimation for multiresponse data with missing observations and unknown variance/covariance
  - Sequential-simultaneous solution approach to deal with potential nonconvexities
    - Observed sequential approach's convergence to optimal solution
    - For example problem, direct simultaneous approach is adequate.
  - Convexity analysis of reduced Hessian
    - Optimize on subset (if necessary) of estimable parameters
- Next step
  - Integrate prior work developed for Design of Experiments for Parameter Estimation
  - Port capability to PyOMO, leveraging s-IPOPT

1st Iteration ($\ell = 0$) Seq-Sim followed by all Sim-Only Iterations ($\ell = 0, 1, 2$)

InitVC - Seq - Sim	k=0a, l=0	k=0b, l=0	k=1a, l=0	k=1b, l=0	k=2a, l=0	k=2b, l=0	k=3, l=0	...	Sim, l=0	...	Sim, l=1	...	Sim, l=2
Variable	Initial	OptVC0	OptMP1	OptVC1	OptMP2	OptVC2	OptALL		OptALL		3MP 2VC		2MP 3VC
ln_Keq	-3.000		-5.351		-6.090		-8.105		-8.105		-8.105		-8.105
Eact	50000		60235		56909		59304		59304		59304		59304
ln_k_f	-12.000		-12.375		-12.610		-12.422		-12.422		-12.422		-12.422
ln_k	-12.000		-38.369		-38.369		-38.368		-36.530		-36.536		-36.530
vc_11	1.000000	0.009774		0.003456		0.001453	0.000341		0.000341		0.000341		0.000341
vc_22	1.000000	0.463976		0.369184		0.384729	0.439630		0.439630		0.439630		0.439630
vc_21	0.000000	0.056194		0.027466		0.016101	-0.001719		-0.001719		-0.001719		-0.001719
sse	290.31		218.35		217.37		216.81		216.81		216.81		216.81
(m+1)*ln(det(VC))	0.00	-19.76		-22.68		-24.34	-26.48		-26.48		-26.48		-26.48
Sum(u, ln(det(VC(u))))	0.00	-3066.12		-3629.77		-3979.72	-4494.27		-4494.27		-4494.27		-4494.27
Sum(uij, sig(uij)*e(ui)*e(uj))	290.31	1006.00	690.02	1006.00	850.47	1006.00	1006.00		1006.00		1006.00		1006.00
S(theta,sigma)	290.31	-2079.88	-2395.86	-2646.45	-2801.97	-2998.06	-3514.75		-3514.75		-3514.75		-3514.75

Convexity Analysis for Reduced Hessians

Reduced Hessian for Four Model Parameters and Three Variance-Covariance Parameters

rH7	ln_K_eq	Eact_rr	ln_k_rr	ln_k_sr	s_vc_11	s_vc_22	s_vc_21	-	EvalrH7	EvalrHmp	EvalrHvc
ln_K_eq	1.45E+06	-1.11E+02	-4.92E+05	-8.93E+04	-6.60E+06	-4.59E+06	-6.63E+06		-1.32E+08	-7.96E-02	
Eact_rr	-1.11E+02	3.41E-02	-7.55E+01	-1.37E+01	-1.01E+03	-7.04E+02	-1.01E+03		1.11E-02	2.03E+04	
ln_k_rr	-4.92E+05	-7.55E+01	6.67E+05	-6.07E+04	-4.48E+06	-3.11E+06	-4.48E+06		3.62E+04	4.52E+05	
ln_k_sr	-8.93E+04	-1.37E+01	-6.07E+04	4.33E+04	-8.18E+05	-5.65E+05	-8.12E+05		6.09E+05	1.69E+06	
s_vc_11	-6.60E+06	-1.01E+03	-4.48E+06	-8.18E+05	1.02E+10	-4.27E+08	1.62E+09		1.76E+06		-3.24E+08
s_vc_22	-4.59E+06	-7.04E+02	-3.11E+06	-5.65E+05	-4.27E+08	5.93E+08	-4.19E+07		5.75E+08		4.50E+05
s_vc_21	-6.63E+06	-1.01E+03	-4.48E+06	-8.12E+05	1.62E+09	-4.19E+07	1.25E+08		1.05E+10		1.43E+10

