# Computer-aided process engineering in developing and operating continuous processes

Salvador García Muñoz, Ph.D.

Lilly

# Agenda

- Formulation development
- Process development
- Process de-risking and transfer
- Process Monitoring
  - MSPC for fault detection and isolation

# Agenda

- Formulation development
- Process development
- Process de-risking and transfer
- Process Monitoring
  - MSPC for fault detection and isolation

# Formulation development

- Mechanistic modeling at this stage requires the ability to predict the bulk-level behavior from the particle-level information in a multi-component mixture.
  - Mechanical behavior
  - Chemical interactions
    - This is very complex and so far non-attainable.
- Mine the data from past experiments

# Formulation development

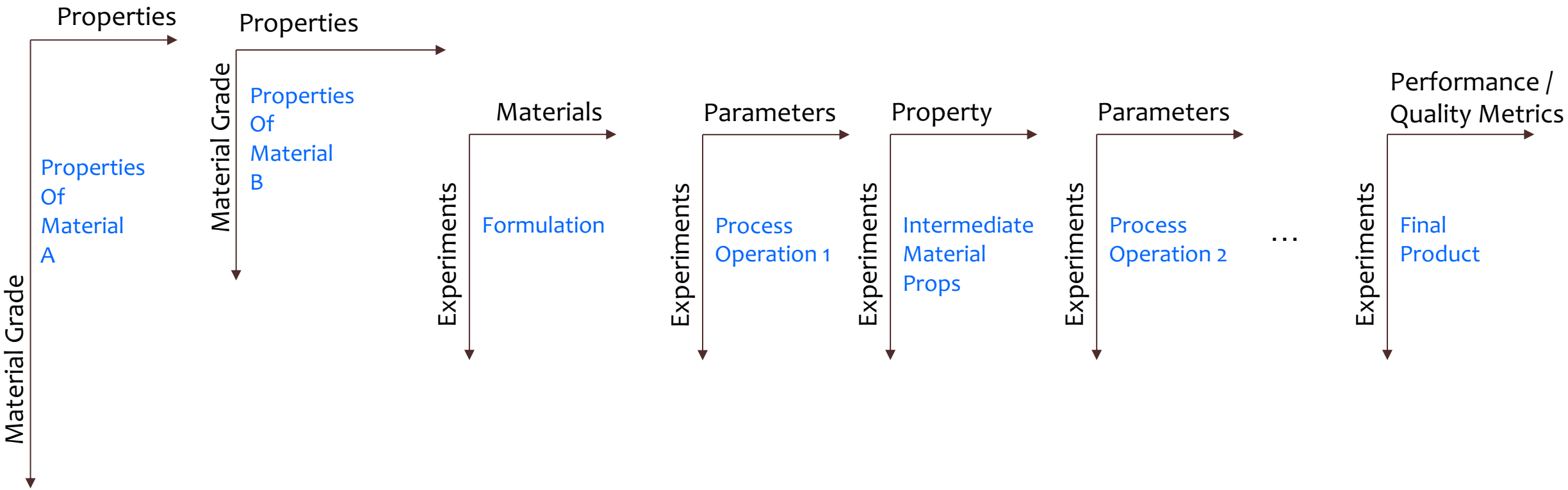
- Mining data from previous experiments.
  - Modeling data from previous products
  - In-silico formulation development
  - Surrogate selection for experimental design
  - Including material variability
    - When you have access to materials
    - When you don’t

# Formulation development

- Mining data from previous experiments.
  - **Modeling data from previous products**
  - In-silico formulation development
  - Surrogate selection for experimental design
  - Including material variability
    - When you have access to materials (Joe K's papers)
    - When you don't (My cloning algorithm)

