

***ON-LINE ESTIMATION OF
DIASTEREOMER COMPOSITION USING
RAMAN: Differentiation in High & Low Slurry
Density PLS Models***

Sze Wing Wong

Advisor: **Christos Georgakis**

Gregory D. Botsaris

***System Research Institute &
Chemical and Biological Engineering Dept,
Tufts University***



Motivation

- ❖ API is often an Enantiomer/Racemates
 - Different Pharmacokinetic & Toxicological effects
- ❖ Enantiomers Separation
 - Preferential Crystallization (Seeding)
 - Achiral Synthesis
 - Diastereomer Resolution
- ❖ On-line Monitoring of Solid Composition
 - Continuous Process Improvement

Objective

- ❖ Find Method for Quantitative Analysis of Diastereomers
 - Raman Spectroscopy
 - PLS Regression Model
- ❖ What Process Variables will Affects Raman spectra?
 - Temperature
 - Slurry Density
- ❖ *Q1: Would Prediction Accuracy Improved by Incorporating Process Variables*

Outline

❖ Background Information

– System Background

- *Chemical System*
- *Raman Spectra*

– Experiments

– Model Background

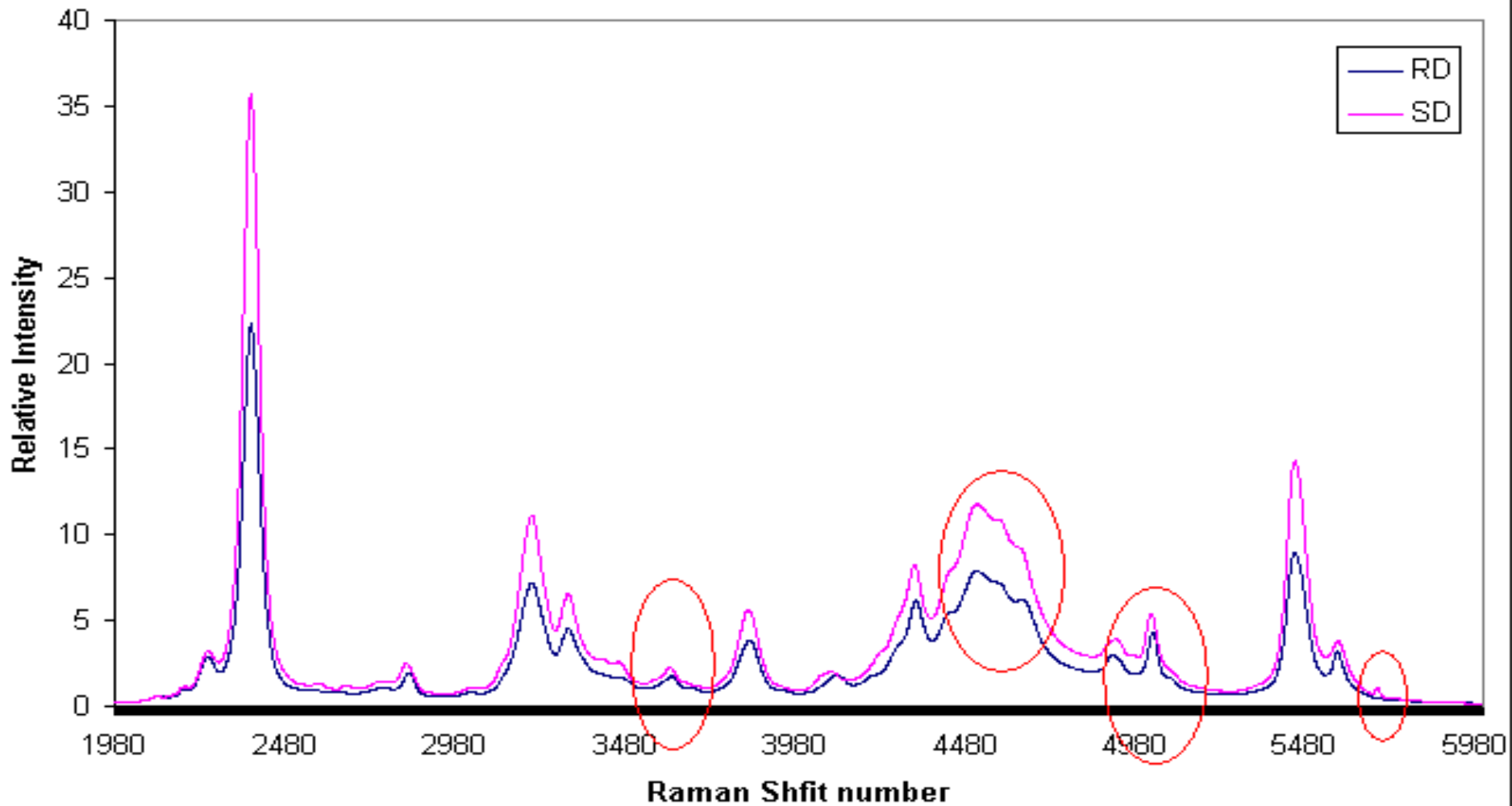
❖ Solution Approach with Chemometric

The Chemical System



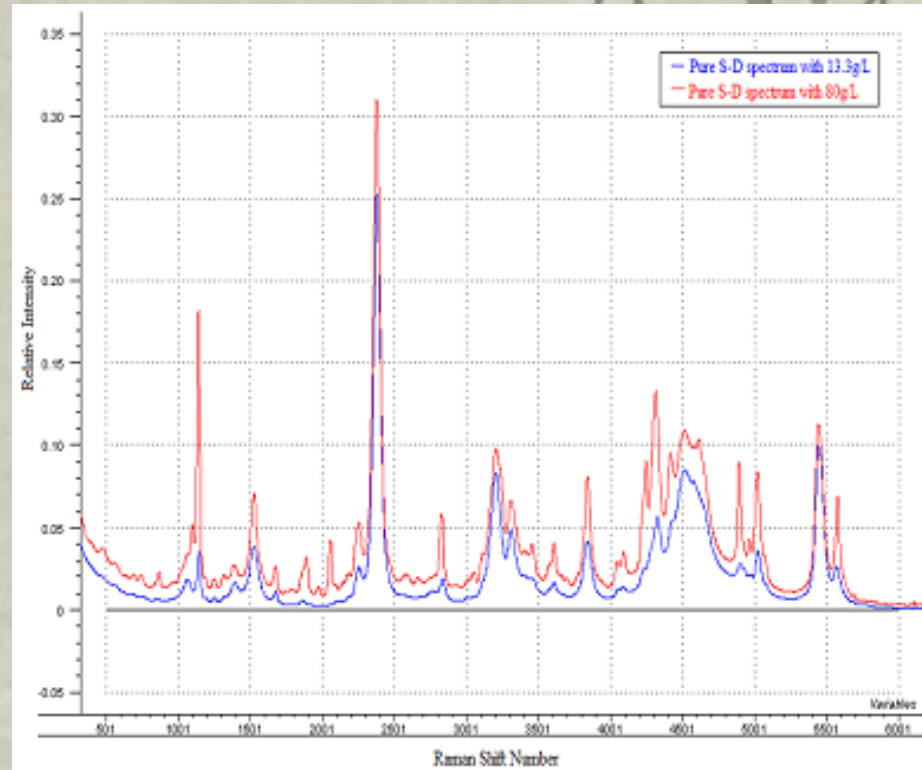
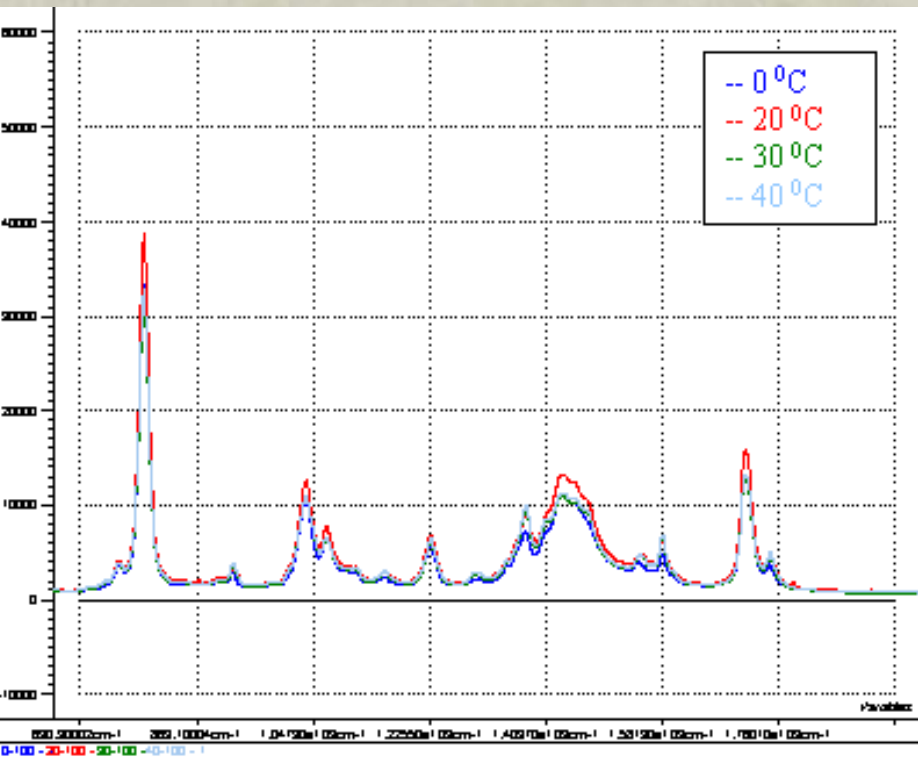
- ❖ Single Enantiomer (*Sepracor* Compound) as Final Drug Form
- ❖ Diastereomers Resolution to Separate Desired Enantiomer
 - Diastereomer Have Different Physical Properties
 - Cooling Crystallization to Separate Desired Diastereomer

