

Initial Design of Experiments to Determine Feasibility to Meet Parameter Estimation Objectives Prior to Data Collection

Mark Daichendt, Pooja Bhalode, Lorenz Biegler

Carnegie Mellon University

Pittsburgh, PA 15217

Ben Weinstein, Robert Johnson, Steve Hodson

Chemical Systems Modeling & Simulation

Procter & Gamble

West Chester, OH 45069

Spring EWO Meeting

14-15 March 2017

- The Modeling & Simulation Group frequently is first engaged to perform a parameter estimation analysis *only after* all the experimental data has been collected (“inherited data”)
 - ◆ Range of experimental conditions may have been insufficient
 - Limited to the *Formulation Design Space*
 - ◆ Some of the responses may be noisier than others
 - *Low Signal-to-Noise Ratio*
 - ◆ Potential responses may not have been measured in order to reduce costs
 - *Information content* may not have been considered
 - ◆ Preferable for Business to *first consult* with the Modeling & Simulation Group to determine:
 - How the *limits of the Experimental Design Space* impact
 - The *precision* of the parameter estimates
 - The *uncertainty* in the prediction
 - How *different responses* contribute to the *business objective*

- **Given** a dynamic model of a reaction network generated from the CheK Library in gPROMS
- **Find** an optimal set of experimental design points
 - ◆ Initial concentrations, temperature
 - ◆ To be *run in parallel due to the long campaign length*
- That **optimize** a selected criterion of the Information Matrix
 - ◆ D/E/A-optimal Designs (maximization)
 - ◆ Covariance (minimize absolute value)
- **While** not adversely impacting the other criteria (parallel design)
 - ◆ Avoid designing experiments at cross purposes
 - E.g. improving one covariant element while making others worse
 - ◆ Calculate all criteria simultaneously, *one optimized, others constrained*
 - Underlying Pareto-optimal aspect to the approach
- In a **“Robust” manner** and **provide “Recourse”**
 - ◆ For future sampling times and selected responses as data is collected

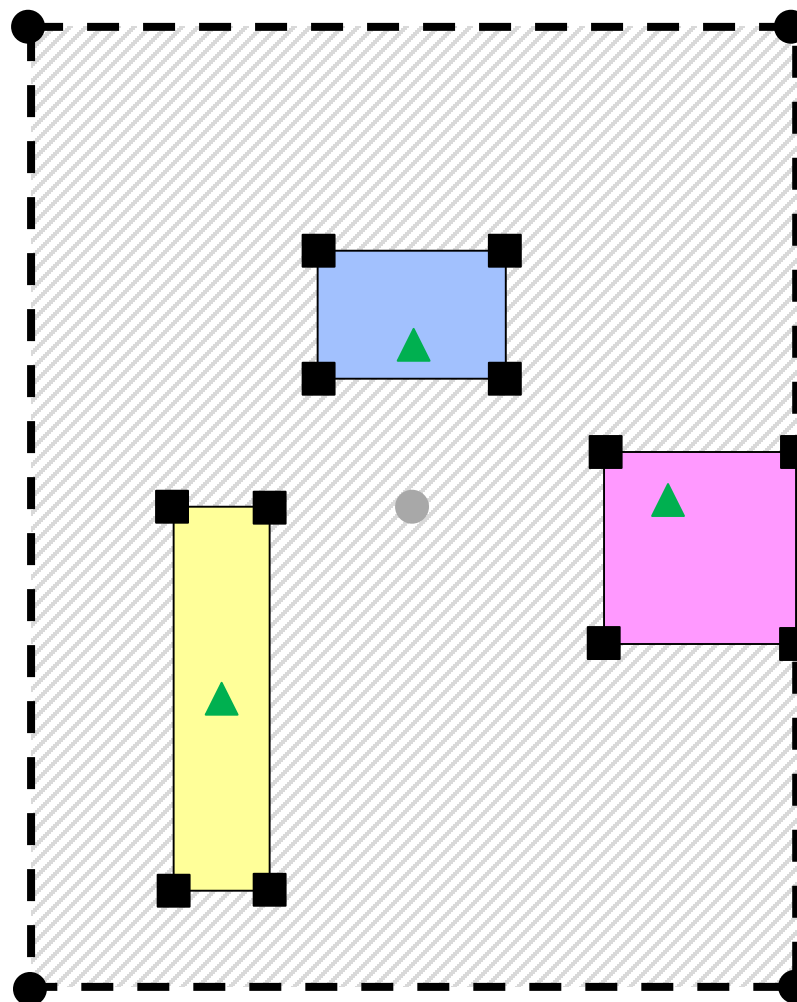
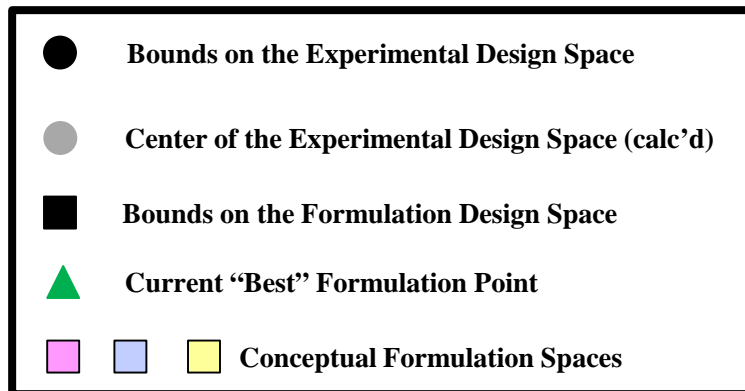
- Initialize multiple (>100) experiments one temporal finite element (FE) at a time while determining optimal finite element length
 - ◆ Uses current parameter estimate and experimental conditions. NEED TO:
 - Perform for worst-case initial conditions for Design of Experiments
 - Perform for worst-case parameter estimates for Robust Designs
 - **Solution:** Find combination that produces the largest absolute value of any component's concentration rate of change (smallest FE size)
- *Generate* the information matrix (IM) for the next-best design point (to be extended to multiple design points) *and optimize*
 - ◆ Integrating the composition profiles along with their sensitivities
 - ☹ Optimizer does not move away from the initial guess
 - Multiple local optima
 - “Egg carton” surface