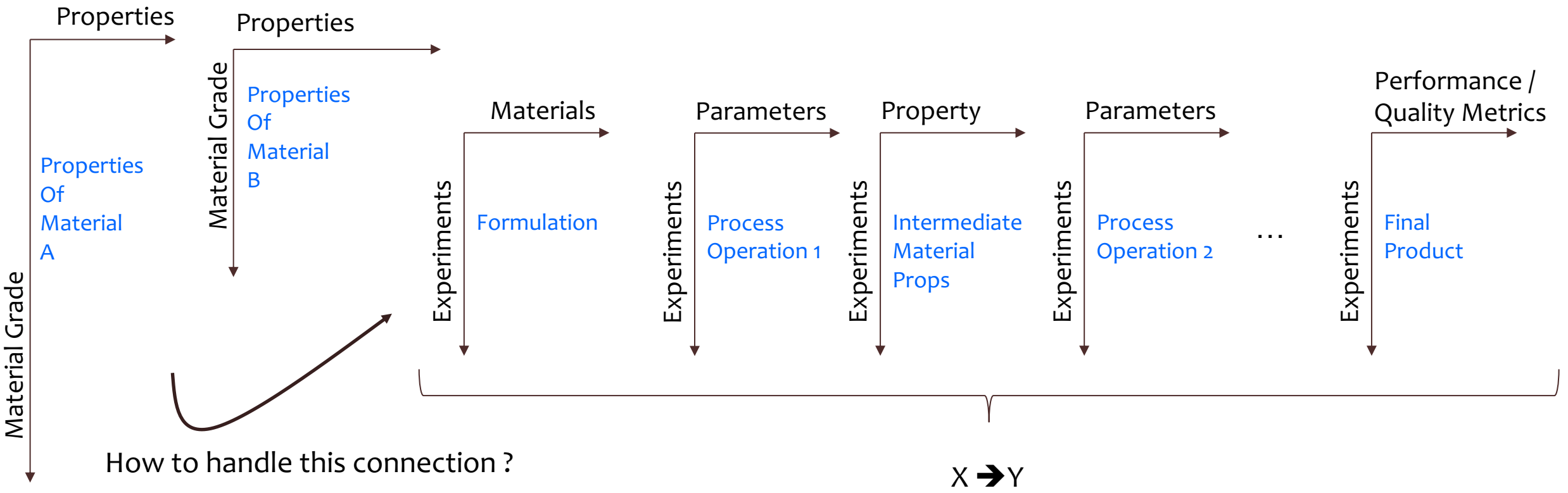
# Formulation development

- Product development data is complex

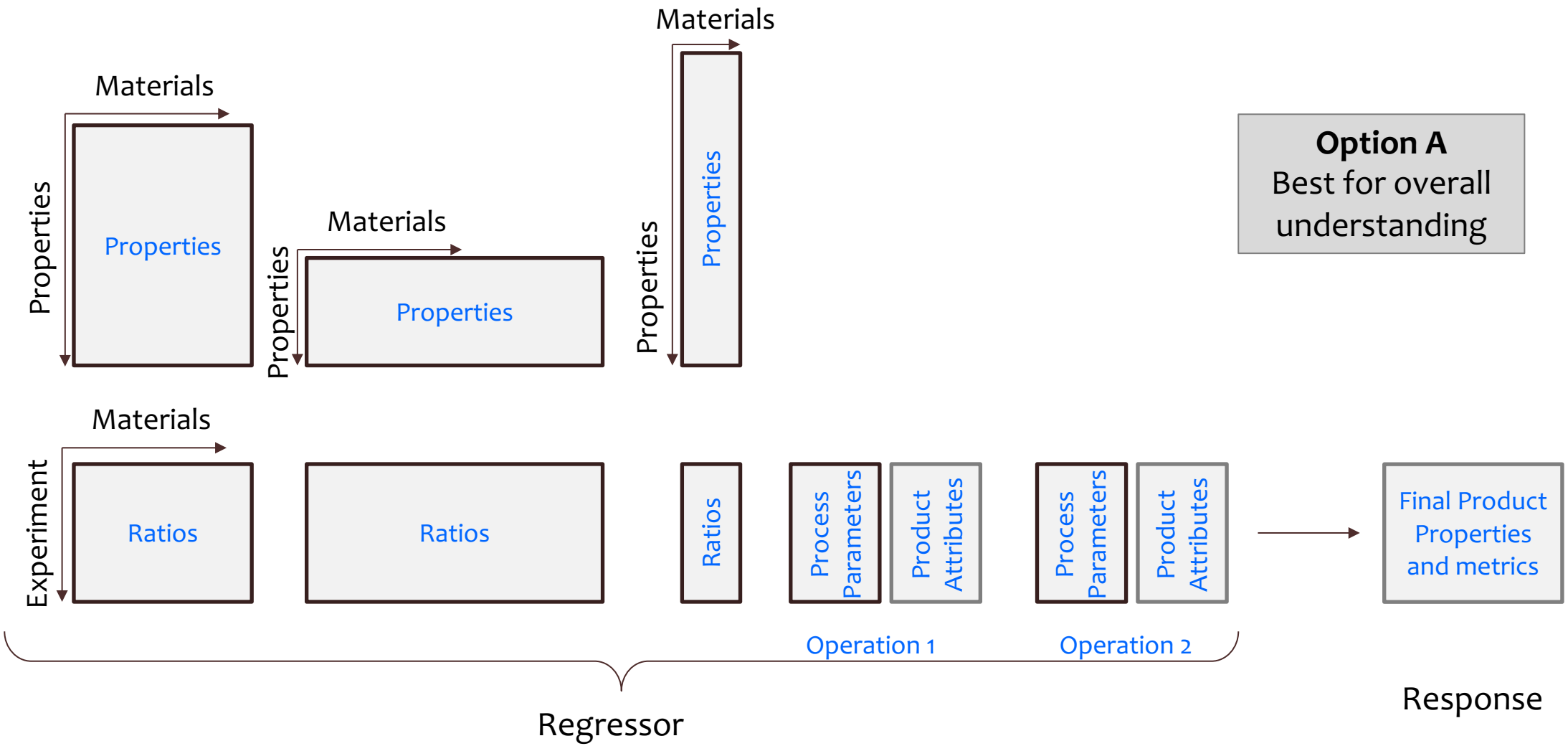


# Formulation development

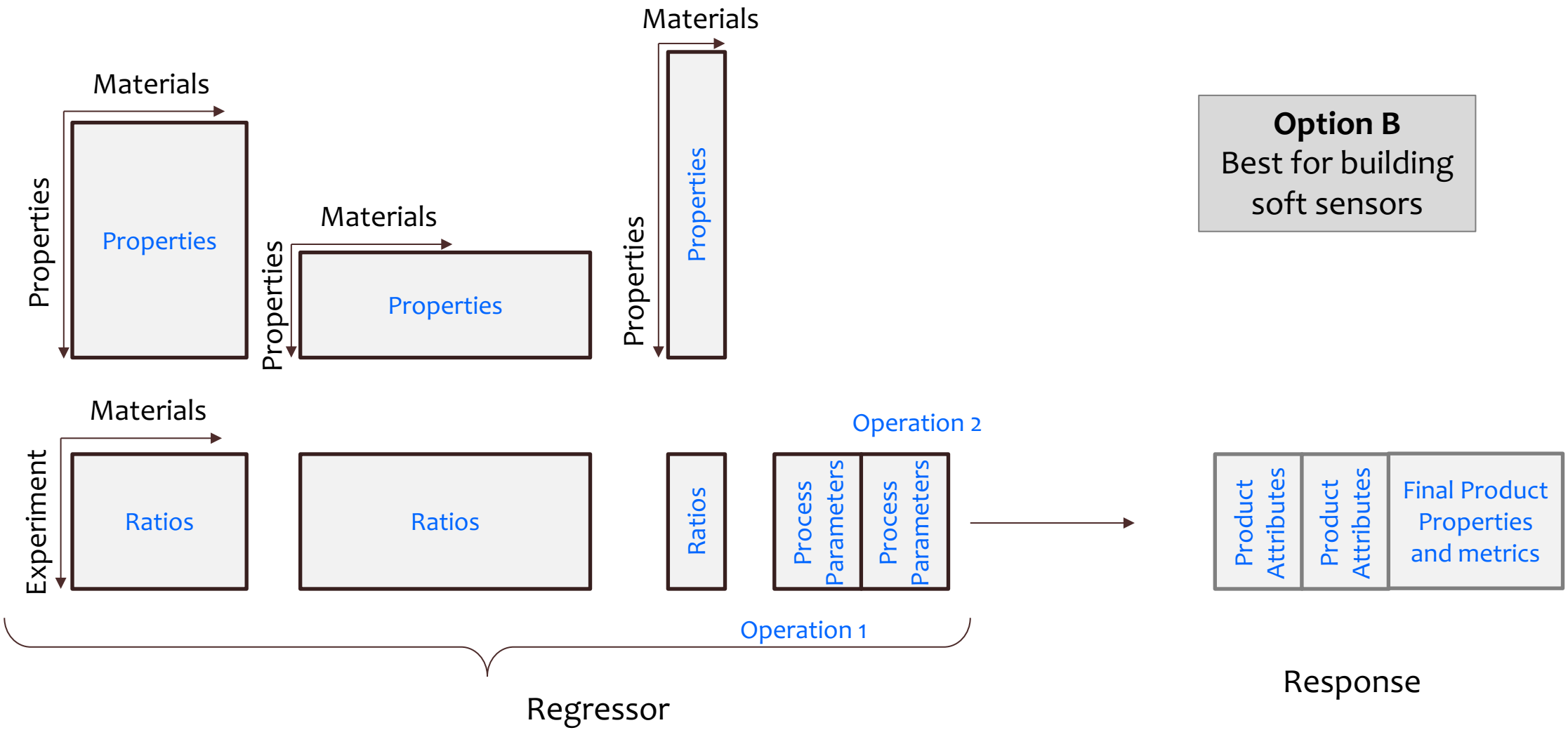
- Product development data is complex



# Formulation development



# Formulation development

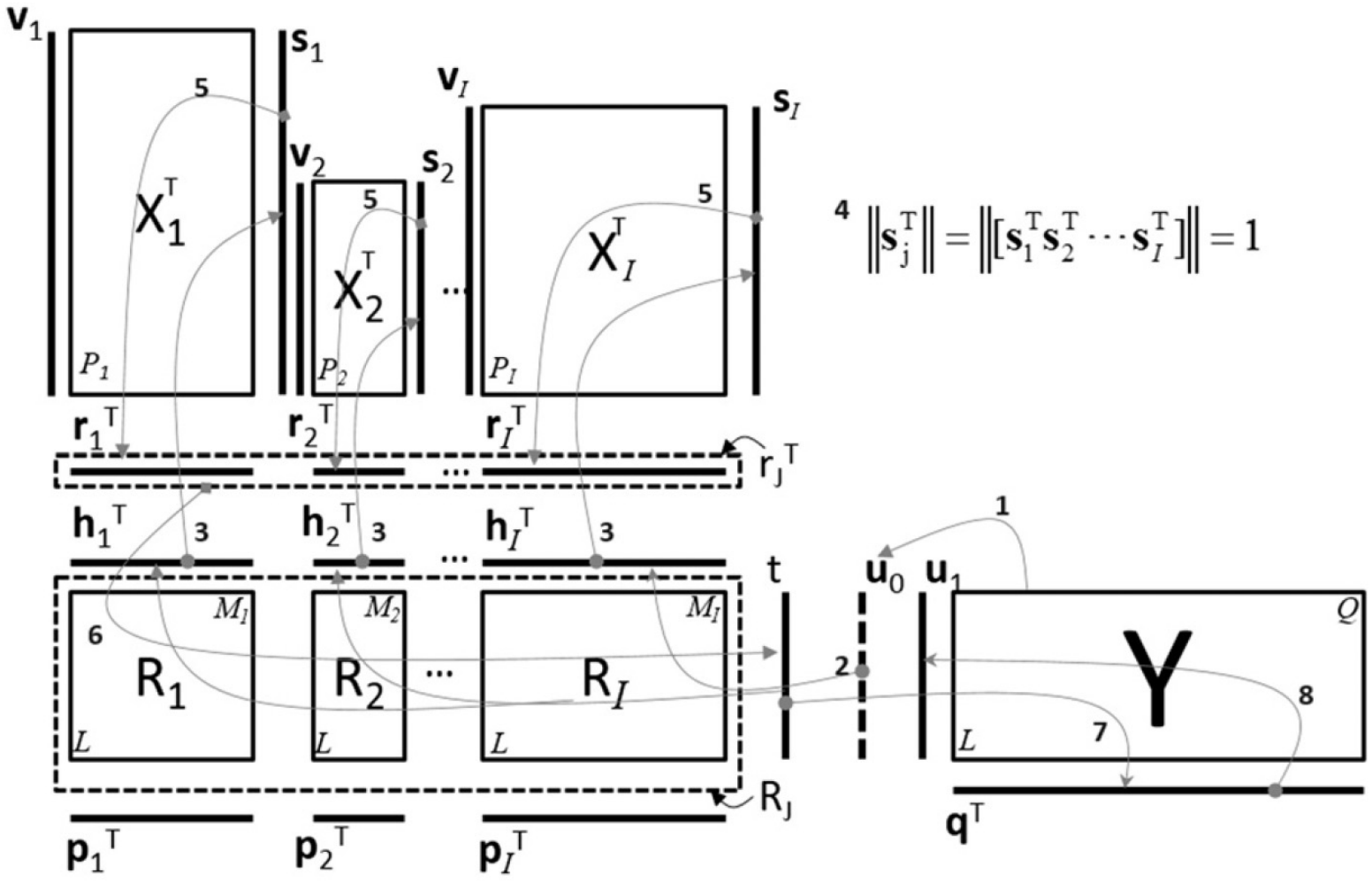



**Option B**
Best for building soft sensors

# Formulation development

*S. Garcia-Munoz / Chemometrics and Intelligent Laboratory Systems 133 (2014) 49–62*

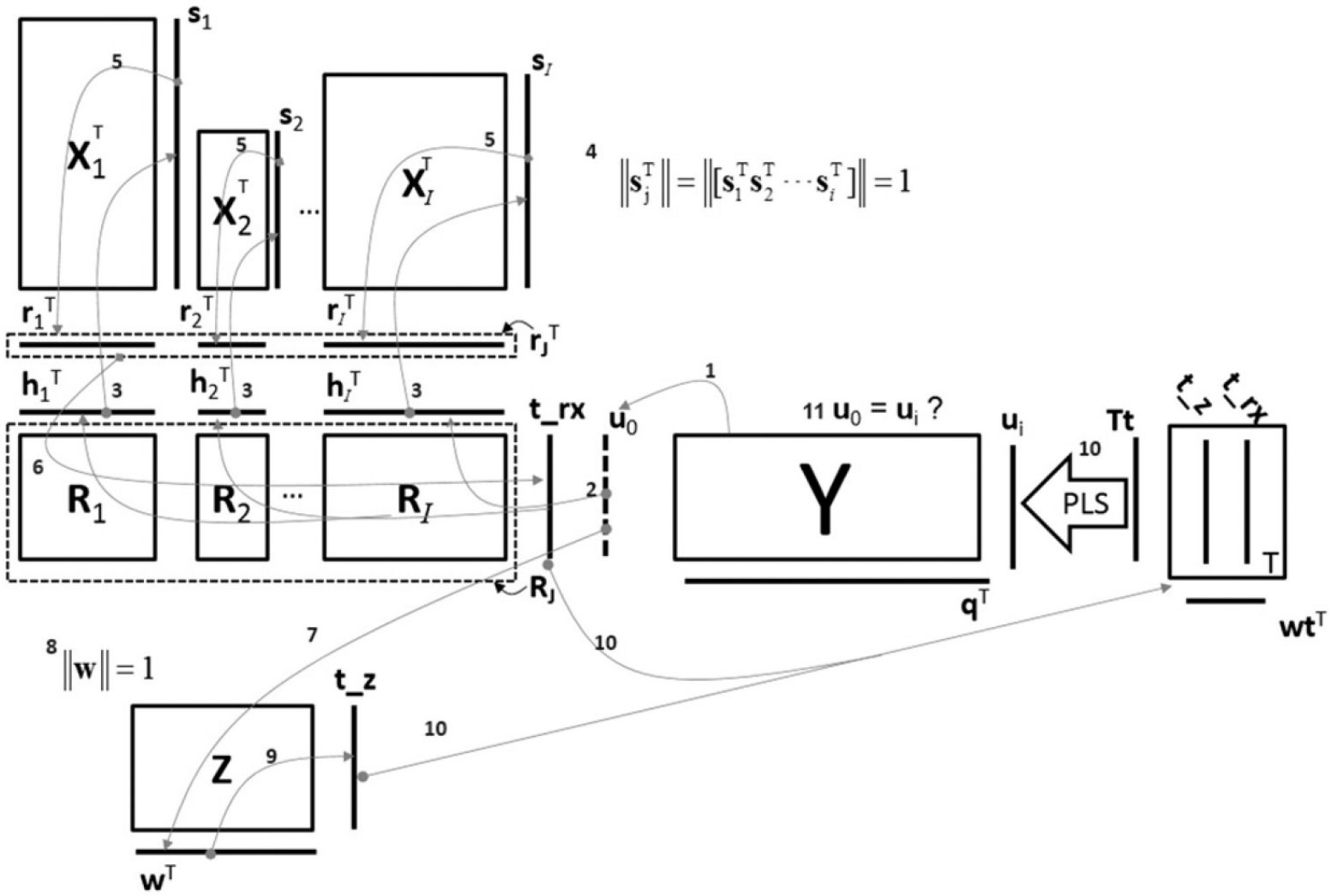
- JRPLS designed to regress effect of materials characterized with different methods.



# Formulation development

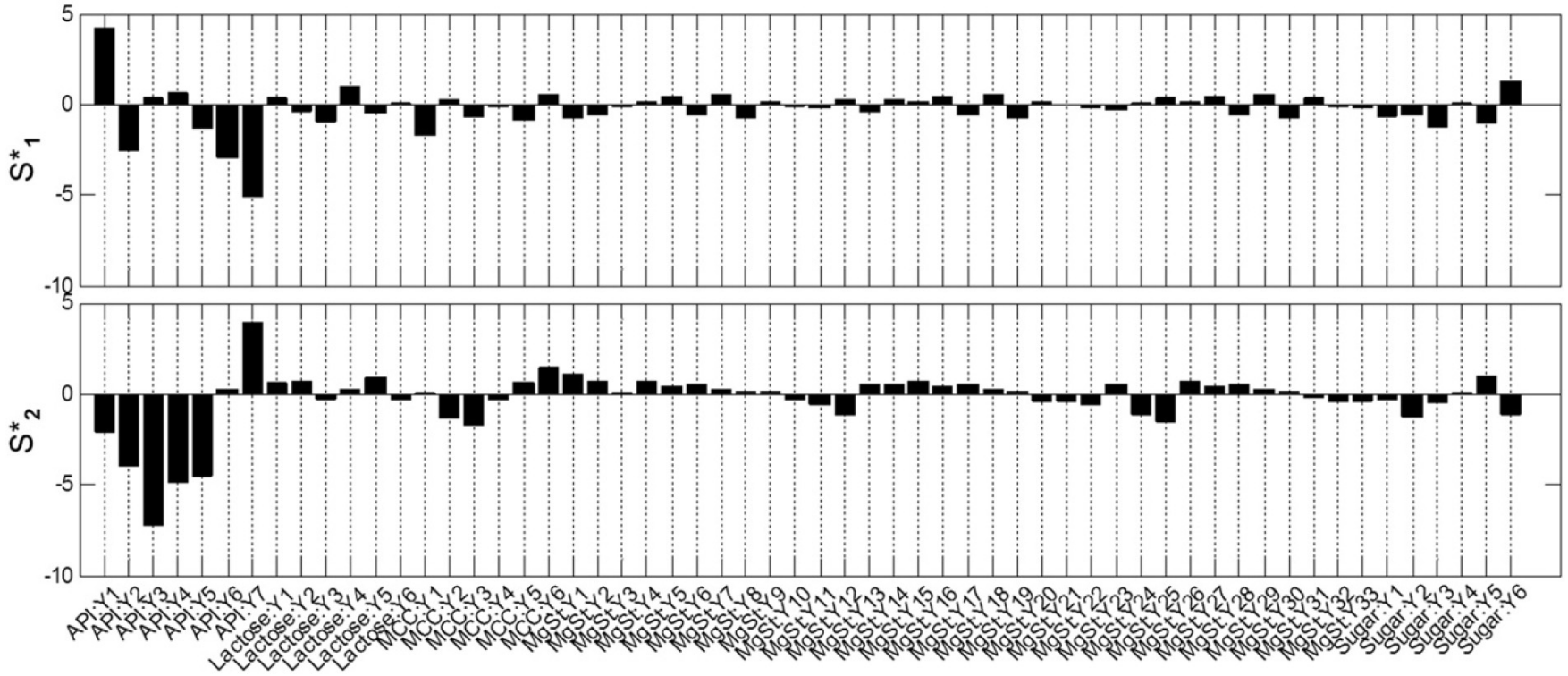
*S. Garcia-Munoz / Chemometrics and Intelligent Laboratory Systems 133 (2014) 49–62*

- TPLS designed to add process information



# Formulation development

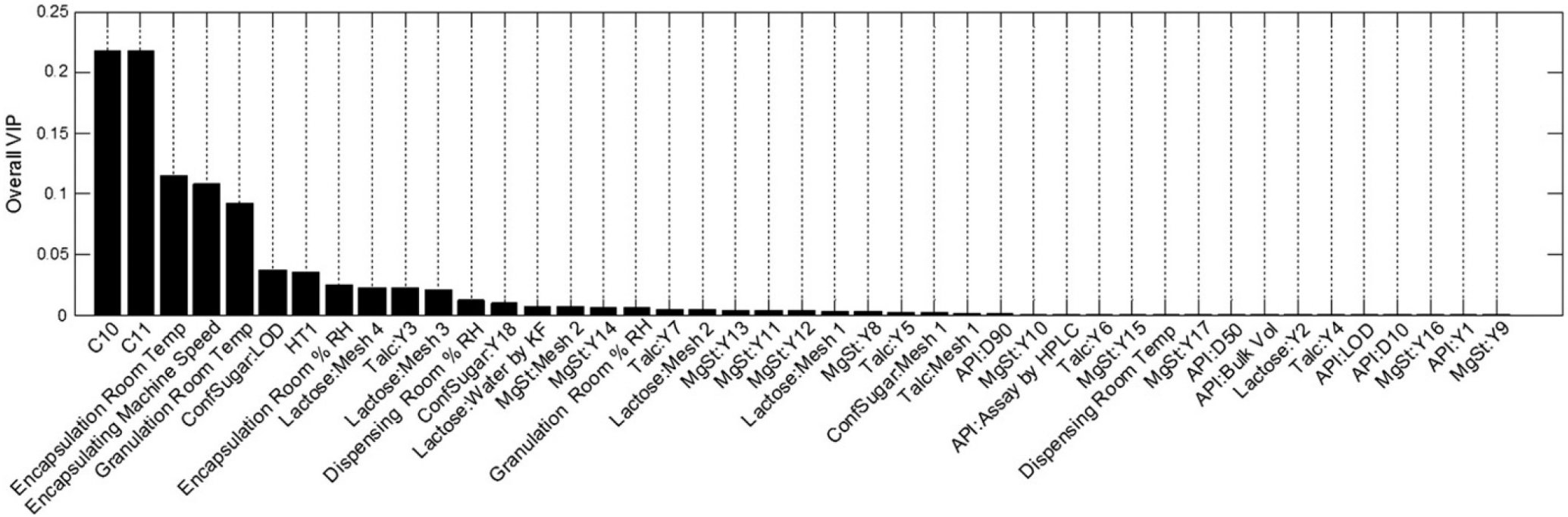
- The objective of using a TPLS or JRPLS model is understanding.




Loadings for material properties across materials

# Formulation development

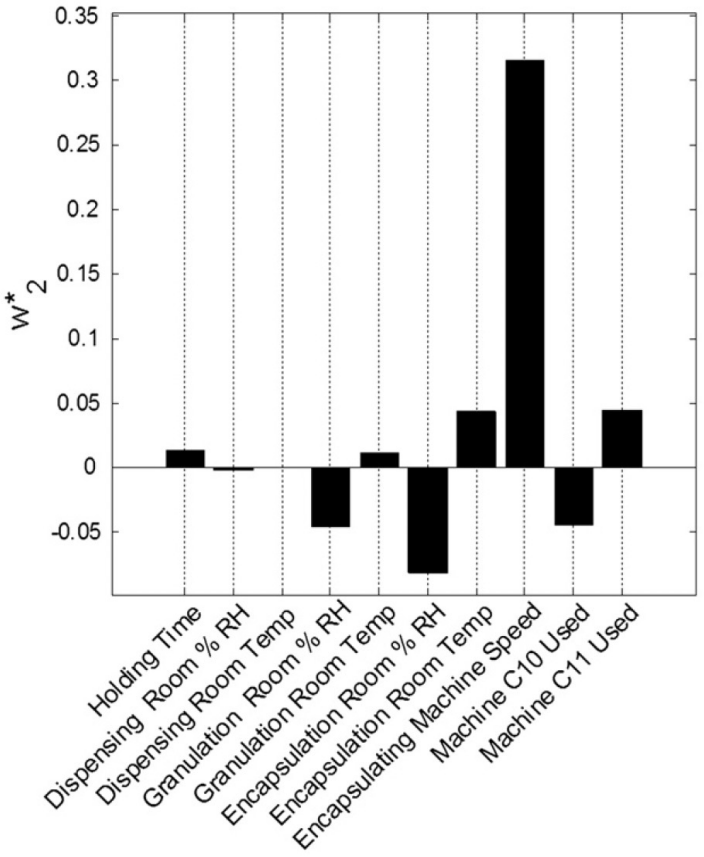
- The objective of using a TPLS or JRPLS model is understanding.