Particle Shape Differences:



Process Variables Effect

- ❖ No Peak Shift – Temperature Effect
- ❖ Intensity Differences Due to Different Solvent/Crystal Ratio – Slurry Density Effect



Outline

- ❖ Objective
- ❖ **Background Information**
 - System Background
 - **Experiments & Modeling**
 - *“Static” Calibration Exp.*
 - *Partial Least Square Regression Models*
 - *“Dynamic” Crystallization Exp.*
- ❖ Solution Approach with Chemometric

Set I:

“Static” Calibration Experiment

- ❖ Pre-saturated Solution at Different Temperature
- ❖ Pre-Mixed Slurry in *15mL* vials
- ❖ Total: **101** Data Points

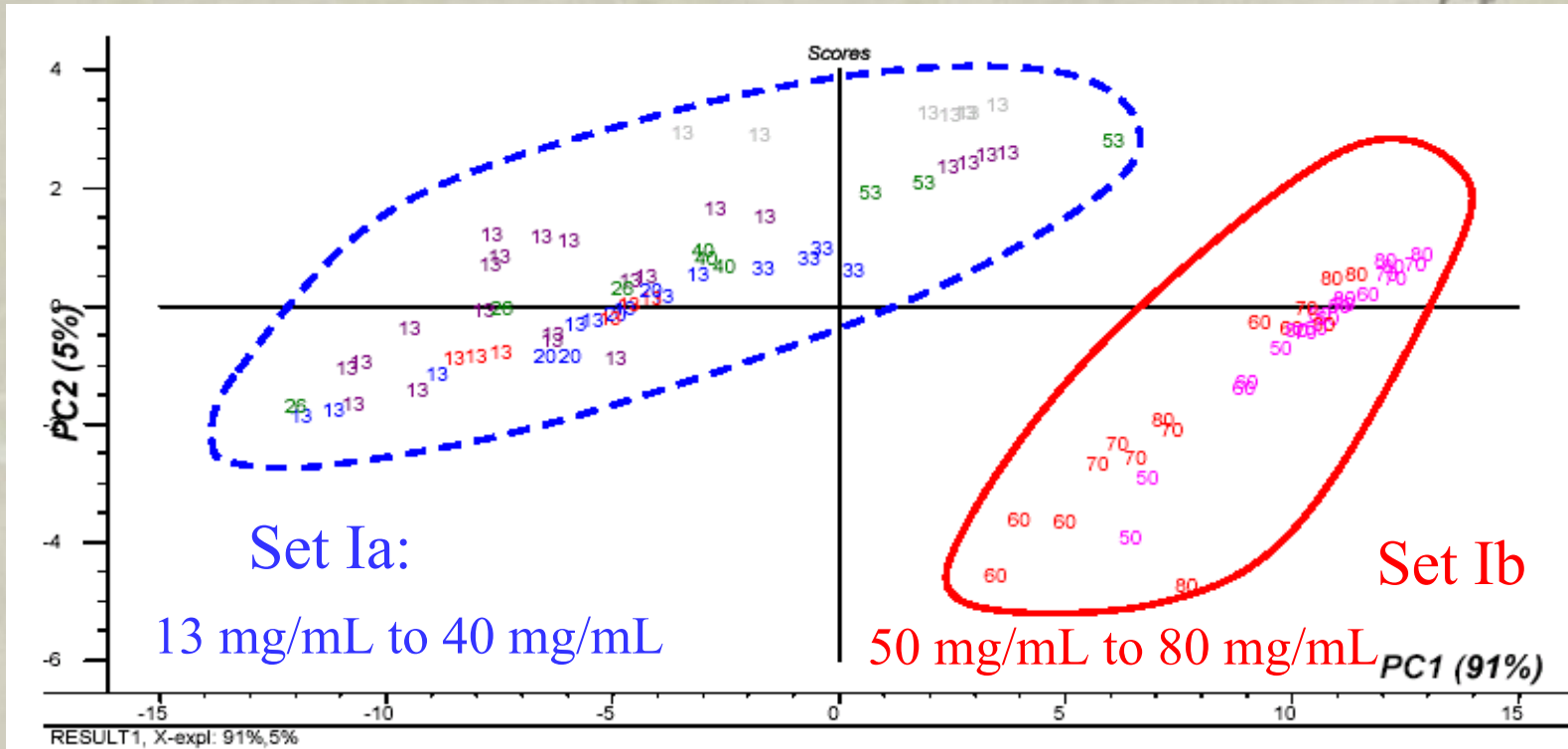


Table 1. Experimental Condition for Standards

Fixed Variable	Fixed Variable	Changing Variable	# of samples
20 °C	13.3 g/L	0-100 % S-D	16
40 °C	13.3g/L	0-100 % S-D	6
30 °C	13.3g/L	0-100 % S-D	6
10 °C	13.3g/L	0-100 % S-D	6
0 °C	13.3g/L	0-100 % S-D	8
0 °C	33.3g/L	80-95 % S-D	4
0 °C	20g/L	70-100 % S-D	4
15 °C	26.7g/L	50-100 % S-D	3
15 °C	40g/L	65-85 % S-D	3
15 °C	53.3g/L	80-100 % S-D	3
5 °C	50 g/L	15-95 % S-D	5
5 °C	60 g/L	15-95 % S-D	5
5 °C	70 g/L	15-95 % S-D	5
5 °C	80g/L	15-95 % S-D	5
10 °C	50 g/L	5-95 % S-D	5
10 °C	60 g/L	5-95 % S-D	5
10 °C	70 g/L	5-95 % S-D	5
10 °C	80g/L	5-95 % S-D	5