- Highly nonlinear high-fidelity adsorption models optimized using a low-order model (LOM)
 - ◆ Choose starting initial condition estimates based on factorial Design of Experiments approaches, denoted by the set U
 - Initial guesses for parameters, $\theta^{k=0}$
 - Initial guesses for data variance-covariance matrix $\sigma^{k=0}$
 - Instrument precision based on manufacturer’s specifications
 - In-house determination (e.g. following IUPAC guidelines)
 - ◆ Demonstrated capability to initialize multiple experiments simultaneously
 - High-fidelity model results
 - ◆ Determine which combination of experimental points $u \in \hat{U} \subseteq U$ leads to a good $IM^\Sigma = \sum_{u \in \hat{U} \subseteq U} IM_u$, where IM_u is the IM for experiment u
 - ◆ Fit points to a quadratic response surface model (LOM “~”)
 - Predicts how the elements of the \tilde{IM}^Σ change with initial conditions \tilde{IM}_u
 - ◆ Optimize the initial conditions using the LOM, denoted by the set U^*
 - ◆ Update $IM^{\Sigma*} = \sum_{u \in \hat{U}^* \subseteq U^*} IM_u^*$ using the high-fidelity model

- Business/Technical partnership to develop a good experimental campaign
 - ◆ Size of the design space is important:
 - Effects the precision of the parameter estimates
 - Experiments can be run more quickly at more aggressive conditions
 - Early initial data → early parameter estimates
 - Recourse for future sampling times and measurements
 - Design new experiments for remaining horizon
 - Early adjustments to formulation
 - ◆ Which responses are measured brings different information content
 - Understand from the onset how including or excluding particular responses affects the information matrix
 - Impacts the rank of the matrix
 - Impacts the precision of the parameter estimate

Design Spaces

- Distinguish between
 - ◆ *Experimental Design Space*
 - ◆ *Formulation Design Space*
- Experimental Design Space depends on
 - ◆ Equipment limitations
 - ◆ Short-term thermal degradation, etc.
- Desire largest possible Experimental Design Space
 - ◆ (Bard) Optimal Design Points usually have at least one variable at a bound
- Variance-Covariance matrices of both the *parameters* and the *data* impact the prediction uncertainty
- Sample data within the Formulation Design Space to
 - ◆ *Assess prediction accuracy*
 - Hold out data
 - ◆ *Fine tune the parameter estimates*



P&G Selected Interface Menus Guide Consult



Design of Experiments with Parameter Estimation

Explorer Bar
Input
Parameters to be estimated

Experimental Design
Initial Exp. Design Points
Responses to be measured

Campaign Structure
Campaign Cutoff Criteria
Robust Design Analysis

Output
Simulate Experiments
Calculate sensitivity
Calculate Info Matrix
Generate best Info Matrix
Calculate Low Order Model
Optimize IC for IM/LOW
Calculate accurate IM based on LDM/IC
Robust Design Output

Main Window

Selection of Parameters to be estimated

Select all the parameters to be estimated:

1. Reversible reaction001

$$A + B \rightleftharpoons D$$

K_eq
 ΔHrxn
 k_f
 Eact_f

2. Simple reaction001

$$A + C \rightarrow E$$

k
 Eact

Order of reaction: Reactant A
 Reactant C

Experimental Design

Explorer Bar
Input
Parameters to be estimated

Experimental Design
Initial Exp. Design Points
Responses to be measured

Campaign Structure
Campaign Cutoff Criteria
Robust Design Analysis

Output
Simulate Experiments
Calculate sensitivity
Calculate Info Matrix
Generate best Info Matrix
Calculate Low Order Model
Optimize IC for IM/LOW
Calculate accurate IM based on LDM/IC
Robust Design Output