B1	Eact_rr	ln_k_rr	ln_k_sr		EvalB1
Eact_rr	3.41E-02	-7.55E+01	-1.37E+01		1.44E-02
ln_k_rr	-7.55E+01	6.67E+05	-6.07E+04		3.75E+04
ln_k_sr	-1.37E+01	-6.07E+04	4.33E+04		6.73E+05
				EvalRatio	4.68E+07

B2	ln_K_eq	ln_k_rr	ln_k_sr		EvalB2
ln_K_eq	1.45E+06	-4.92E+05	-8.93E+04		2.03E+04
ln_k_rr	-4.92E+05	6.67E+05	-6.07E+04		4.52E+05
ln_k_sr	-8.93E+04	-6.07E+04	4.33E+04		1.69E+06
				EvalRatio	8.34E+01

B3	ln_K_eq	Eact_rr	ln_k_sr		EvalB3
ln_K_eq	1.45E+06	-1.11E+02	-8.93E+04		1.44E-02
Eact_rr	-1.11E+02	3.41E-02	-1.37E+01		3.77E+04
ln_k_sr	-8.93E+04	-1.37E+01	4.33E+04		1.46E+06
				EvalRatio	1.01E+08

B4	ln_K_eq	Eact_rr	ln_k_rr		EvalB4
ln_K_eq	1.45E+06	-1.11E+02	-4.92E+05		-3.73E-05
Eact_rr	-1.11E+02	3.41E-02	-7.55E+01		4.30E+05
ln_k_rr	-4.92E+05	-7.55E+01	6.67E+05		1.69E+06

rHvcii	s_vc_11	s_vc_22			EvalrHv0
s_vc_11	1.02E+10	-4.27E+08			5.74E+08
s_vc_22	-4.27E+08	5.93E+08			1.02E+10

Reduced Hessian for Three Model Parameters and Two Variance-Covariance Parameters

rH5	ln_K_eq	ln_k_rr	ln_k_sr	s_vc_11	s_vc_22		EvalrH5	EvalrH3mp	evalrH2vc
ln_K_eq	7.91E+02	2.35E+02	8.56E+04	-5.40E+03	1.25E+03		-1.26E+05	-1.26E+05	
ln_k_rr	2.35E+02	8.64E+02	5.82E+04	-5.19E+03	1.06E+03		6.23E+02	6.23E+02	
ln_k_sr	8.56E+04	5.82E+04	-4.15E+04	7.78E+05	5.42E+05		8.52E+04	8.55E+04	
s_vc_11	-5.40E+03	-5.19E+03	7.78E+05	1.01E+10	-3.85E+08		5.20E+08		5.20E+08
s_vc_22	1.25E+03	1.06E+03	5.42E+05	-3.85E+08	5.35E+08		1.01E+10		1.01E+10

Subset	ln_K_eq	ln_k_rr		EvalrH2mp
ln_K_eq	7.91E+02	2.35E+02		5.90E+02
ln_k_rr	2.35E+02	8.64E+02		1.06E+03

Reduced Hessian for Two Model Parameters and Two Variance-Covariance Parameters / Inverse

rH4	ln_K_eq	ln_k_rr	s_vc_11	s_vc_22		EvalrH4
ln_K_eq	5.93E+02	2.01E+02	-4.95E+03	3.75E+03		3.67E+02
ln_k_rr	2.01E+02	5.46E+02	-5.19E+03	1.70E+03		7.72E+02
s_vc_11	-4.95E+03	-5.19E+03	1.01E+10	-3.85E+08		5.20E+08
s_vc_22	3.75E+03	1.70E+03	-3.85E+08	5.35E+08		1.01E+10

