Data-Driven Modeling of Batch Processes: Two Methodological Generalizations

Time-Varying Inputs and Time-Resolved Outputs

CHRISTOS GEORGAKIS
Systems Research Institute for Chemical and Biological Processes & Department of Chemical and Biological Engineering
TUFTS UNIVERSITY
MEDFORD MA 02155, USA

What I am Going to Tell You?

- Generalization of the Design of Experiments
  - Design of Dynamic Experiments
  - Dynamic Response Surface Models
- Use DoDE - DRSM to:
  - Model Processes Not Well Understood
    - Mostly Batch but also Continuous Processes
  - Optimize them
    - Almost as Well as with a Knowledge-Driven Model (KDM)
  - Even ... Proceed towards a KDM
- Industrial Applications
The Limitations of DoE

DoE a Very Powerful Methodology 50 Years Young!
- Full and Fractional Factorial Designs, ANOVA
- RSM: Interpolative and Linear and Nonlinear Models
  - Linear in Parameters

Two Major Limitations of DoE
- Inputs Do NOT Vary with Time
  - Why Keep Reaction Temperature Constant?
  - Why Keep Co-reactant Flow Constant?
- Outputs Measurements at End of Experiment
  - We Take On-Line Spectral and Other Measurement VERY frequently.

Our Answer is DoDE and DRSM
PART A: The DoDE Approach

- Applicable to ANY Time-Varying Input Factor, \( u(t) \)
  - Define Coded variable, \( z(\tau) \)
    
    \[
    u(\tau) = u_0(\tau) + \Delta u(\tau) z(\tau)
    \]
    
    \[
    z(\tau) = \left( \frac{u(\tau) - u_0(\tau)}{\Delta u(\tau)} \right) = \frac{u_0(\tau) + \frac{u_{\text{max}}(\tau) + u_{\text{min}}(\tau)}{2}}{\Delta u(\tau)} - \frac{u_{\text{max}}(\tau) - u_{\text{min}}(\tau)}{2}
    \]
    
    \(-1 \leq z(\tau) \leq +1, \quad \tau = t/t_b\)

- Parameterize Input: \( z(\tau) \)
  - Using: \( P_1(\tau) = \) Shifted Legendre Polynomials
    
    \[
    P_i(\tau) = 1, \quad P_1(\tau) = -1 + 2\tau, \quad P_2(\tau) = 1 - 6\tau + 6\tau^2, \ldots
    \]
  - Orthogonality: \( \int_0^1 P_i(\tau) P_j(\tau) d\tau = 0 \) for \( i \neq j \)

- Dynamic Sub-factors: \( x_1, x_2, \ldots, x_n \); \(-1 \leq x_1 \pm x_2 \pm \ldots \pm x_n \leq +1\)

DoDE with \( n=2 \): a \( 3^2 \) Design

- Dynamic Factor: \( z(\tau) \)
  - Dynamic Subfactors: \( x_1 \) and \( x_2 \)
    
    \[
    z(\tau) = x_1 P_0(\tau) + x_2 P_1(\tau) = x_1 + x_2 (2\tau - 1); \quad -1 \leq x_1 \pm x_2 \leq +1
    \]
The nine (9) runs within the Region

Quadratic Time Profiles

- **The $2^3=8$ Full Factorial DoDE**
- Dynamic Factor: $z(\tau)$
  - Dynamic Subfactors: $x_1$, $x_2$, and $x_3$

\[
z(\tau) = x_1 P_0(\tau) + x_2 P_1(\tau) + x_3 P_2(\tau) = x_1 + x_2 (2\tau - 1) + x_3 (1 - 6\tau + 6\tau^2)
\]

&

\[-1 \leq x_1 \pm x_2 \pm x_3 \leq +1\]

so that

\[-1 \leq z(\tau) \leq +1\]
DoE & DoDE - Response Surface Models

- The DoE Steps
  - Design of Experiments ➔ Data ➔ Multilinear Regression ➔ Response Surface Model ➔ Optimization

- Response Surface Model (RSM)
  \[ y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \beta_{ii} x_i^2 \]

- Design of Dynamic Experiments: The Same!
  - Parameterize Time-Varying Input \( z(\tau) \); \((\tau=t/t_b)\)
  - \( z(\tau) = \sum_{i=1}^{n} \tau_i P_i(\tau) \) - Shifted Legendre Polynomials