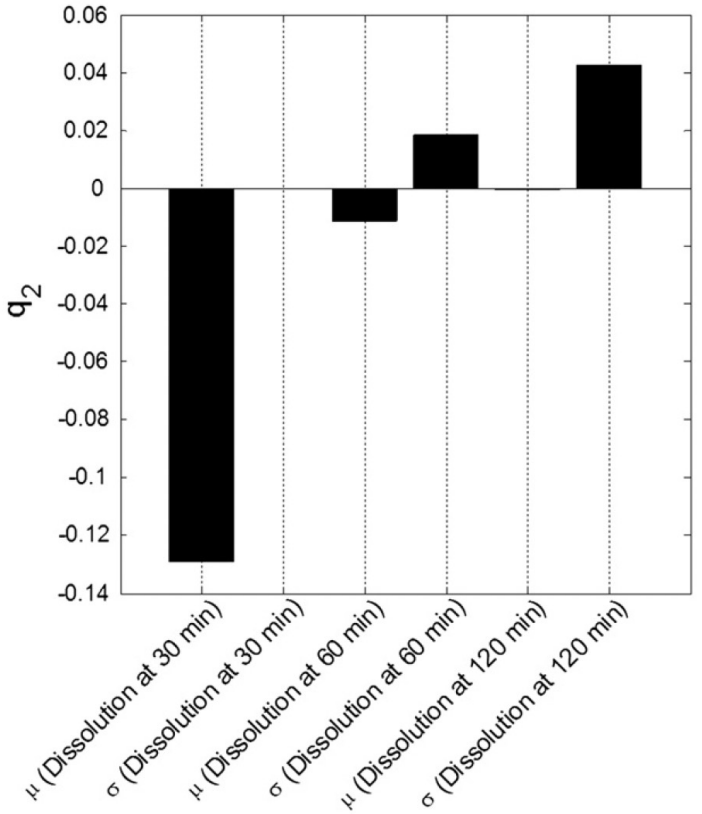



VIP for variables across product and process variables

# Formulation development

- The objective of using a TPLS or JRPLS model is understanding.



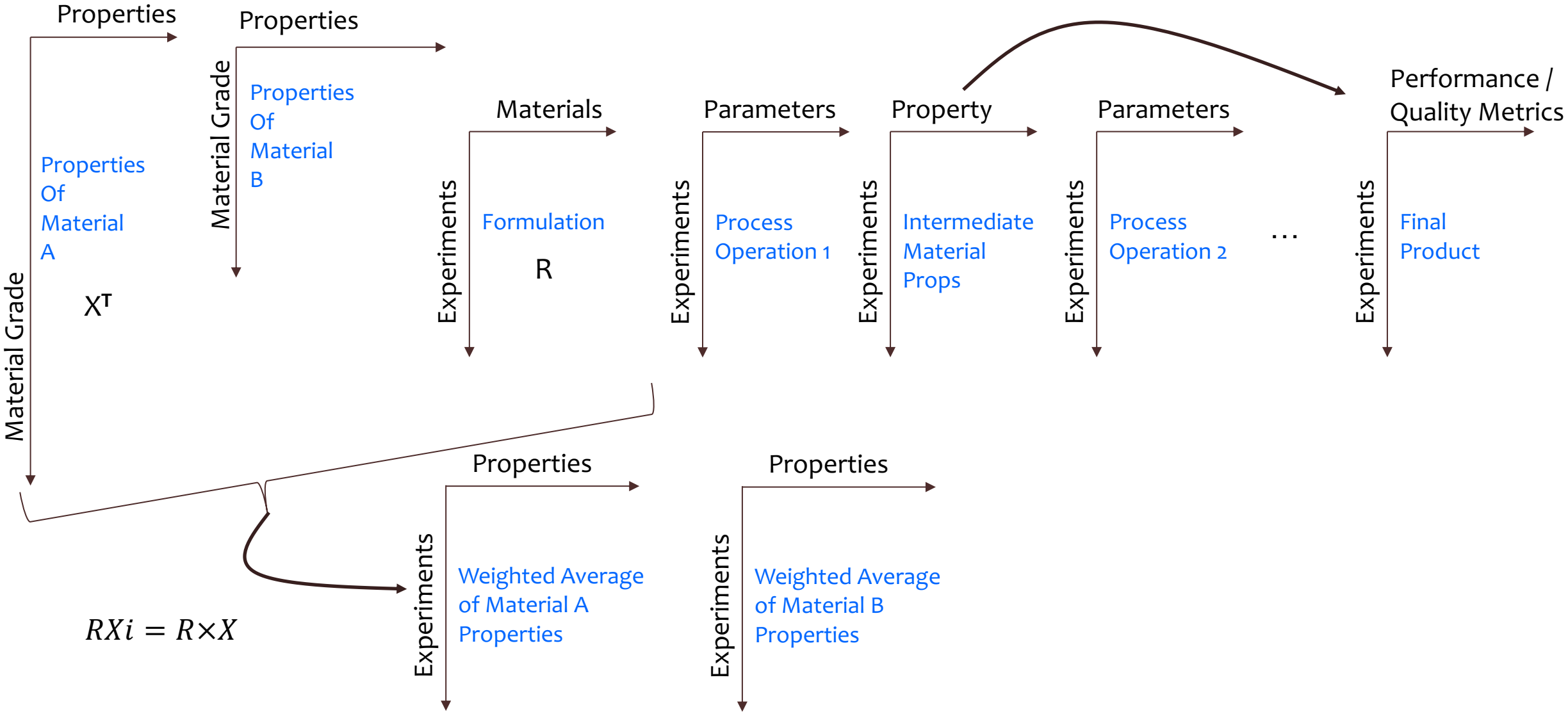




Loadings for process and Quality

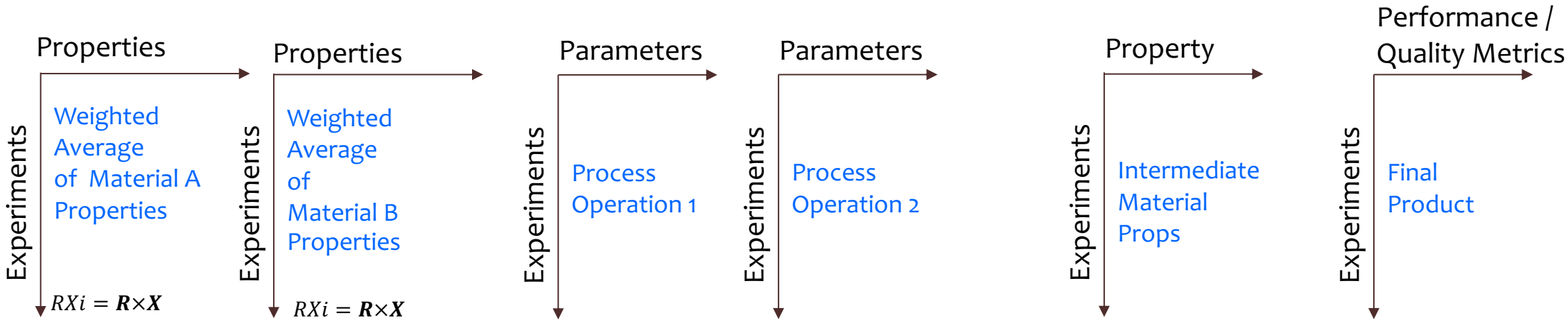
# Formulation development

- A new product likely introduces a new material for which there is not experimental data yet.
- TPLS and JYPLS assign a loading per material.
- No material = No loading
- Without a loading we cannot predict
- Need a different approach

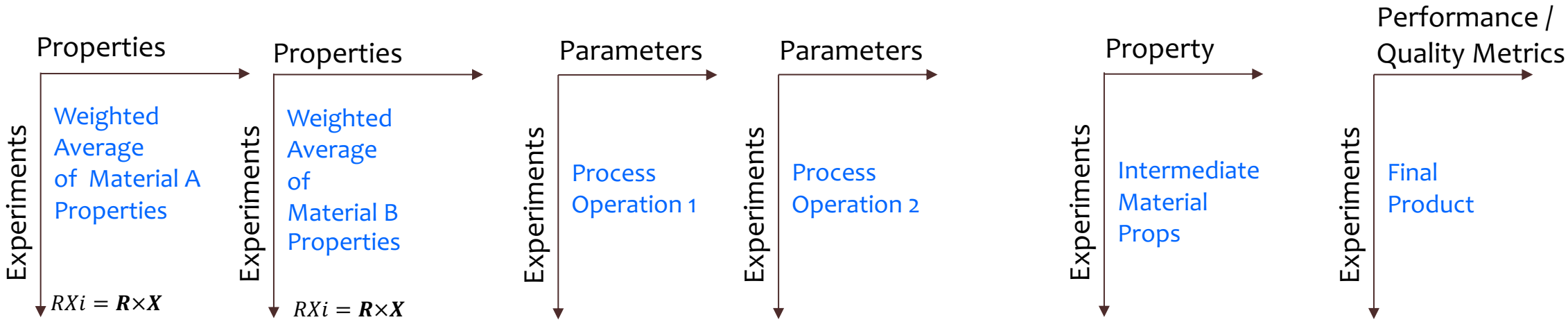
# Formulation development



# Formulation development

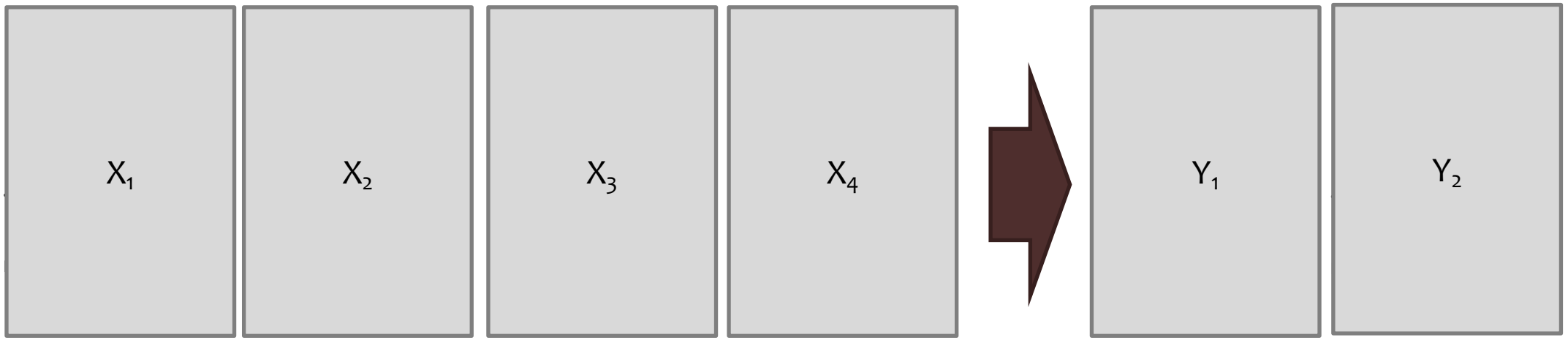


# Formulation development




- This data can be regressed with a MBPLS model

# Formulation development



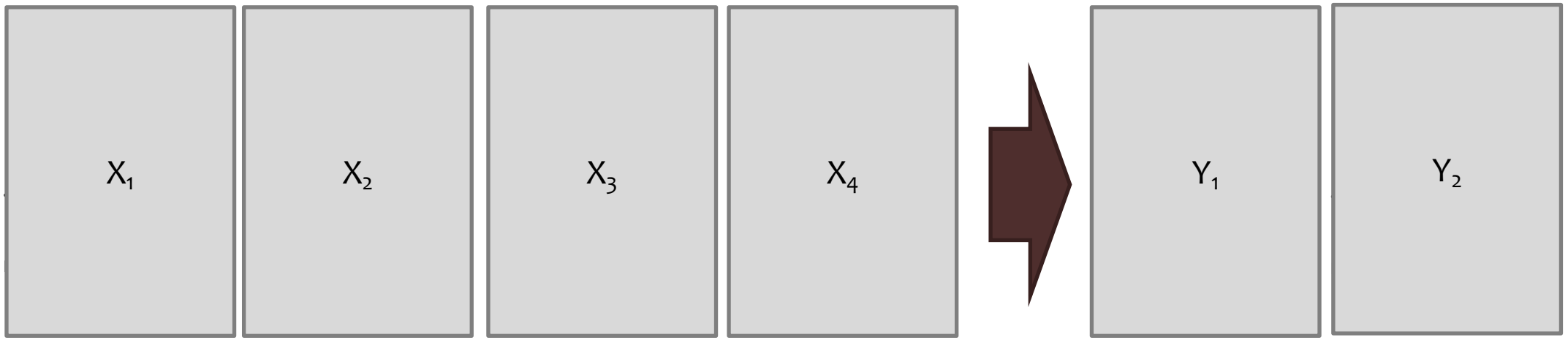

- This data can be regressed with a MBPLS model

*Analysis of Multiblock and Hierarchical PCA and PLS Models - J. Chemometrics 12, 301–321 (1998)*

# Formulation development

- Although the MBPLS model does not explicitly model the relationships across material and product characteristics, it allows the exploration of new materials.
- Given that they are characterized with the same properties as their predecessors
- Loadings are assigned to properties not directly to materials.

# Formulation development




$$Int.Mat.Propsi, Y = f(RXi_1, RXi_2, \dots, RXi_m, Process_1, Process_2, \dots, Process_n, IMPn)$$

$$RXi_m = \underbrace{\mathrm{R_m} \times Xm}$$

*This allows us to use this model formulation to predict new materials*

*Analysis of Multiblock and Hierarchical PCA and PLS Models - J. Chemometrics 12, 301–321 (1998)*

# Formulation development

- Mining data from previous experiments.
  - Modeling data from previous products
  - **In-silico formulation development**
  - Surrogate selection for experimental design
  - Including material variability
    - When you have access to materials
    - When you don’t

# Formulation development

- Once we have a predictive model where

$$Int.Mat.Propsi, Y = f(RXi_1, RXi_2, \dots, RXi_m, Process_1, Process_2, \dots, Process_n, IMPn)$$

$$RXi_m = \mathrm{R_m} \times Xm$$

- The available inputs to the model are:
  - $\mathrm{R_m}, Process_1, Process_2, \dots, Process_n$

Choices of materials

# Formulation development

- How to use such model in developing a new formulation:

A. Human driven trial and error: generate different choices of materials and ratios ($R_i$) and process conditions and examining the model predictions.