InvrH4	ln_K_eq	ln_k_rr	s_vc_11	s_vc_22
ln_K_eq	1.93E-03	-7.10E-04	1.55E-10	-1.11E-08
ln_k_rr	-7.10E-04	2.09E-03	6.83E-10	-1.17E-09
s_vc_11	1.55E-10	6.83E-10	1.02E-10	7.32E-11
s_vc_22	-1.11E-08	-1.17E-09	7.32E-11	1.92E-09

Results of Test Problem

$$A + B \rightleftharpoons C$$

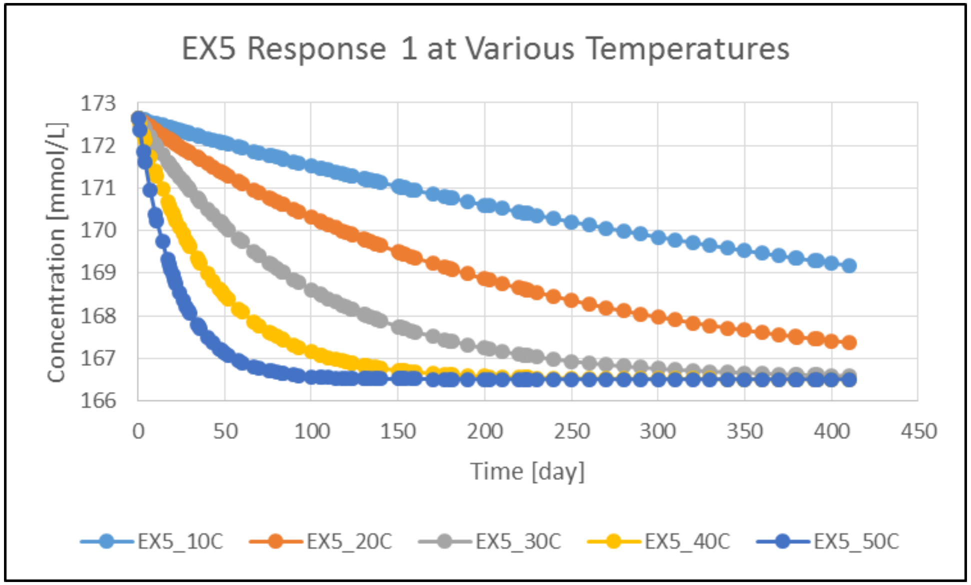
$$A + D \rightarrow E$$

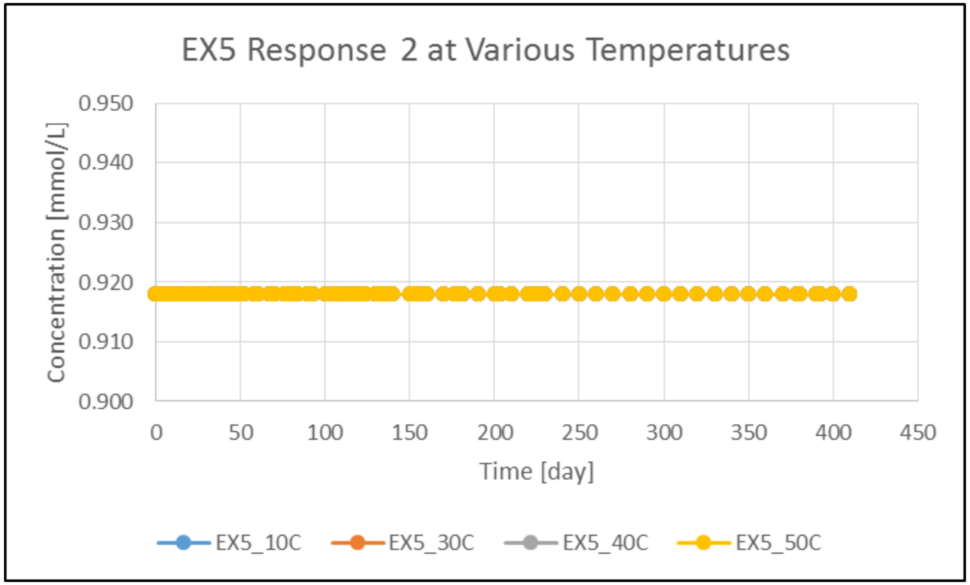
Optimization is performed in transformed variable space (e.g. ln_K_eq, ln_k_rr, ln_k_sr, ln_conc, but Eact_rr).