PARAMETERIZE, DoDE experiments, RSM, OPTIMIZE

DoDE Example: Batch Reactor

- Batch Reversible Reaction \([15 < T < 50 \, ^\circ C]\)
  - \( A_1 \rightleftharpoons A_2 \) \( k_i = k_{i0} \exp(-E_i/RT) \) with \( E_2 > E_1 \)

Model-based Optimum: Decreasing Temperature Profile

Optimum Conversion = 74.57% at \( t_b = 2.0 \) hr

DoE or DoDE ???
Reactor Optimization via DoE & DoDE

- **Single Factor: Reactor Temperature**
  - Data: Conversion at 2hr + Error (±3%)
  - Five DoE Experiments at T=15, 32.5 (3), and 50 °C
  - T constant with time!
  - Nine DoDE Experiment (T(t) linear in Time)
  - Between 15 and 50°C

- **Optimization: Maximum Conversion**
  - DoE Optimum: x=71.44 at T*= 36.25 °C
  - DoDE Optimum: x=74.32, T* from 50 to 28°C
  - Model-Based (True) Optimum = 74.57%

---

DoDE on Isothermal Semi-Batch Reactor

- **Reaction Example:**
  - Run1: A + B → C, \( r_1 = k_1 C_A C_B \), \( k_1 = 2 \text{ mol}^{-1} \text{h}^{-1} \)
  - Run2: 2B → D, \( r_2 = k_2 C_B^2 \), \( k_2 = 1 \text{ mol}^{-1} \text{h}^{-1} \)
  - Run2: C → E, \( r_2 = k_3 C_C \), \( k_3 = 1 \text{ h}^{-1} \)

- **DoDE Runs: Feeding B**
- **Optimal Runs**
Sepracor Pharmaceutical Reaction System

Asymmetric Catalytic Hydrogenation

<table>
<thead>
<tr>
<th>Reactant</th>
<th>Catalytic Hydrogenation → trans-Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reactant</td>
<td>Catalytic Hydrogenation → cis-Product</td>
</tr>
</tbody>
</table>

Project Specific Goals:
- Optimize Reaction Conditions
  - Selectivity of Asymmetric Hydrogenation
  - Minimize Catalyst Loading
- Performance Criterion
  - Profit = Value of Product - Cost of Reactants

Experiments and Analysis performed by Fenia Makrydaki, PhD candidate

Sepracor Experimental System

Advantages
- Accurate Measurements
- Precise Pressure Control
- H₂ Consumption Monitoring
- Minimize Mass Transfer Limitations

Gas Controllers
Raman RXN1
Temperature Controllers
HEL ChemLog Temperature Logger
Agitator
Raman Probe
Design of Dynamic Experiments – DoDE

8=2^3 experiments with 2 Levels & 3 (2+1dynamic) Factors Full Factorial

- Time Variant Experiments: Temperature Profile
- Advantages: Additional degrees of freedom

Figure A: 2 level, 2 factor, full factorial case

Figure B: 2 level, 2 factor, full factorial case for two time horizons.

DoDE & DRSM: Generalizations of DoE

DoE Design Table & Responses

D-Optimal Experimental Design -17 Runs

<table>
<thead>
<tr>
<th>Run</th>
<th>x1 (T)</th>
<th>x2 (RE)</th>
<th>x3 (CL)</th>
<th>x4 (BT)</th>
<th>DE (%)</th>
<th>Y (%)</th>
<th>PI ($/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1.67</td>
<td>-1</td>
<td>97.4</td>
<td>85.5</td>
<td>562.2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1.67</td>
<td>1</td>
<td>93.7</td>
<td>75.7</td>
<td>497.7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>97.7</td>
<td>98.6</td>
<td>153.7</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>1.67</td>
<td>1</td>
<td>97.3</td>
<td>96.1</td>
<td>668</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>99</td>
<td>99.7</td>
<td>56.2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>97.7</td>
<td>98.6</td>
<td>155.7</td>
<td></td>
</tr>
</tbody>
</table>