B. Use optimization methods.

# Formulation development

- Optimization based formulation
  - Components of an optimization formulation
    1. **Objective:** Quantitative scalar that we seek to minimize or maximize
    2. **Degrees of freedom:** Variables that the optimizer will use to achieve its objective
    3. **Constraints**: Conditions that the optimal solution must comply with ($=, \geq or \ \leq statements$)

# Formulation development

- Optimization based formulation
  - Illustrative Example
    1. **Objective:** maximize dissolution at 30 min
    2. **Degrees of freedom:** Choice of disintegrant and lubricant and percentages in formula and tablet press conditions.
    3. **Constraints**:
       1. The Model [*this is what relates the df with the obj*]
       2. Min. hardness ≤ Tablet hardness ≤ Max. hardness
       3. Tablet Weight RSD ≤ 2 %
       4. Formulation needs to add up to 100%

# Formulation development

- Optimization based formulation
  - How do optimizers work?
    - Gradient based:  will use derivatives (or approximation of the derivatives) of the objective function with respect to the degrees of freedom to find the optimal solution.
    - Blind search (popular with ML): Use of extensive sampling to search the solution space and find a better solution than the initial guess.

# Formulation development

- Optimization based formulation
  - If there are multiple objectives
    - One solution is to use the sum of the weighted objectives
  - The mathematical formulation of the problem is very important
    - "Choosing materials" is much better when binary variables are involved in the formulation
    - Use realistic constraints

# Formulation development

- Actual example from Computers and Chemical Engineering 60 (2014) 396– 402

$$\min\left(\sum_{n}(\hat{y}_i(n)-y_i^{\text{target}})^2\right)$$

$s.t.$

$$\hat{\mathbf{y}}(n)=\mathbf{Q}\boldsymbol{\tau}_{\text{new}}(n)$$

$$\boldsymbol{\tau}_{\text{new}}(n)=\mathbf{W}^{*T}\left[\mathbf{z}^T(n)\mathbf{rxi}^T(n)\right]^T$$

$$\mathbf{rxi}^T(n)=[\mathbf{rxi}_{api}^T(n)\mathbf{rxi}_{ex1}^T(n)\mathbf{rxi}_{ex2}^T(n)\mathbf{rxi}_{ex3}^T(n)\mathbf{rxi}_{ex4}^T(n)]$$

$$spe_X(n)=\sum\left([\mathbf{z}^T(n)\mathbf{rxi}^T(n)]^T-\mathbf{P}\boldsymbol{\tau}_{\text{new}}(n)\right)^2$$

$$\text{Hot}\,T^2(n)=\sum_{a=1}^{A}\left(\frac{\tau\text{new}a(n)}{\sigma_a}\right)^2$$

$$spe_X(n)\le spe\_\text{upper}\_\text{lim}\,it$$

$$\text{Hot}\,T^2(n)\le \text{hot2}\_\text{upper}\_\text{lim}\,it$$

$$z_l(n)\le z\_max_l$$

$$z_l(n)\ge z\_min_l$$

$$z_l(n)=z\_\text{fixed}_l$$

$$\hat{y}_i(n)\le y\_max_i$$

$$\hat{y}_i(n)\ge y\_min_i$$

$$\forall j=[api,ex1,ex2,ex3,ex4]$$

$$\mathbf{rxi}_j(p_j,n)=\sum_{m_j}\mathbf{r}_j(m_j,n)\mathbf{x}_j(p_j,m_j)$$

$$\mathbf{r}_j(m_j,n)\le\mathbf{rbinary}_j(m_j,n)$$

$$\sum_{m_j}\mathbf{r}_j(m_j,n)=1$$

$$\mathbf{mass}_j(m_j,n)=\mathbf{r_j}(m_j,n)\times\text{mass\_required}_j$$

$$\sum_{n}\mathbf{mass}(m_j,n)\le\text{mass\_available}(m_j)$$

$$\sum_{m_j}\mathbf{rbinary}_j(m_j,n)\le\text{max\_num\_lots\_to\_blend}_j$$

# Formulation development

- Examples, references
  - *International Journal of Pharmaceutics 418 (2011) 235– 242*
  - *Ind. Eng. Chem. Res. 2012, 51, 12886–12900*
  - *Ind. Eng. Chem. Res. 2013, 52, 5934–5942*
  - *Ind. Eng. Chem. Res. 2013, 52, 8260–8271*
  - *Computers and Chemical Engineering 60 (2014) 396– 402*
  - *Chemical Engineering Research and Design 92 ( 2014) 534–544*

# Formulation development

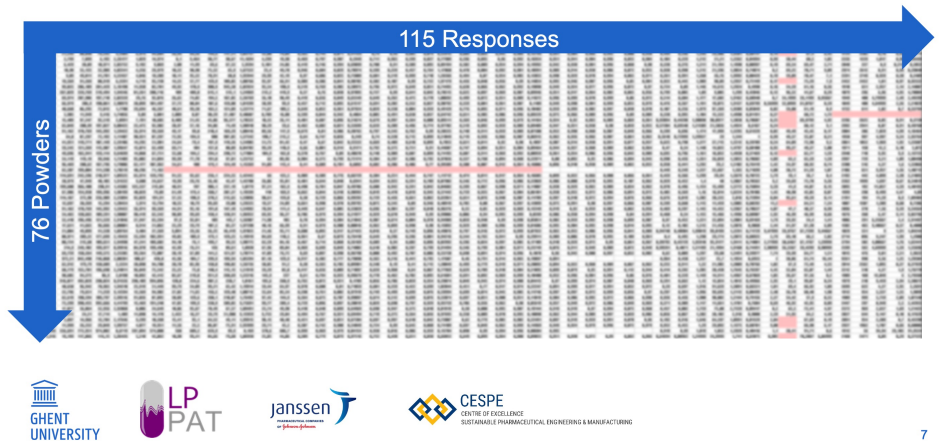
- Mining data from previous experiments.
  - Modeling data from previous products
  - In-silico formulation development
  - **Surrogate selection for experimental design**
  - Including material variability
    - When you have access to materials
    - When you don’t

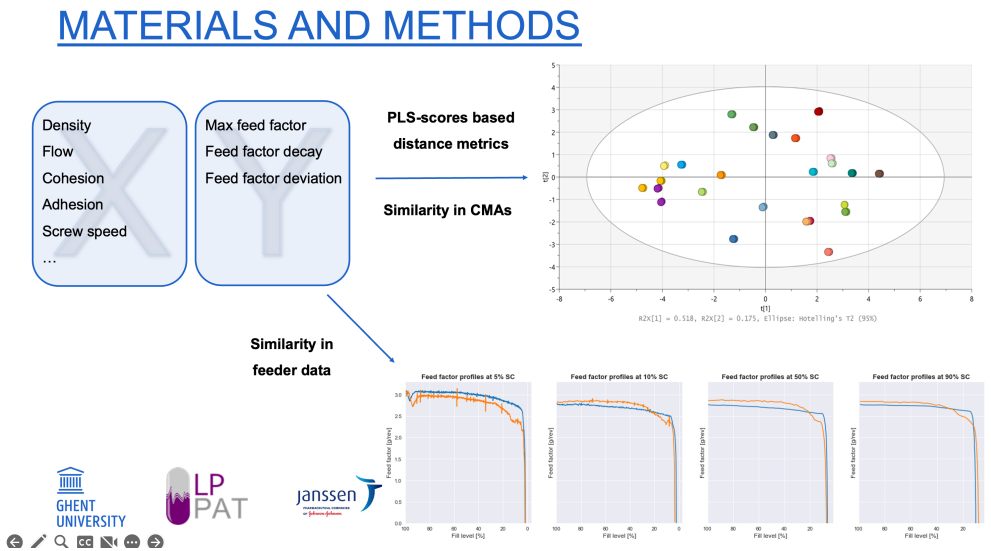
# Formulation development

- **Surrogate selection**: Choose an alternative (more accessible, safe or cost effective) material that is similar to the drug to carry out development work.
- Great idea!
  - Define similar?
  - With respect to what ?
  - How similar is “similar enough” ?

# Formulation development

- All interesting questions being explored by the research group of Prof. Thomas DeBeer @Ghent University.



# Formulation development

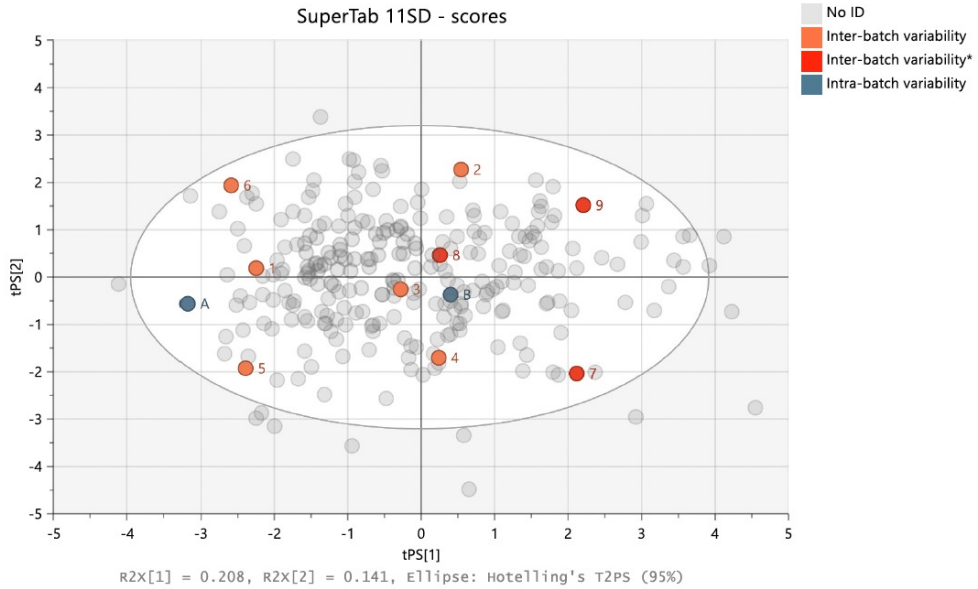
- Mining data from previous experiments.
  - Modeling data from previous products
  - In-silico formulation development
  - Surrogate selection for experimental design
  - Including material variability
    - When you have access to materials
    - When you don’t

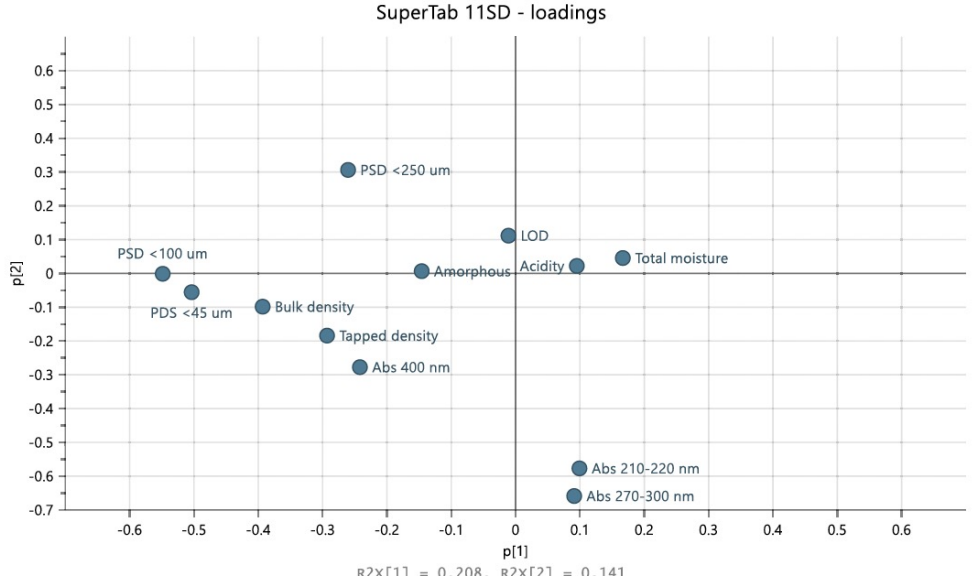
# Formulation development

- Material variability is perhaps one of the most challenging risks to address at the R&D stage.
  - Not many lots are consumed in development
  - Vendors are getting involved and greatly helping their customers address this question.

# Formulation development

- Use PCA to summarize material properties.
- Use scores to help select what areas of the variation need to be explored.




*Powder Technology 409 (2022) 117776*

DFE Pharma

# Formulation development

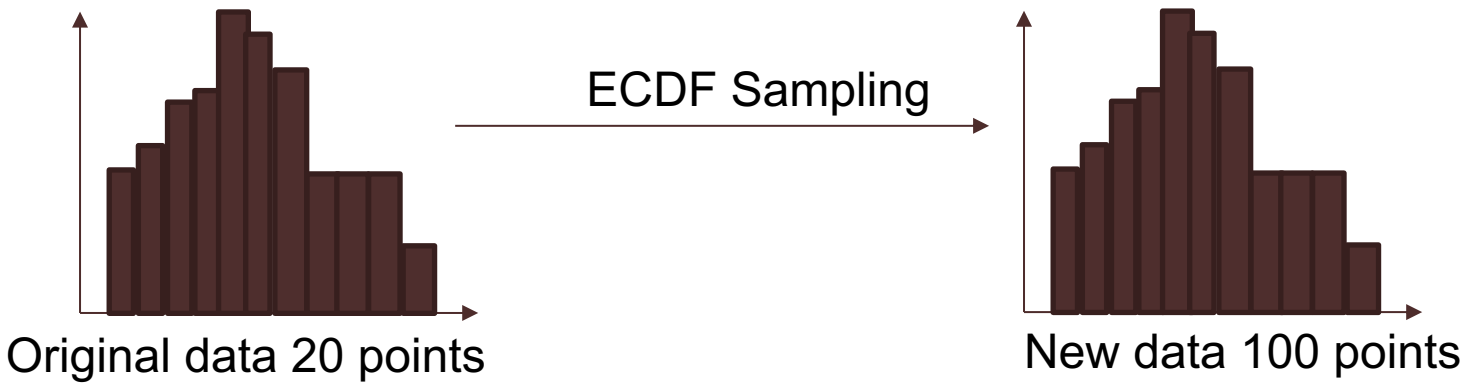
- Mining data from previous experiments.
  - Modeling data from previous products
  - In-silico formulation development
  - Surrogate selection for experimental design
  - Including material variability
    - When you have access to materials
    - When you don’t

# Formulation development

- Problem: We don’t have enough replicas to well represent variability.
- Objective: Create **credible clones** of the available data preserving the existing correlation across variables and the uncertainty distributions.

# Formulation development

- Cloning data using PCA and ECDF
- The Empirical Cumulative Distribution Function can be used to draw new samples from a population while preserving the population distribution.