Main Window

1. What is the available Design Space?
(Variables to be taken into consideration)

Concentration
 Temperature

Experimental Range
For Parameter Estimation

	Average	Lowest	Highest	Units
<input checked="" type="checkbox"/> A	150	100	200	μmol/L
<input checked="" type="checkbox"/> B	200	150	250	μmol/L
<input checked="" type="checkbox"/> C	30	10	50	μmol/L

Formulation Range
For Prediction Accuracy

	Anticipated	Lowest	Highest	Units
<input checked="" type="checkbox"/> A	125	110	135	μmol/L
<input checked="" type="checkbox"/> B	180	160	210	μmol/L
<input checked="" type="checkbox"/> C	20	15	30	μmol/L

Design of Experiments with Parameter Estimation

Explorer Bar
Input
Parameters to be estimated

Experimental Design
Initial Exp. Design Points
Responses to be measured

Campaign Structure
Campaign Cutoff Criteria
Robust Design Analysis

Output
Simulate Experiments
Calculate sensitivity
Calculate Info Matrix
Generate best Info Matrix
Calculate Low Order Model
Optimize IC for IM/LOW
Calculate accurate IM based on LDM/IC
Robust Design Output

Main Window

Measurement of responses
Concentration values

Select the reactants for measurement:

A
 B
 C

Analytical Instrumentation Precision:
+/- (95% confidence limits)

1. Reactant A:
 Percentage of measured value (Relative):
 Concentration value (Absolute): %

2. Reactant B:
 Percentage of measured value (Relative):
 Concentration value (Absolute): mmol/L

3. Reactant C:
 Percentage of measured value (Relative):
 Concentration value (Absolute):

Design of Experiments with Parameter Estimation

Explorer Bar
Input
Parameters to be estimated

Experimental Design
Initial Exp. Design Points
Responses to be measured

Campaign Structure
Campaign Cutoff Criteria
Robust Design Analysis

Output
Simulate Experiments
Calculate sensitivity
Calculate Info Matrix
Generate best Info Matrix
Calculate Low Order Model
Optimize IC for IM/LOW
Calculate accurate IM based on LDM/IC
Robust Design Output

Main Window

Campaign Cutoff Criteria

Enter the maximum number of experiments that can be performed:
Lower limit = Number of parameters to be estimated.

Campaign optimality objective:

A-Optimality
 D-Optimality
 E-Optimality
 Parameter Co-variance

Convergence Criteria (Relative change in the optimality criteria):
Minimum relative change in optimized criteria to have an additional experiment: D-optimality:

Note: Resource/ Cost associated with each experiments needs to be taken into consideration.

Calculating Info Matrix for each experiment

Select the experiment to display the information matrix:

Info Matrix (IM) for Experiment 1

	θ ₁	θ ₂	θ ₃	θ ₄	
θ ₁	0	0	0	0	K_eq
θ ₂	0	0	0	0	k_f
θ ₃	0	0	0	0	Eact_f
θ ₄	0	0	0	0	k

- ◆ Large-scale optimization capability of IPOPT
- ◆ Port to open-source PyOMO / Python platform
- ◆ Dynamic optimization using Orthogonal Collocation on Finite Elements
- ◆ Simultaneous direct evaluation of sensitivities
- ◆ Simultaneous calculation (to optimize or constrain) the various DoE objectives of the information matrix
- ◆ Parallel experimental campaign structure due to long duration
- ◆ Design for multiresponse data with “missing observations”
 - Account for both cost and information content of a measured response
- ◆ Incorporate “Robust” experimental design methodologies
 - Reduce dependence on initial parameter guess
 - Best design for worst-case parameters values
- ◆ Provide “Recourse” by
 - Performing parameter estimates throughout the campaign
 - Updating sampling times for specific responses
 - Designing additional experiments for remaining campaign horizon

- Create the incentive for the Business Area to *first consult* with the Modeling & Simulation Group to determine, *before any experiments are conducted and measurements are taken*:
 - ◆ Whether or not the contemplated *experimental campaign* can achieve the desired *business objective*
 - ◆ If not, utilize the tool to determine the impact of:
 - Extending the limits on the *Experimental Design Space*
 - Including *responses* that may be *less noisy* and/or provide significantly *better information content*
- **Reduce cost** of running an “ad-hoc” (versus “model-based” or “systematic”) campaign with the associated cost and time that
 - ◆ May generate *data with limited information content*
 - ◆ *Cannot be analyzed* to determine if the business objective can be met
- Paradigm shift to start with a parallel Design of Experiments before gathering any data by leveraging the reaction network model