The 95% confidence intervals shown here are transformed into "physical" variable space and are not symmetric.

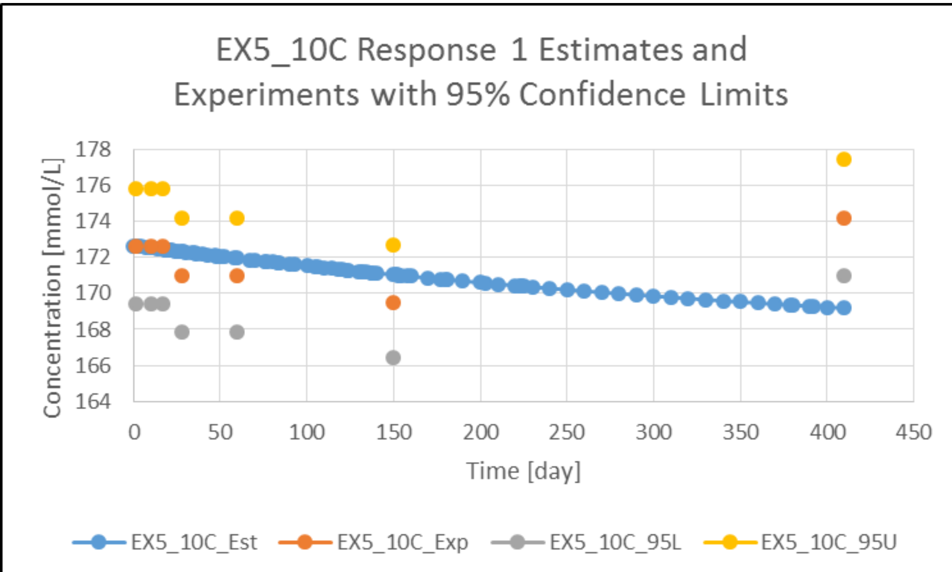
Note that the confidence intervals for concentration are relative to the measured concentration.

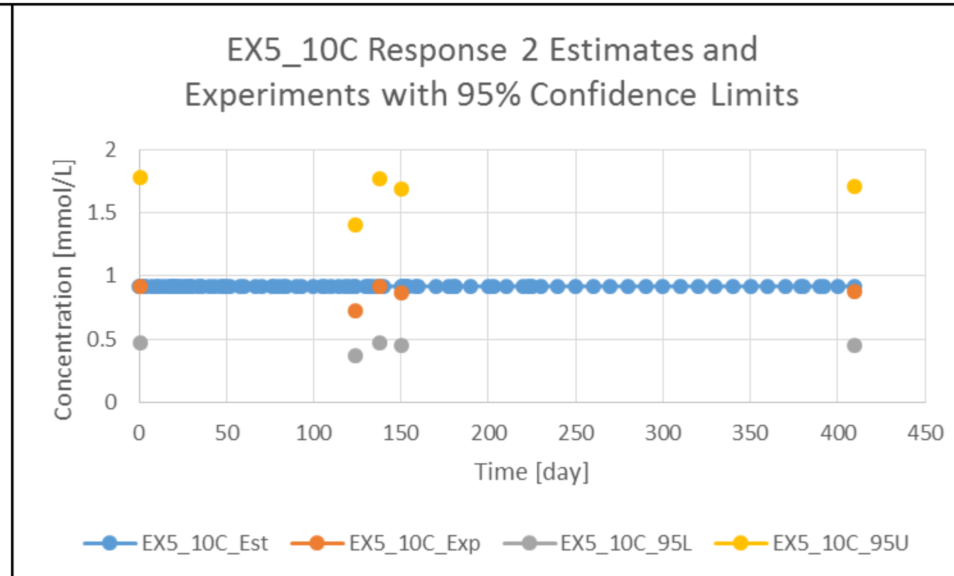
	Optimum	Lower	Upper
Keq	3.02E-04	2.77E-04	3.29E-04
Eact	59304	NA	NA
k_rr	4.03E-06	3.68E-06	4.41E-06
k_sr	1.37E-16	NA	NA
Sqrt(sig11)/c1	0.01846	0.01829	0.01863
Sqrt(sig22)/c2	0.7127	0.4847	0.9407