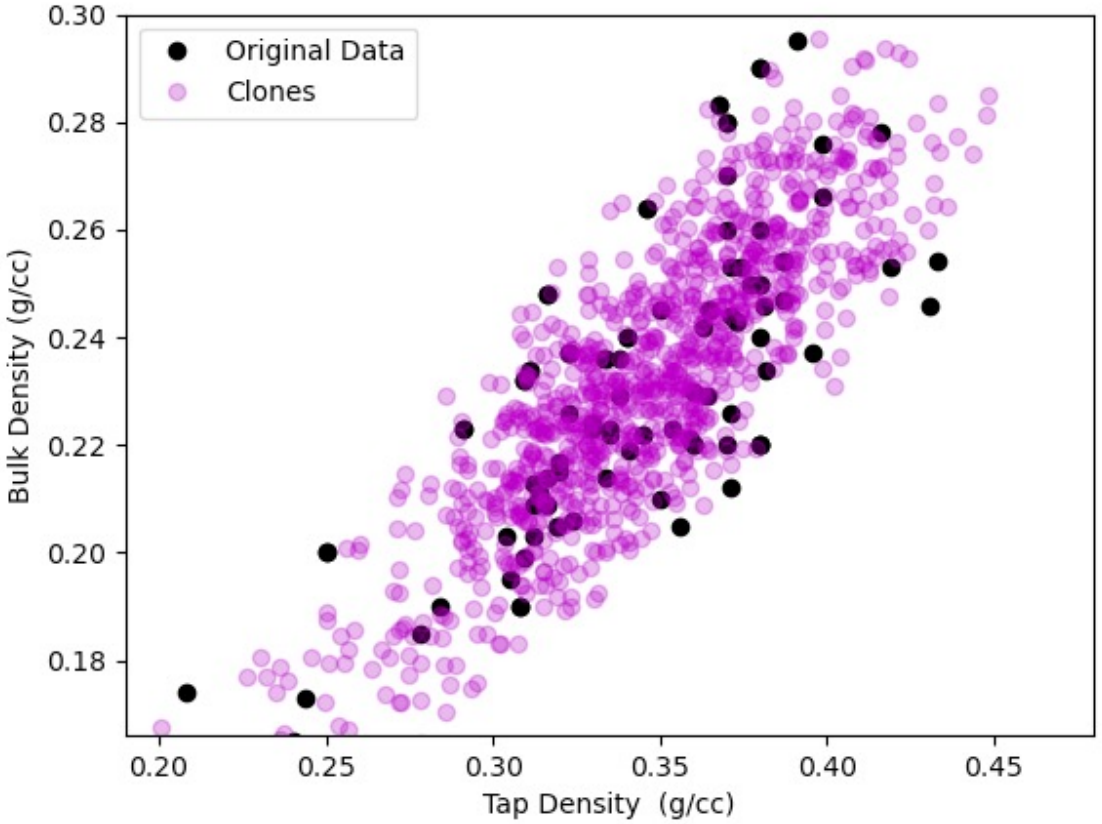


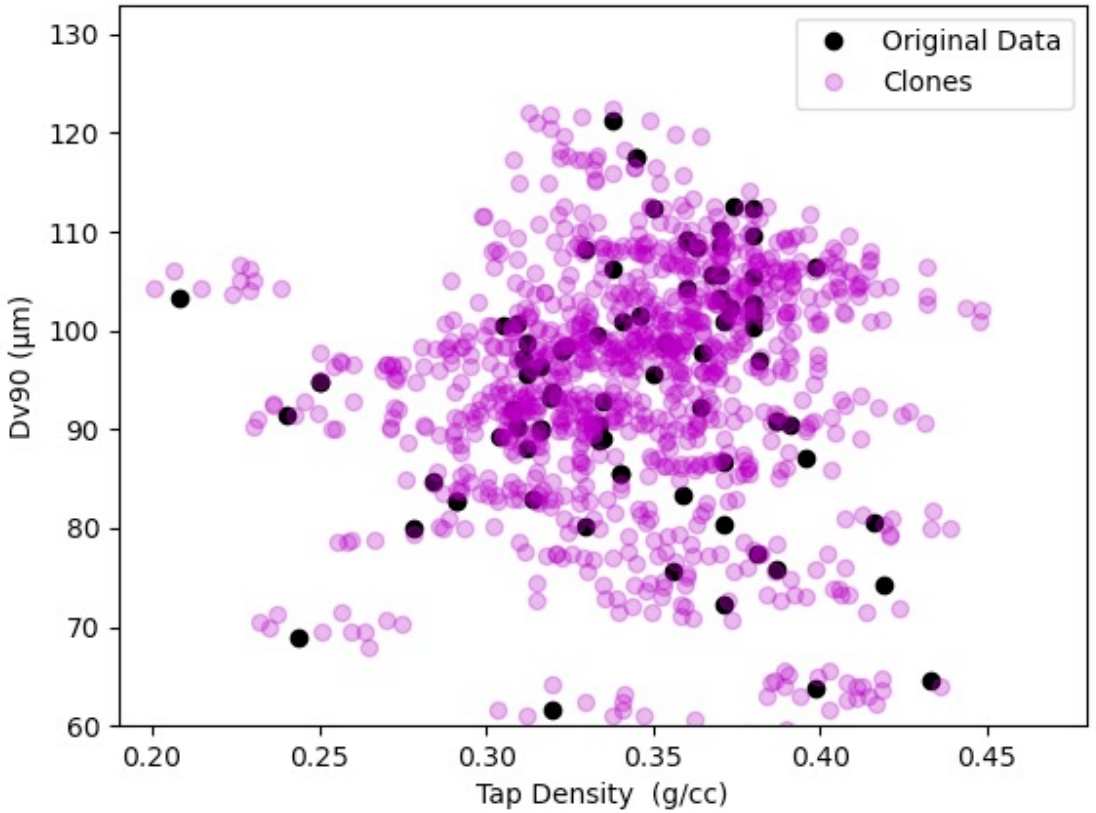
# Formulation development

1. Perform PCA on X
2. Create a copy of the reconstructed portion of $\hat{X} = TP^T$
3. To each column of $\hat{X}$ add new residuals, sampled from the population of the original residuals using ECDF.
4. Ready.

# Formulation development

- Example



# Agenda

- Formulation development
- Process development
- Process de-risking and transfer
- Process Monitoring
  - MSPC for fault detection and isolation

# Process Development

- Computational tools:
  - Modeling of feeders
  - Modeling of mixing extent and residence time distribution (RTD)

# Modeling of loss in weight feeders

- Addressed in plenty of papers in literature:
  - *Yu, Y. Theoretical modelling and experimental investigation of the performance of screw feeders. Ph.D. Thesis, University of Wollongong, 1997.*
  - *Boukouvala et al.,Computers and Chemical Engineering 42 (2012) 30–47*
  - *Rogers et al. Ind. Eng. Chem. Res. 2014, 53, 13, 5128−5147*
  - *Jia, J. Pharm. Innovation 2009, 4, 174–186.*
  - *Bascone, Industrial & Engineering Chemistry Research, 59(14), pp.6650-6661.*
  - *Johnson, International Journal of Pharmaceutics, 621, p.121776.*

# Modeling of loss in weight feeders

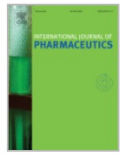
- Most recent paper from RCPE
  - Studied the process with multiple modeling approaches.

ELSEVIER

International Journal of Pharmaceutics

Available online 2 April 2023, 122915

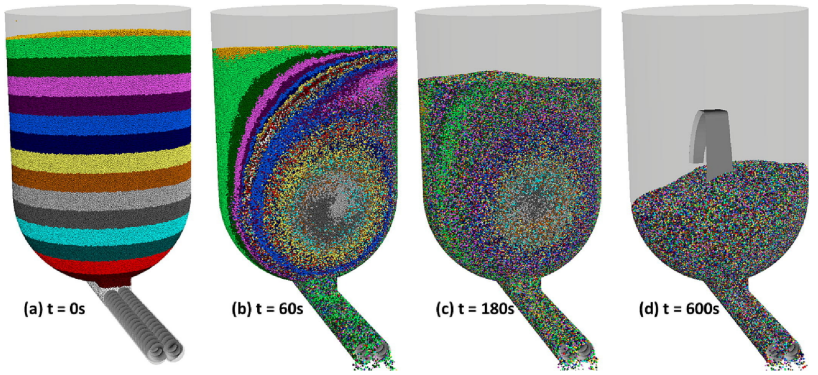
In Press, Journal Pre-proof ⓘ What's this? ↗



"Even a very simple model that assumes perfect mixing inside the hopper is a decent approximation of the real dynamics. This model works well for the given twin-screw feeder geometry because the agitator mixes a large portion of the material inside the feeder. Different feeder designs agitate different portions of the hold-up mass and thus have different material survival functions. The perfect mixing model is not a universal law across all feeders."

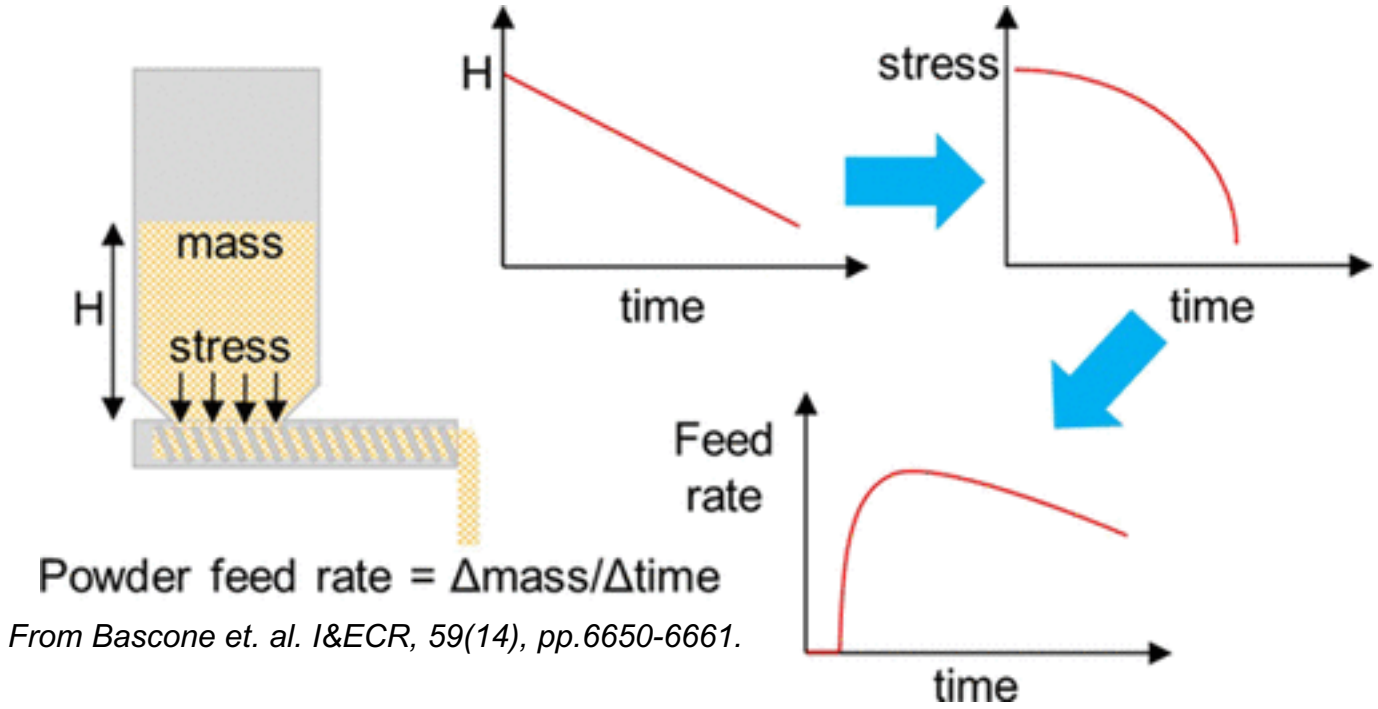
## A DEM Model to Evaluate Refill Strategies of a Twin-Screw Feeder

Peter Toson [a] ✉, Johannes G. Khinast [a b] ✉



# Modeling of loss in weight feeders

- Mixing behavior apart, the inherent challenge in modeling loss in weight feeders is estimating the densification of material in the base of the hopper as a function of hold up.




*From Bascone et. al. I&ECR, 59(14), pp.6650-6661.*

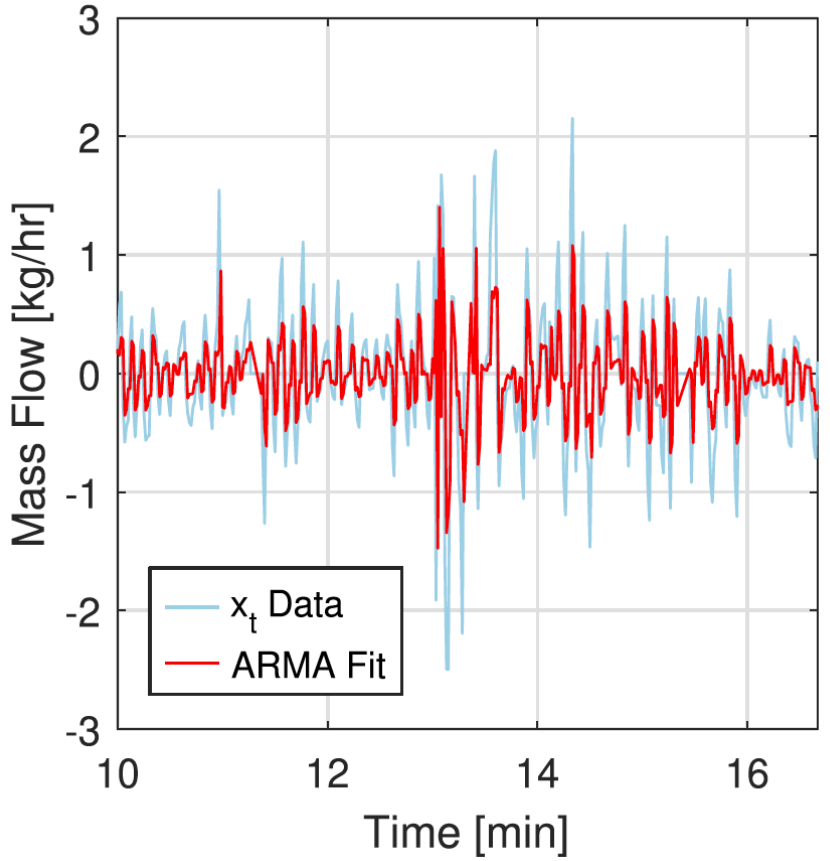
# Modeling of loss in weight feeders

- Some work has been done to predict this relation through the estimation of the feed-factor from powder properties.