BEST DoE Run

| 7 | -1 | 1.67 | -1 | 0 | 96.8 | 95.4 | 726.1 |

DoDE & DRSM: Generalizations of DoE
**DoDE** Design Table & Responses

- 3 Static & 2 Dynamic Factors, 21 Runs +3 CPs

<table>
<thead>
<tr>
<th>Run</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$ (RE)</th>
<th>$x_4$ (BT)</th>
<th>DE (%)</th>
<th>Y (%)</th>
<th>PI ($/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.88</td>
<td>0.13</td>
<td>1</td>
<td>1</td>
<td>4.1</td>
<td>97.3</td>
<td>98.7</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
<td>0.06</td>
<td>1</td>
<td>1</td>
<td>7.5</td>
<td>97.2</td>
<td>98.4</td>
</tr>
<tr>
<td>3</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>4</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>5</td>
<td>0.06</td>
<td>0.06</td>
<td>1</td>
<td>1</td>
<td>7.5</td>
<td>97.2</td>
<td>98.4</td>
</tr>
<tr>
<td>6</td>
<td>-0.11</td>
<td>0.11</td>
<td>1</td>
<td>1</td>
<td>7.5</td>
<td>97.2</td>
<td>98.4</td>
</tr>
<tr>
<td>7</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>8</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>9</td>
<td>0.06</td>
<td>0.06</td>
<td>1</td>
<td>1</td>
<td>7.5</td>
<td>97.2</td>
<td>98.4</td>
</tr>
<tr>
<td>10</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>11</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>12</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>13</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>14</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>15</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>16</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>17</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>18</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>19</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>20</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>21</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>22</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>23</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>24</td>
<td>0.15</td>
<td>0.15</td>
<td>1</td>
<td>1</td>
<td>2.42</td>
<td>98.1</td>
<td>98.8</td>
</tr>
</tbody>
</table>

**What Did I Tell You so Far?**

**Design of Dynamic Experiments (DoDE)**
- First Generalization of DoE
  - New Set of Inputs: TIME-VARYING
- Effective Optimization of Processes
- SMALL distance from Model-Based Optimum
PART B: Dynamic RSM (DRSM)

- Use Time-Resolved Output Data
- Classical RSM: 
  \[ y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \beta_{ii} x_i^2 \]
- Dynamic RSM: 
  \[ y(\tau) = \beta_0(\tau) + \sum_{i=1}^{n} \beta_i(\tau) x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{ij}(\tau) x_i x_j + \sum_{i=1}^{n} \beta_{ii}(\tau) x_i^2 \]
- Parameterization of 
  \[ \beta_i(\tau) = \gamma_{ij} P_0(\tau) + \gamma_{ij} P_1(\tau) + \ldots + \gamma_{ij} P_{R-1}(\tau) \]
  \[ R = \# \text{ of Polynomials} \]
- \# of Model parameters < \# of Data \[ R < K \]

DRSM for Simple Batch Reaction

\[ A \xrightarrow{k_1} B \]
\[ r = k_1[A] - k_2[B] \]
\[ k_1 = k_{10} \exp\left( -\frac{E_1}{RT} \right) \]
\[ k_{10} = 1.32 \times 10^8 \text{ h}^{-1} \]
\[ E_1 = 10,000 \text{ kcal} \]
\[ k_2 = k_{20} \exp\left( -\frac{E_2}{RT} \right) \]
\[ k_{20} = 5.25 \times 10^{13} \text{ h}^{-1} \]
\[ E_2 = 20,000 \text{ kcal} \]
Statistical Measure of Accuracy

- **Unmodeled Variance:**

  \[ SS_{unm}(R,K) = \sum_{i=1}^{m} \frac{\left( \sum_{j=1}^{n} (y_{exp}(t_j;i,R,K) - y_{exp}(t_j)) \right)^2}{\sum_{j=1}^{n} \sum_{i=1}^{m} (y_{exp}(t_j))^2} \]

- **Normal Variability:**

  \[ SS_{over} = \sum_{i=1}^{m} \frac{\left( \sum_{j=1}^{n} (y_{i}(t_j) - y_{i}(t_j))^2 \right)}{\sum_{j=1}^{n} \sum_{i=1}^{m} (y_{i}(t_j))^2} \]

- **Hypothesis Testing:**

  *Null Hypothesis* --- \( H_0 \): \( SS_{over} = SS_{exp}(R,K) \)

  *Alternative Hypothesis* --- \( H_1 \): \( SS_{over} < SS_{exp}(R,K) \)

  \[ F_0(R,K) = \frac{SS_{exp}(R,K)/n_k}{SS_{over}/n_2} = \frac{SS_{exp}(R,K)/(MK-Q)}{SS_{over}/(n_{K^2}-1)} \]