PCA – Cluster Analysis

- ❖ Two Clusters Separated by Slurry Density
- ❖ *Q2: Does 1 PLS Model Suffice to Represent the Data Variance?*



12 PLS Estimation Model (4x3)

- ❖ **Q1: Does Process Variables Improve Accuracy**
 - Four Sets of Measurements
 1. $A = [Spectra]$
 2. $B = [Temp \ \& \ Spectra]$
 3. $C = [Density, \ \& \ Spectra]$
 4. $D = [Temp, \ Density, \ \& \ Spectra]$
- ❖ **Q2: 1 or 2 PLS model(s) – Three (3) Data Sets:**
 - Ia (LD), Ib (HD), or $Ia+Ib$ (ALL)
- ❖ Comparison of 12 PLS Models

$$\% \text{ Error} = \left| \frac{(y_{i \text{ estimate}} - y_{i \text{ experiment}})}{y_{i \text{ experiment}}} \right| \times 100\%$$

$$RMS \text{ of } \%E = \sqrt{\frac{\sum_{i=1}^n (\%Error)^2}{n}}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{i \text{ estimate}} - y_{i \text{ experiment}})^2}{n}}$$

Data Fusion

❖ Normalized Different Measurements

- Autoscale Temperature, Slurry Density, and Percent Composition of S-D diastereomer

$$\hat{T}_i = \frac{T_i - T_{avg}}{\sigma_T}$$

$$\hat{D}_i = \frac{D_i - D_{avg}}{\sigma_D}$$

$$\hat{P}_i = \frac{P_i - P_{avg}}{\sigma_P}$$

- Normalize Spectra with Under Curve Area

- Spectral Data Pre-processing – Take 1st Derivative for Baseline Correction

$$F_i = \int f_i(w)dw$$

$$F_{avg} = \frac{\sum_{i=1}^n F_i}{n}$$

$$\tilde{S}_i = \frac{f_i(w) - F_{avg}}{\sigma_F}$$

Selection of Latent Variables

Training Group

1. All = 86 Data Points (Ia+Ib)
2. LD = 46 Data Points (Ia)
3. HD = 40 Data Points (Ib)

Testing Group

1. All = 15 Data Points (Ia+Ib)
2. LD = 11 Data Points (Ia)
3. HD = 4 Data Points (Ib)

PLS Models:

I. Cross-Validation:

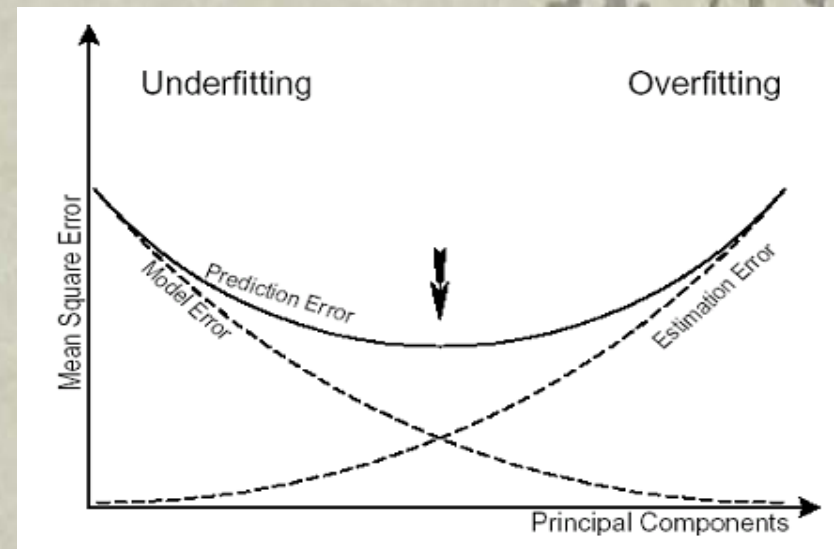
- 84:2 / 44:2 / 38:2

II. Over/Under Fitting

- Testing Groups

III. Model Comparison:

- “Dynamic” Crystallization Exp.