  - Bostijn, N.Int. J.Pharm. 2019, 557, 342–353.
  - Wang, Powder Technol. 2017,308, 135–148.

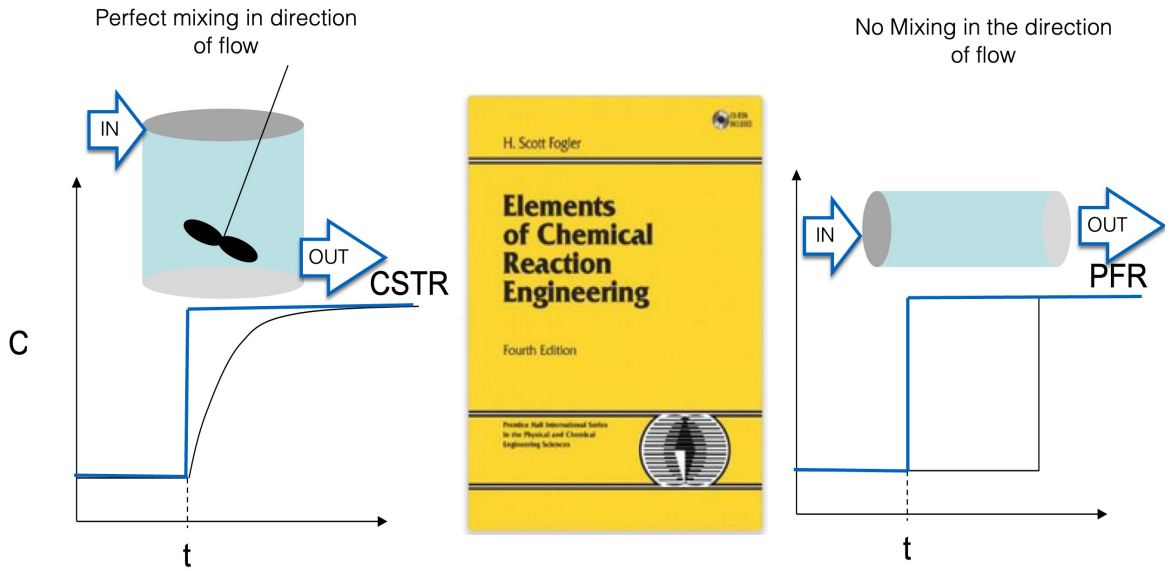
# Modeling of loss in weight feeders

- Others have focused on characterizing and predicting the stochastic behavior of a feeder in gravimetric mode.
- More work is needed in this area to generate a truly forward-predictive model of the feeding behavior expected for a new material



# Modeling the Residence Time Distribution

- Very large amount of publications in this topic.
- Application of reaction engineering concepts to determine the extent of mixing along the direction of flow of a powder.



# Modeling the Residence Time Distribution

- Two modeling approaches:
  - White box model
    - Calculate the speed of transit from geometry and linear speed, which implies the knowledge of a density [difficult for powders]
  - Black box model:
    - Fit the RTD curve to time explicit functions

# Modeling the Residence Time Distribution

- Two modeling approaches:
  - White box model

    *García-Muñoz, S. AIChE Journal, 64(2), pp.511-525.*
  - Black box model [most recent development]:

    *Toson and Doshi, Processes 2019, 7, 615*

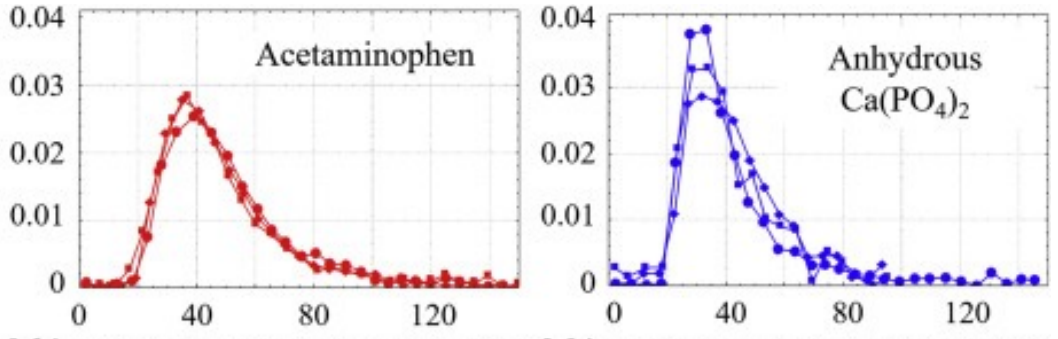
$$\mathrm{RTD}_{n,\tau}(t) = \frac{t^{n-1}}{\Gamma(n)} \cdot \left(\frac{n}{\tau}\right)^{n} \cdot \exp\left\{-\frac{t\,n}{\tau}\right\}$$

# Modeling the Residence Time Distribution

- Either approach requires data!
- Two different experiments can be done
  - Tracer experiment:

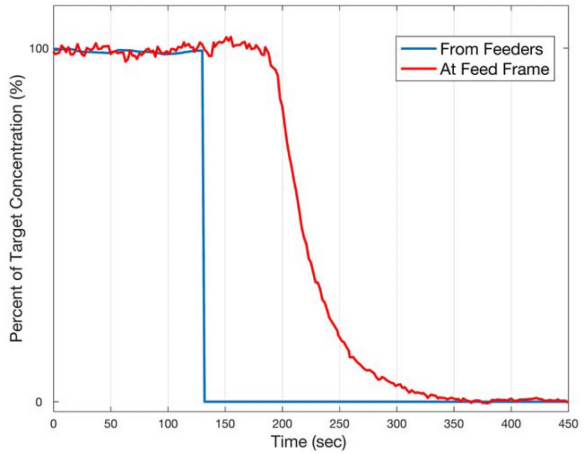
*Care is needed to select the proper tracer*

*From Escotet-Espinosa et.al., Powder Tech. (342),15,744-763,2019*




  - Step experiment:

*Cost effective and informative*




*From Garcia-Munoz et al., AICHE J. (64),2,516,2018*

# Modeling the Residence Time Distribution

- The RTD model can then produce a funnel plot

- Most Funnel plots are built as a function of fed **concentration disturbances**.
- Behavior not 100% symmetric when using a white box model.
- Behavior is 100% symmetrical with black box models

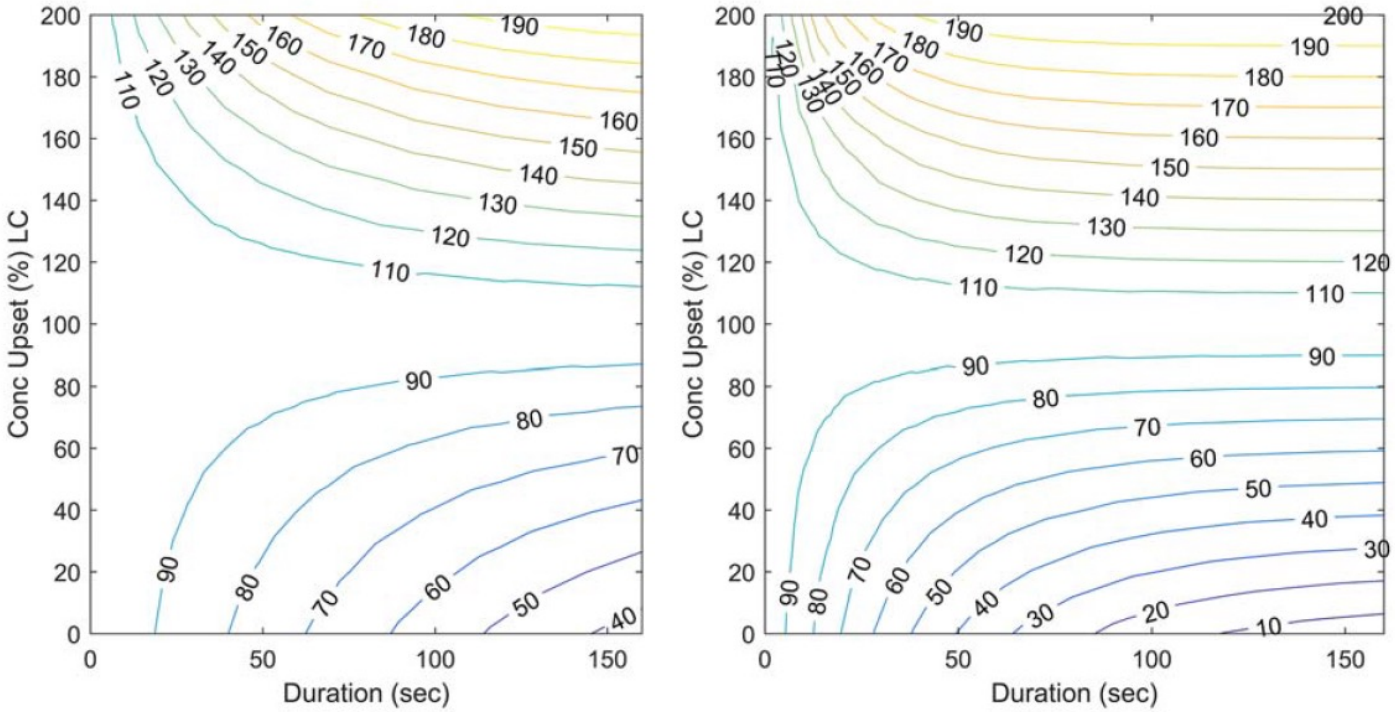



**Figure 13. Two funnel plots, for low throughput (left) and high throughput (right).**

*From Garcia-Munoz et al., AICHE J. (64),2,516,2018*

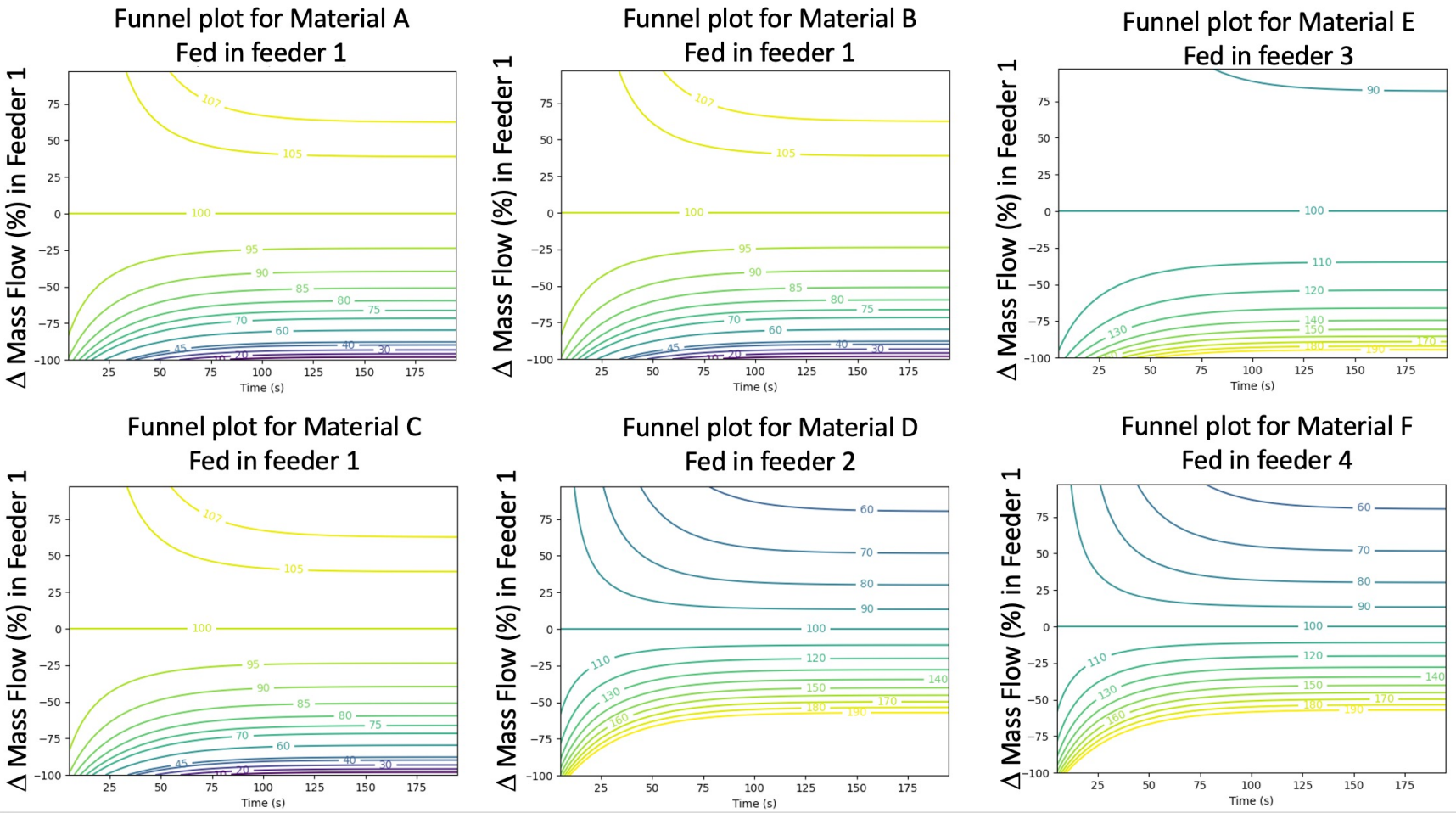
# Modeling the Residence Time Distribution

- Funnel plots as a function of mass flow disturbance

- FP are non-linear.
- and dependent on relative mass-flows from one feeder to the other.

- Feeder with greatest mass-flow has largest effect.

- Example assumes all other feeders are kept at target.



# Modeling the Residence Time Distribution

- Uses:
  - Disturbance detection and control.
  - Genealogy tracking for incoming material.