- **F-Statistic:**

Reactor Example:

K=Measurements, R=Polynomials

<table>
<thead>
<tr>
<th>R</th>
<th>K</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>0.97</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>0.51</td>
<td>0.99</td>
<td>0.90</td>
<td>0.95</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.82</td>
<td>0.34</td>
<td>0.48</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.47</td>
<td>0.47</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.58</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( K = \) number of time-resolved measurements; \( R = \) number of polynomials

**Table 3.** \( p \)-values of F-test for DRSM of A \( \rightarrow \) B conversion in batch reactor

**if** \( p(R,K) \leq 0.95 \) the Null Hypothesis **fails** to be rejected \( \Rightarrow \) **Model GOOD**

**if** \( p(R,K) > 0.95 \) the Null Hypothesis **is rejected**
D-RSM Model (R=7, K=14)

More Complex Semi-Batch Case

- Three inter-related reactions
- C is the desired product
  - Reactant B is fed in semi-batch mode

Rxn1: \( A + B \rightarrow C \), \( r_1 = k_1[A][B] \) with \( k_1 = 2 \text{ l mol h}^{-1} \)
Rxn2: \( 2B \rightarrow D \), \( r_2 = k_2[B]^2 \) with \( k_2 = 1 \text{ l mol h}^{-1} \)
Rxn3: \( C \rightarrow E \), \( r_3 = k_3[C] \) with \( k_3 = 1 \text{ h}^{-1} \)

- DRSMs for A(t), B(t), C(t), D(t), and E(t)
Table 6. Corresponding F-test p values for DRSM of product [C] in semi-batch 3-reaction network

<table>
<thead>
<tr>
<th>R</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>2.33E-06</td>
<td>8.06E-06</td>
<td>2.89E-05</td>
<td>1.45E-02</td>
<td>2.22E-05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.18E-26</td>
<td>1.22E-29</td>
<td>5.18E-27</td>
<td>6.14E-37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6.89E-36</td>
<td>9.60E-38</td>
<td>1.07E-50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>9.60E-38</td>
<td>1.07E-50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.07E-50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

K = number of time-resolved measurements; R = number of polynomials

Excellent Model

Some C(t) Profiles

- Excellent fits:
- No Significant Difference Using Stepwise Regression

Simulated Data - DRSM Prediction
Part C: DRSM Usage → Door to Knowledge

- Revisit Semi-Batch Reactor Example
  - Five DRSMs at Hand
    $$c_i(t) = \beta_{\omega}(t) + \sum_{i=1}^{n} \beta_{\omega}(t)x_i + \sum_{j=1}^{m} \beta_{\omega}(t)x_{ij}^2$$
    $$c_j(t) = \beta_{\omega}(t) + \sum_{i=1}^{n} \beta_{\omega}(t)x_i + \sum_{j=1}^{m} \beta_{\omega}(t)x_{ij}^2$$
    $$c_k(t) = \beta_{\omega}(t) + \sum_{i=1}^{n} \beta_{\omega}(t)x_i + \sum_{j=1}^{m} \beta_{\omega}(t)x_{ij}^2$$
    $$c_l(t) = \beta_{\omega}(t) + \sum_{i=1}^{n} \beta_{\omega}(t)x_i + \sum_{j=1}^{m} \beta_{\omega}(t)x_{ij}^2$$
    $$c_m(t) = \beta_{\omega}(t) + \sum_{i=1}^{n} \beta_{\omega}(t)x_i + \sum_{j=1}^{m} \beta_{\omega}(t)x_{ij}^2$$

- Can Calculate Derivatives wrt Time
  $$\frac{dc_i(t)}{dt} = \frac{d\beta_{\omega}(t)}{dt} + \sum_{j=1}^{m} \frac{d\beta_{\omega}(t)}{dt} x_{ij}^2 + \sum_{j=1}^{m} \frac{d\beta_{\omega}(t)}{dt} x_{ij}^2$$
  ... for ALL experiments

Calculate Rate of Appearance (Disappearance)

- Calculate at 100 time points in each Run:
  - T=0.01, 0.02, ..., 0.99, 1.00
- Can plot the Rates vs. Time
  - Can Understand what is Happening