Set II:

“Dynamic” Crystallization Experiments

- ❖ **Unseeded Crystallization**
(250mL)
 - Low Density (Exp 1)
 - High Density (Exp 2)
- ❖ **Seeded Crystallization**
(250mL)
 - S-D diastereomer seed (Exp 3)
 - R-D diastereomer seed (Exp 4)



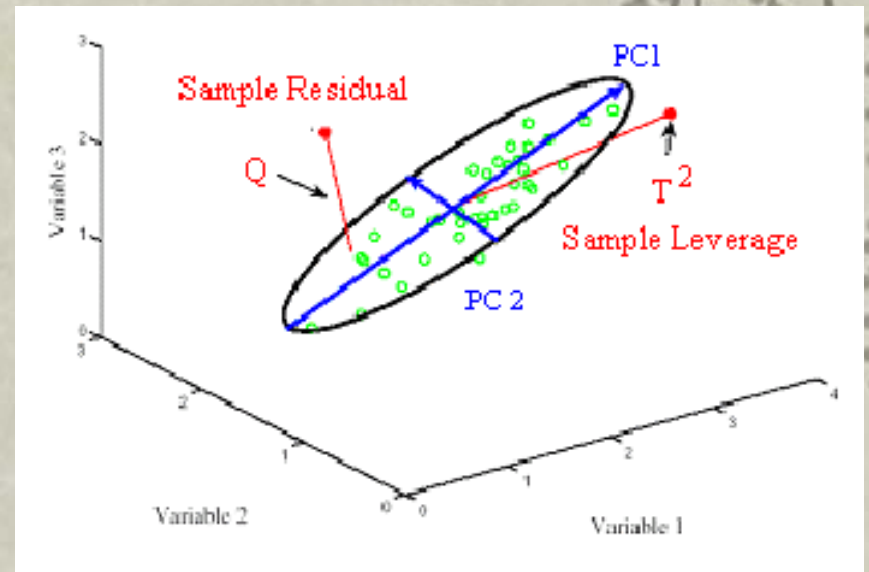
Model Selection: Set Ia vs. Ib

- ❖ Model from TG-LD or TG-HD used in Experiment?
- ❖ How to Quantify Similarity between Samples and Models?

$$Q_i = x_i (I - P_k P_k^T) x_i^T$$

$$T_i^2 = x_i P_k \lambda_k^{-1} P_k^T x_i^T$$

$$d_i = \sqrt{(Q_i)^2 + (T_i^2)^2}$$



Outline

- ❖ Objective
- ❖ Background Information
- ❖ **Solution Approach with Chemometric**
 - **Model Building & Testing (PLS)**
 - **Crystallization Experiments**
 - *LD vs. HD*
 - **Model Selection**

Models Validation

- ❖ RMSE – Cross-Validation vs. Testing Group
 - No over-fitting of Models
- ❖ Data from Testing Groups Not Included in Model Building
 1. Test-All = 15 Data Point
 2. Test-LD = 11 Data Point
 3. Test-HD = 4 Data Point

Model Performance on “Dynamic” Crystallization Experiments?

RMSE of Cross-Validation Result

Data Set	PLS Models w/ Diff Measurements			
	S	T,S	D,S	T,D,S
la	5.10	6.64	6.70	6.18
lb	7.74	7.70	7.90	7.90
la + lb	6.45	7.28	6.55	4.47

RMSE - Prediction Error from Testing Group

Data Set	PLS Models w/ Diff Measurements			
	S	T,S	D,S	T,D,S
la	9.91	10.53	8.87	9.60
lb	5.08	5.18	5.73	5.43
la + lb	9.30	9.73	8.28	8.16

Dynamic Exp. – 1 or 2 Model(s)?