# Agenda

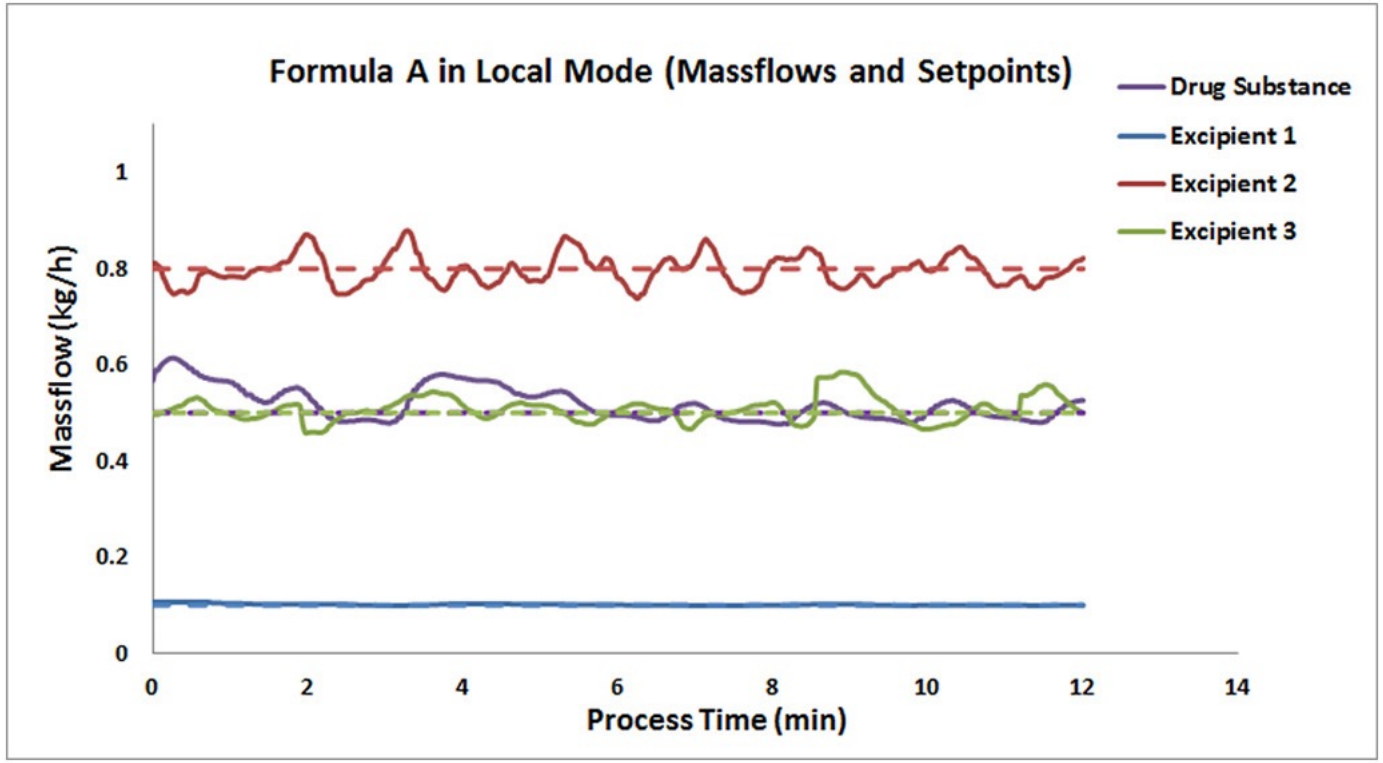
- Formulation development
- Process development
- Process de-risking and transfer
- Process Monitoring
  - MSPC for fault detection and isolation

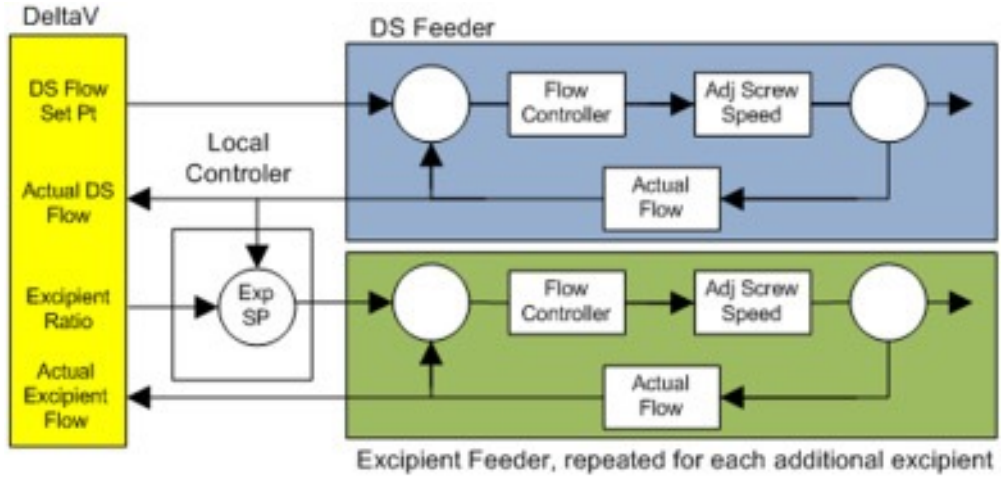
# Process de-risk and transfer

- Areas of risk are very few:
  - Main potential source of disturbances is the dispensing of materials [feeder performance]
- Mitigation strategies
  - Use ratio control
  - Quantitation of disturbance effects

# Process de-risk and transfer

- Ratio-control ensures concentration at the cost of small variation in mass-flow.

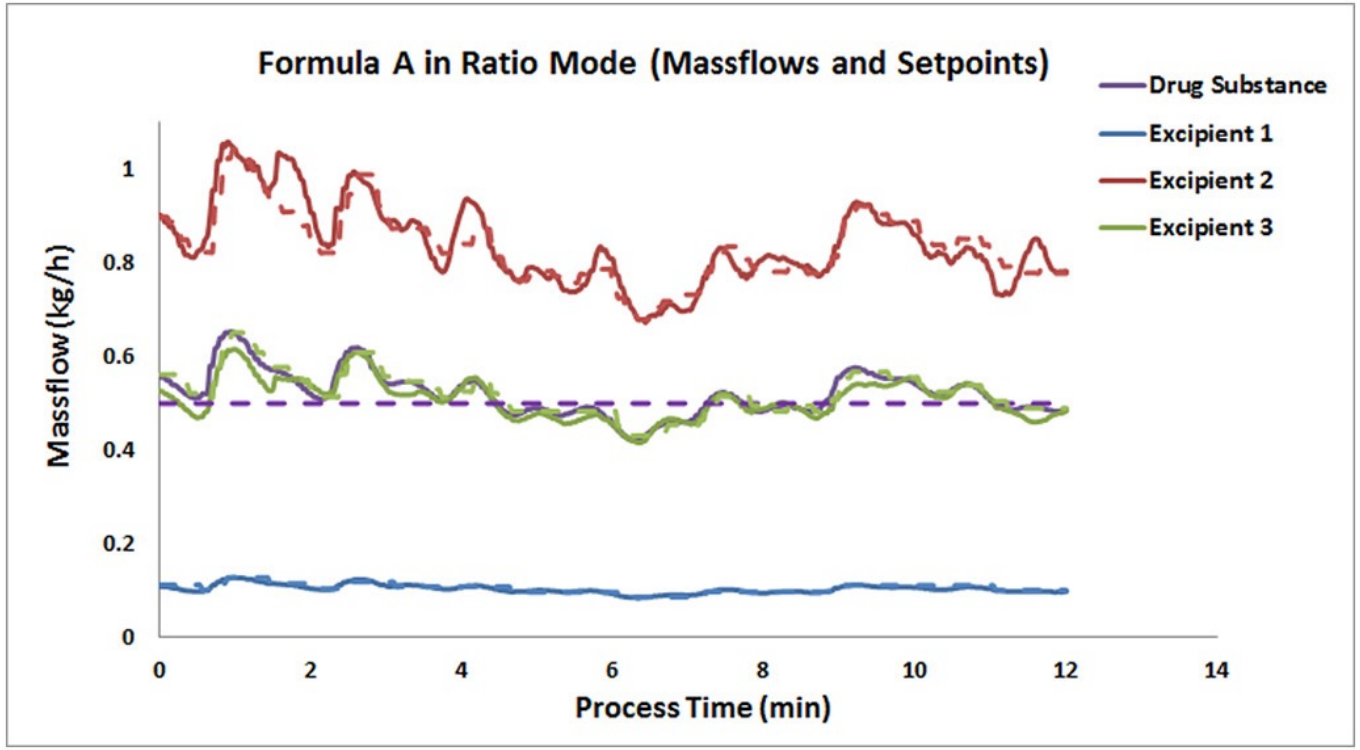


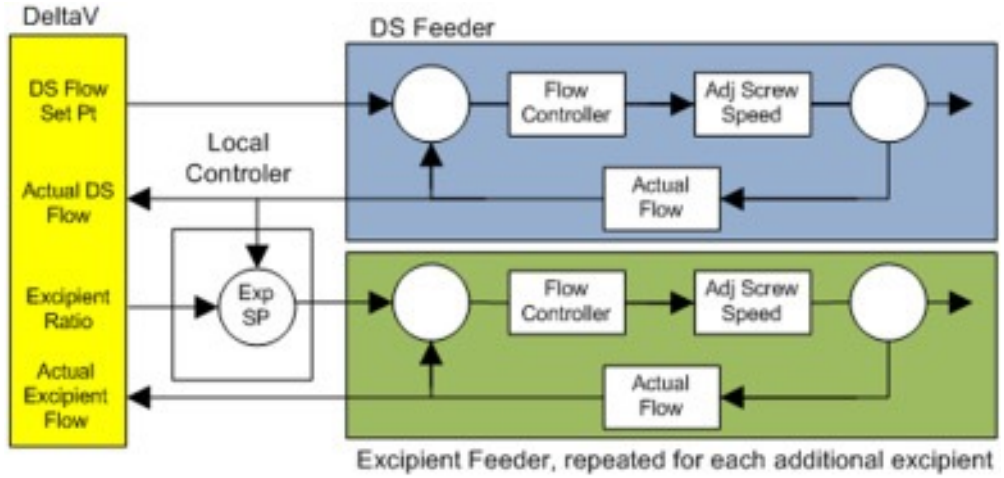




*From Hanson, Powder Tech.(331),15,236-243,2018*

# Process de-risk and transfer

- Ratio-control ensures concentration at the cost of small variation in mass-flow.




*From Hanson, Powder Tech.(331),15,236-243,2018*

# Process de-risk and transfer

- Quantitation of disturbance effects
  - Parse though data acquired during development and clinical manufacture.
  - Characterize all [even small] disturbances from set-point.
  - Place disturbances in funnel plot.

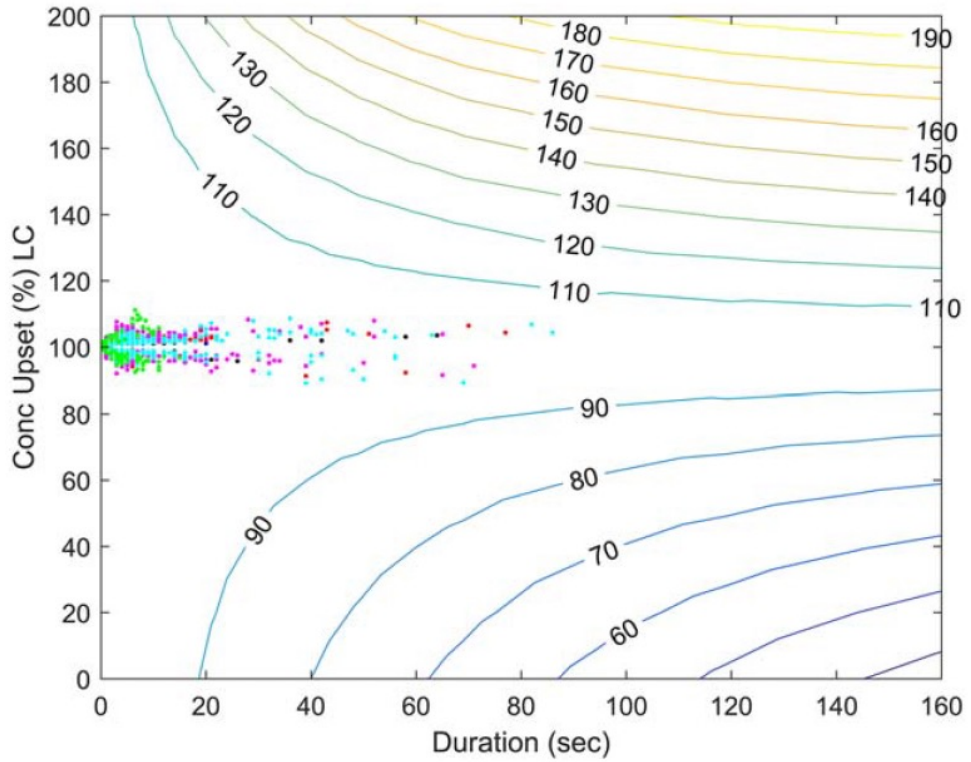



**Figure 14. Funnel plot overlay with all events (i.e., disturbances) from historical data supporting process development.**

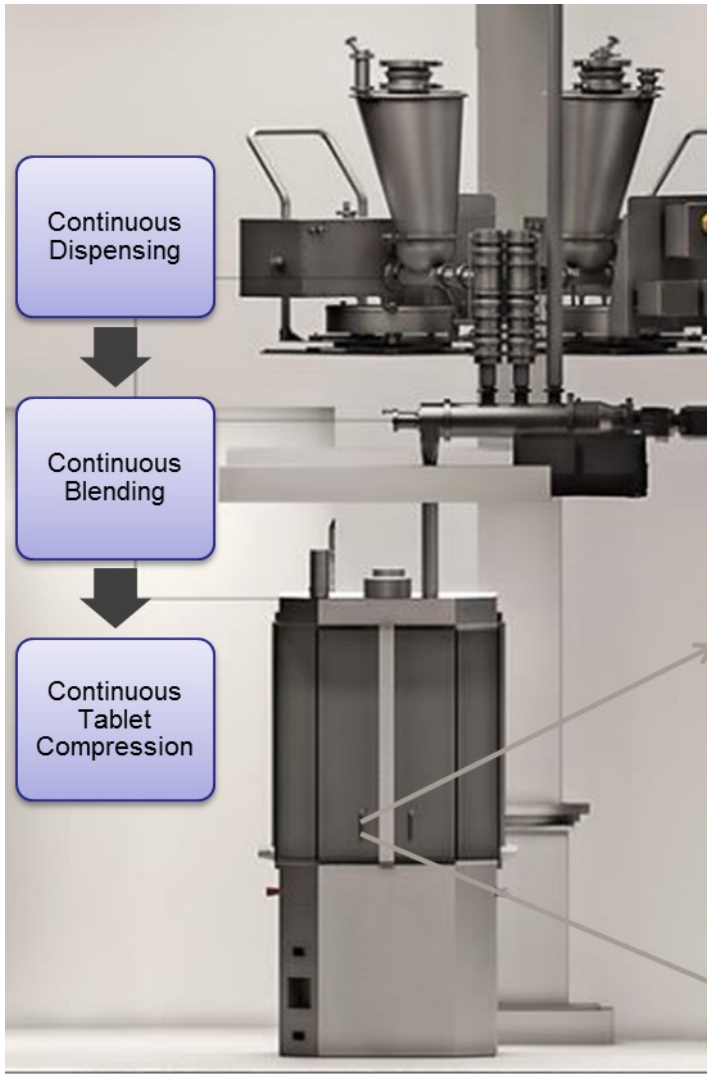
# Agenda

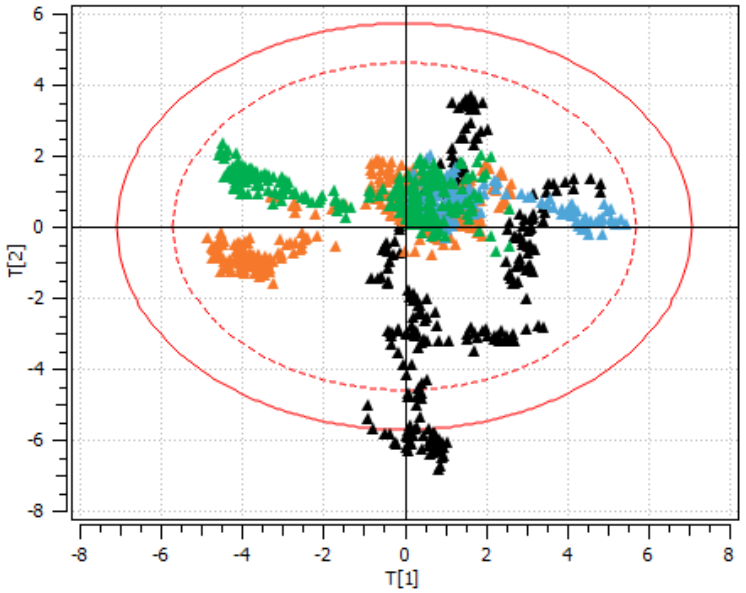
- Formulation development
- Process development
- Process de-risking and transfer
- Process Monitoring
  - MSPC for fault detection and isolation