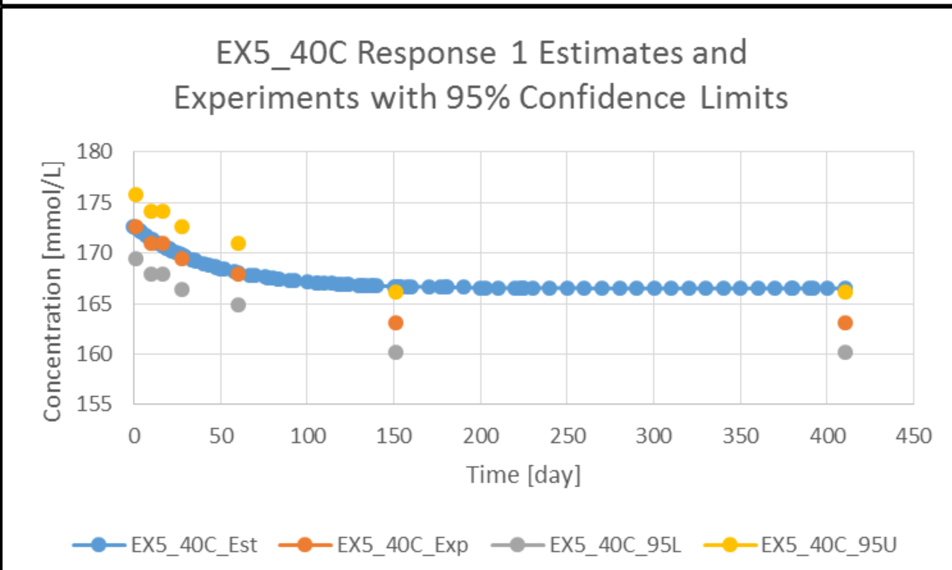


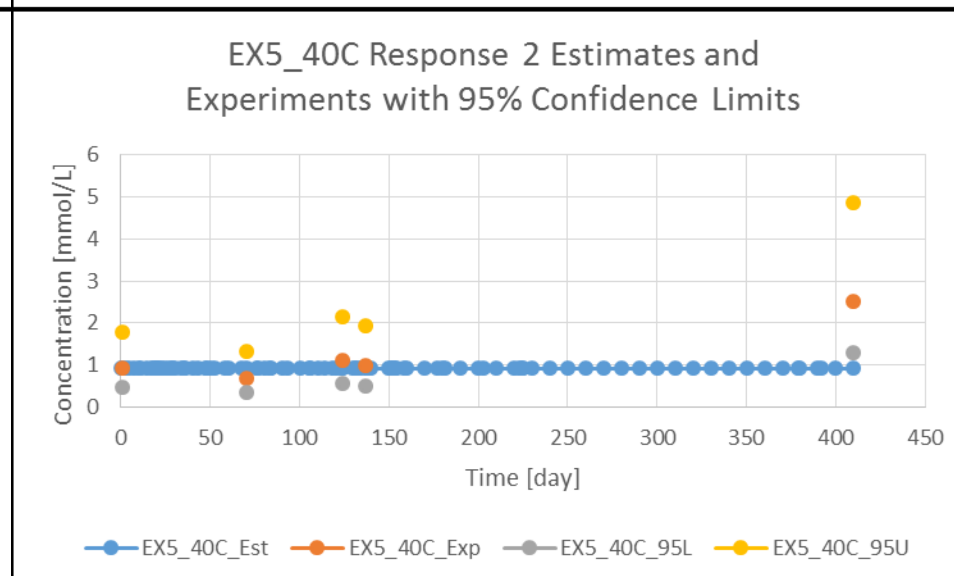


Results of Test Problem EX5









Results of Test Problem EX11