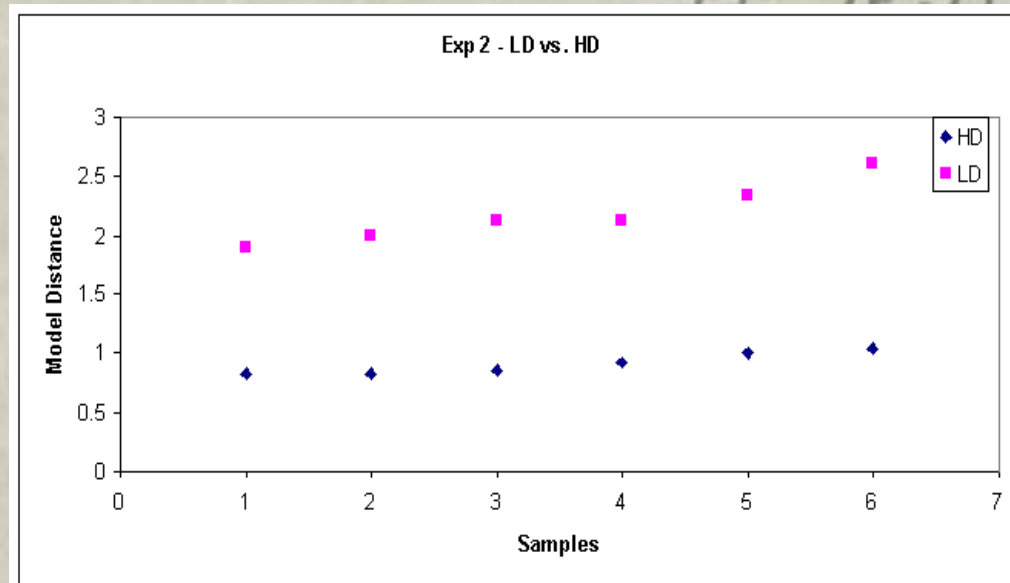
- ❖ *Q2: Does 1 PLS Model Suffice to Represent the Data Variance?*
 - Which PLS Models Perform Better? **ALL vs. LD/HD**

Detail Information of the PCA Model

	PCA-LD	PCA-HD
# of PC	3	3
% Variance	95%	97%

Model Selection

	Exp 1	Exp 2	Exp 3	Exp 4
TG-LD	✓			✓
TG-HD		✓	✓	



Experimental Result

❖ Models from All data Set

- Poor Accuracy
- Does Not Correlate Well with Slurry Density

❖ Models from HD or LD

- Prediction Accuracy Improves w/ **2 Data Sets**
- Models with **T, D, S** Performed Best

RMSE of PLS Models from TG-All

	Model 1 (S)	Model 2 (T, S)	Model 3 (D, S)	Model 4 (T, D, S)	
Exp 1	19.6	16.2	34.2	41.6	LD
Exp 2	76.8	87.5	84.0	50.9	HD
Exp 3	23.0	36.5	14.2	17.2	HD
Exp 4	26.5	12.8	35.9	37.5	LD

RMSE of PLS Models from TG-LD or TG-HD

	Model w/ S	Model w/ T, S	Model w/ D, S	Model w/ T, D, S	
Exp 1 (LD)	18.9	10.8	11.1	7.4	✓
Exp 2 (HD)	7.7	8.0	10.1	9.9	
Exp 3 (HD)	8.9	9.4	9.0	8.8	✓
Exp 4 (LD)	38.3	35.3	5.7	5.4	✓