Experiment 1

Experiment 8
Discover the Stoichiometry

Define Big Rate Data Matrix: DataM

\[
\text{DataM} = \begin{bmatrix}
D_1 \\
D_2 \\
\vdots \\
D_n
\end{bmatrix}
\]

- \( D_i \) = Data from \( i \)-th Experiment

\[ Q = 1 + n + 0.5n(n-1) + n + 6(3) \]

for \( n = 2 \) \( \Rightarrow Q = 12(9) \)

- DataM is a 909 x 5 matrix !!!
- SVD of DataM

\[ \text{SVD(DataM)} = \text{USV}^T \]

\[ \text{S has three Dominant Singular Values} \rightarrow \text{three Reactions} !! \]

- Matrix V is KEY to Stoichiometry

Testing Stoichiometries (measure A, B, C, D, and E)

SVD Results

\[
V^T = \begin{bmatrix}
0.4072 & 0.8365 & -0.2557 & -0.2145 & -0.1516 \\
0.2596 & 0.2086 & 0.7629 & -0.2333 & -0.5027 \\
0.5955 & -0.2826 & 0.0140 & 0.4417 & -0.5084
\end{bmatrix}
\]

\[\begin{align*}
R_1: & \quad A + B \rightarrow C \\
R_2: & \quad 2B \rightarrow D \\
R_3: & \quad C \rightarrow E
\end{align*}\]

- TEST:

\[ N_v = \text{NNV}^T; \text{ is } N_v = N ??? \]

- Test TRUE Stoichiometry

\[ \text{Score: 99.74%} \]

- Test Incorrect Stoichiometry

\[ \text{Score: 64.97%} \]

- Can Test One Reaction at a Time
Testing Stoichiometries (measure A, B, C, and D - NO E)

- SVD Results
  \[ V^T = \begin{pmatrix} 0.4125 & 0.8456 & -0.2608 & -0.2164 \\ -0.2841 & 0.3233 & 0.8503 & -0.3029 \\ 0.7620 & -0.1172 & 0.4571 & 0.4435 \end{pmatrix} \]
  \[ R_1: A + B \rightarrow C \]
  \[ R_2: 2B \rightarrow D \]
  \[ R_3: C \rightarrow E \]

- Test TRUE Stoichiometry
  - Score: 99.67%

- Test Incorrect Stoichiometry
  - Score: 64.87%

- Can Test One Reaction at a Time

Calculate Reaction Rates

- From: \( r_A(\tau), r_B(\tau), r_C(\tau), r_D(\tau), r_E(\tau) \)
  - TO: \( r_1(\tau), r_2(\tau), r_3(\tau) \)

Reaction Rates Experiment 4

Reaction Rates Experiment 8
**Derive Kinetic Laws**

- From: \( r_1(\tau), C_A(\tau), C_B(\tau) \)
  - TO: Kinetic Rates \( r_1(\tau) = f(C_A(\tau), C_B(\tau)) \)

- Work in Progress
  - Stay ... Tuned

---

**What did I just Tell you?**

- **DRSM is a Generalization of RSM**
  - Using Time-Resolved Measurements
    - Excellent Approximation of Composition Profiles
  - Stepping Stone to Stoichiometry and a Kinetic Model
Industrial Applications

- Dow Batch Polymerization Reactor
  - Use DoDE to Increase Productivity by 20%
  - Presented at the Houston AIChE Meeting
  - Can Give you the Highlights

- Pfizer Pharmaceutical Reaction System
  - Develop DRSMs & Discover Complex Stoichiometry
  - 10 Species involved in 8 reactions

- ExxonMobil Continuous Polymerization Process
  - Develop Meta-Models of KDM
  - Make Them More Accurate with Plant Data
  - USE for Optimization and Control Between SS Transition

---

DRSM Model for the Optimization and Control of Batch Processes

DYCOPS 2016

Z. Wang, N. Klebanov & C. Georgakis
Tufts University
Medford, MA 02155, USA
Control for Maximum Yield

What You Should Remember Tomorrow

- **DoDE**: First Generalization of DoE
  - Time-varying Inputs
- **DRSM**: Second Generalization of DoE
  - Using Time-Resolved Outputs
    - Excellent Approximation of Composition Profiles
    - Stepping Stone to a Kinetic Model
- Towards Non-Linear Models for Control
- Potential Benefits: **Substantial**

**Thank You**
May I answer your questions?