Prediction Error of Dynamic Exp

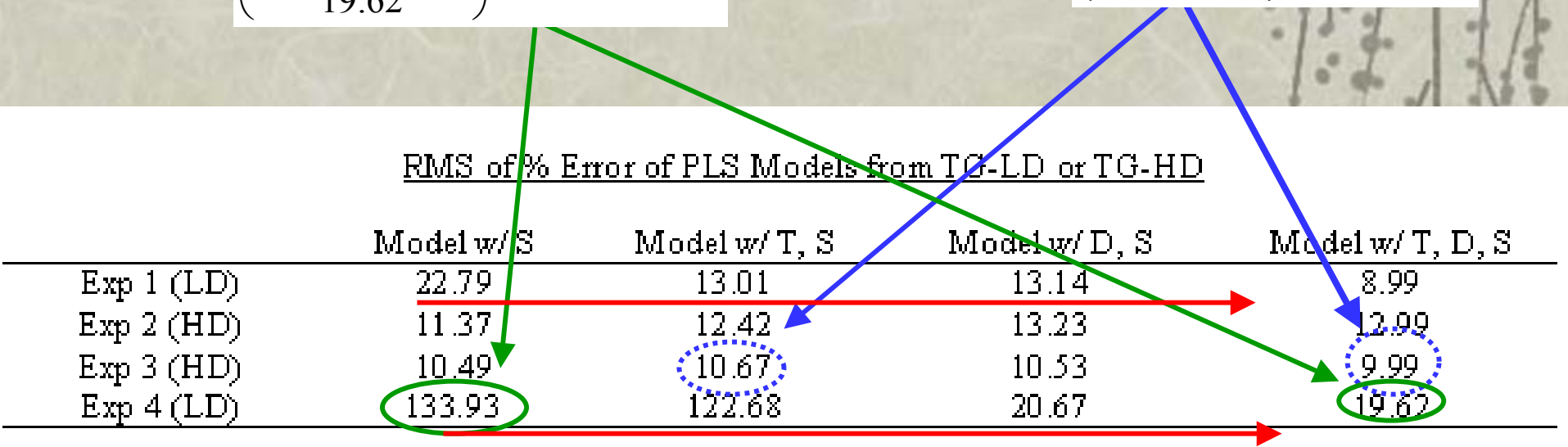
- ❖ Model Accuracy Increased with Process Variables
 - Models are Sensitive with Slurry Density
- ❖ *Raman Spectra ALONE is Not Enough*

$$\left(\frac{133.93 - 19.62}{19.62} \right) \times 100\% = 583\%$$

$$\left(\frac{10.67 - 9.99}{9.99} \right) \times 100\% = 7\%$$

RMS of % Error of PLS Models from TG-LD or TG-HD

	Model w/S	Model w/T, S	Model w/D, S	Model w/T, D, S
Exp 1 (LD)	22.79	13.01	13.14	8.99
Exp 2 (HD)	11.37	12.42	13.23	12.99
Exp 3 (HD)	10.49	10.67	10.53	9.99
Exp 4 (LD)	133.93	122.68	20.67	19.62



Conclusion

- ❖ Raman is Capable in **Differentiating** both Diastereomers
- ❖ Raman Spectra **ALONE**
 - Poor Estimate of % Diastereomers in “Dynamic” Crystallization Experiment
- ❖ Model Accuracy **Improve** with Process Variables Measurement
 - *Temperature & Slurry Density*

Future Challenge

- ❖ Sampling-Loop to Measure Slurry Density:
Not Reliable
 - Use **FBRM**, & **IR** in Place of Slurry Density*
 - Convert Spectral Data into Information: **% S-D**, **Solute Concentration of S-D & R-D**, and **Slurry Density**
- ❖ Model Crystallization Kinetic
 - Model Predictive Control
 - Monitor & Control Optical Purity of Separation Product (i.e. % S-D)

* Wong et. al. (2006). *On-Line Estimation of Diastereomeric Resolution with the Use of FBRM, Raman Spectroscopy, and ATR-FTIR*, AIChE, San Francisco, CA, Section #13f

Acknowledgement

❖ **Sepracor:**

- Kostas Saranteas
- Roger Bakale
- Patrick Mousaw
- Mike Sizensky

❖ **Camo Inc (The Unscrambler)**

- DongSheng Bu

❖ **Financial Support From Sepracor Inc.**