# Process Monitoring - MSPC

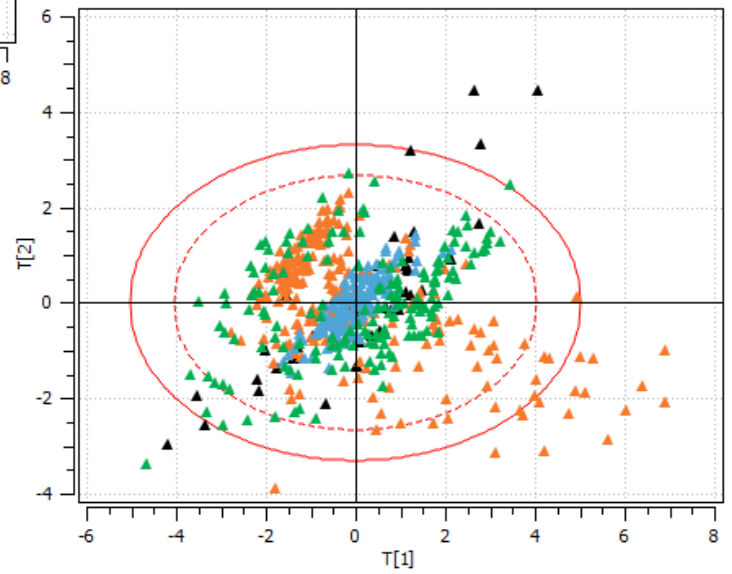
- Multivariate Statistical Process Control (MSPC) is an established method for monitoring and fault detection.
- Established in 1994.
- Plenty of software available commercially.
- Mostly implemented in bio-pharmaceutical processing.

# Process Monitoring - MSPC

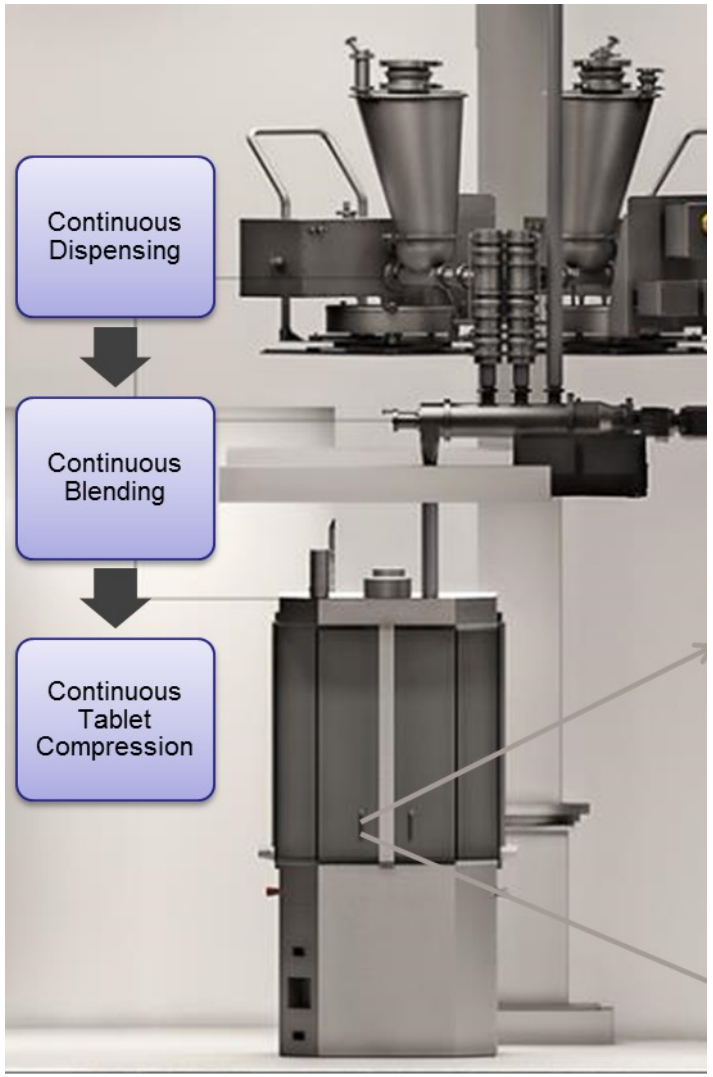



Model built with data from 6 different molecules

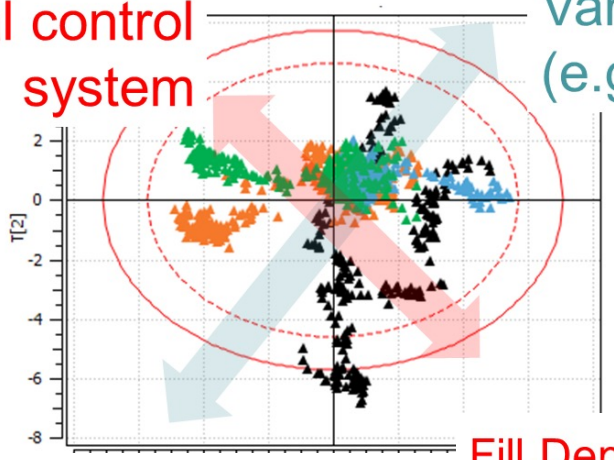


# Process Monitoring - MSPC




Variability in internal control system

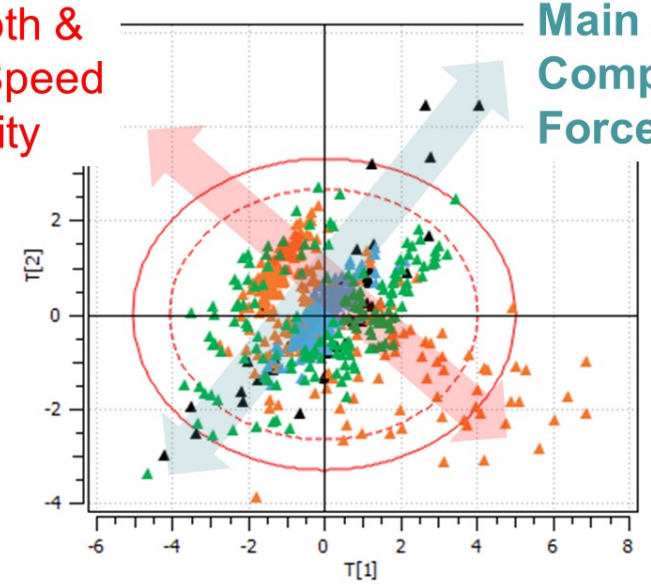
Mass Flow Variability (e.g. Start up)




Model built with data from 6 different molecules

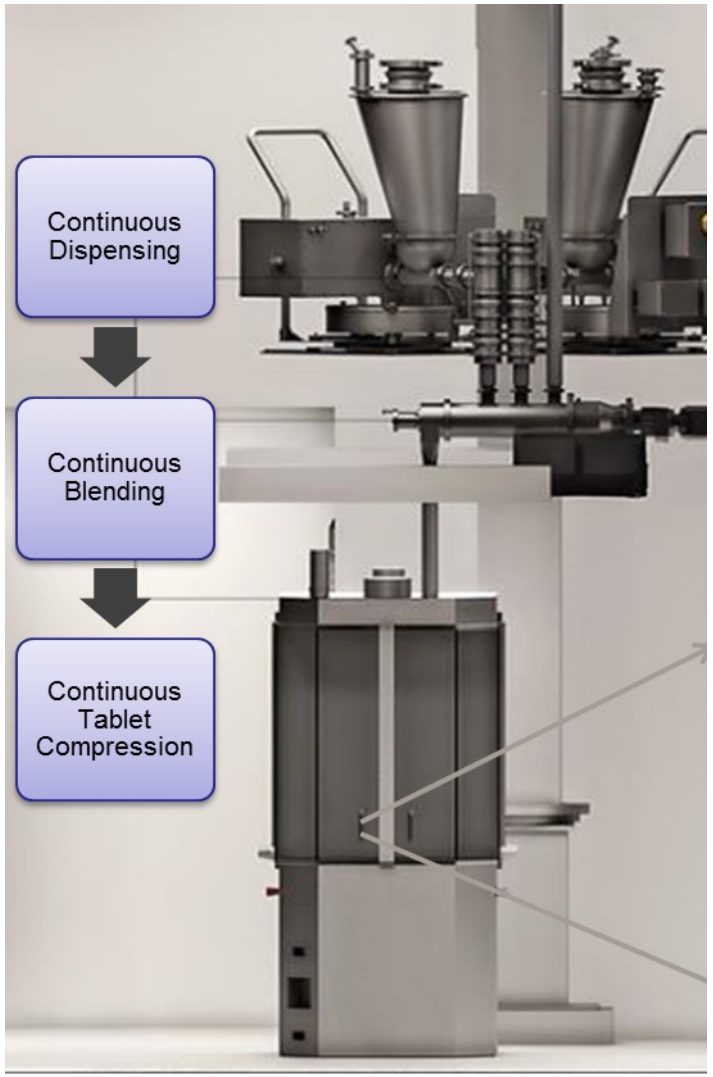
Fill Depth & Rotor Speed Variability

Main & Pre-Compression Force Variability



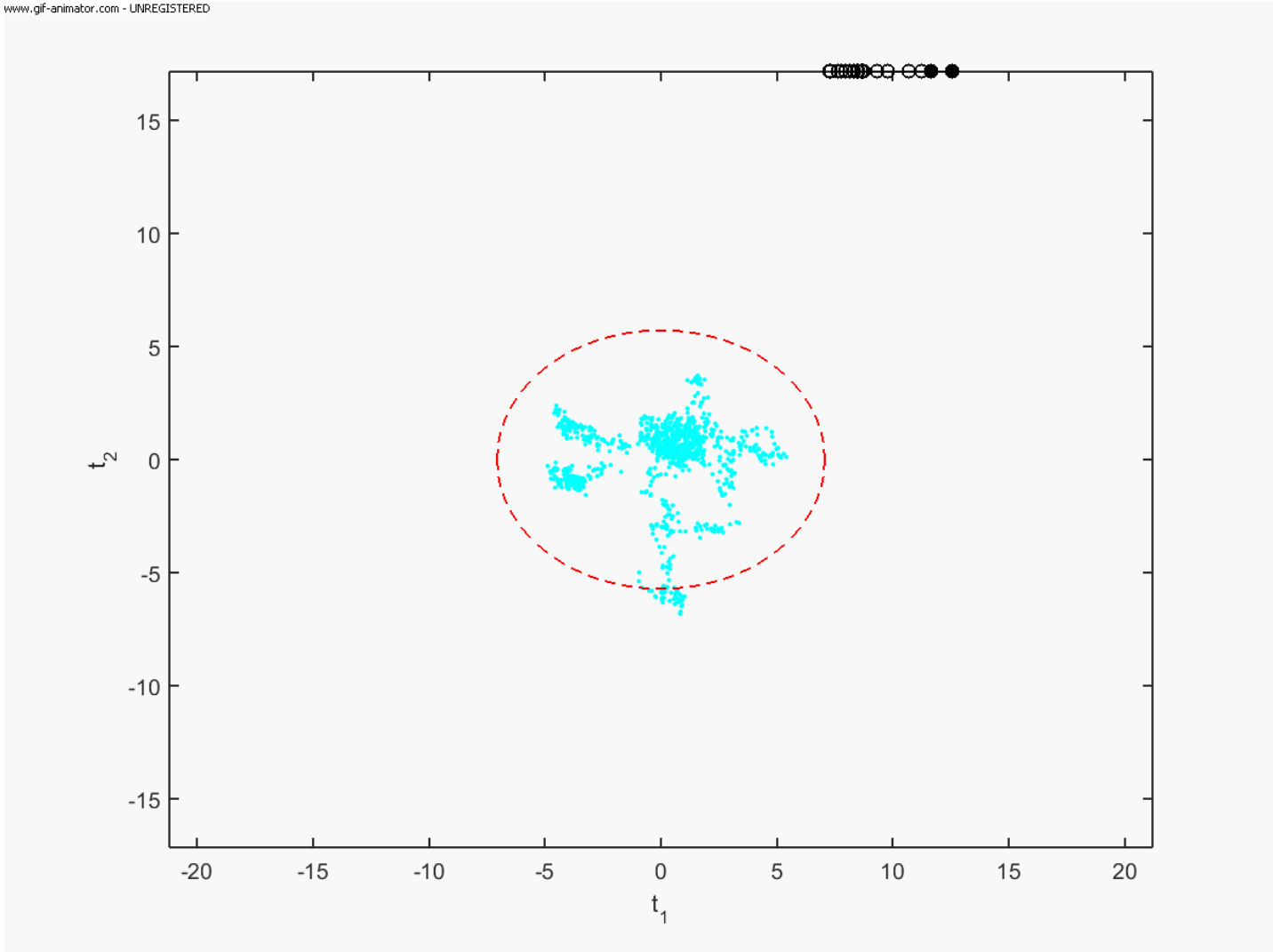

acturing

# Process Monitoring - MSPC

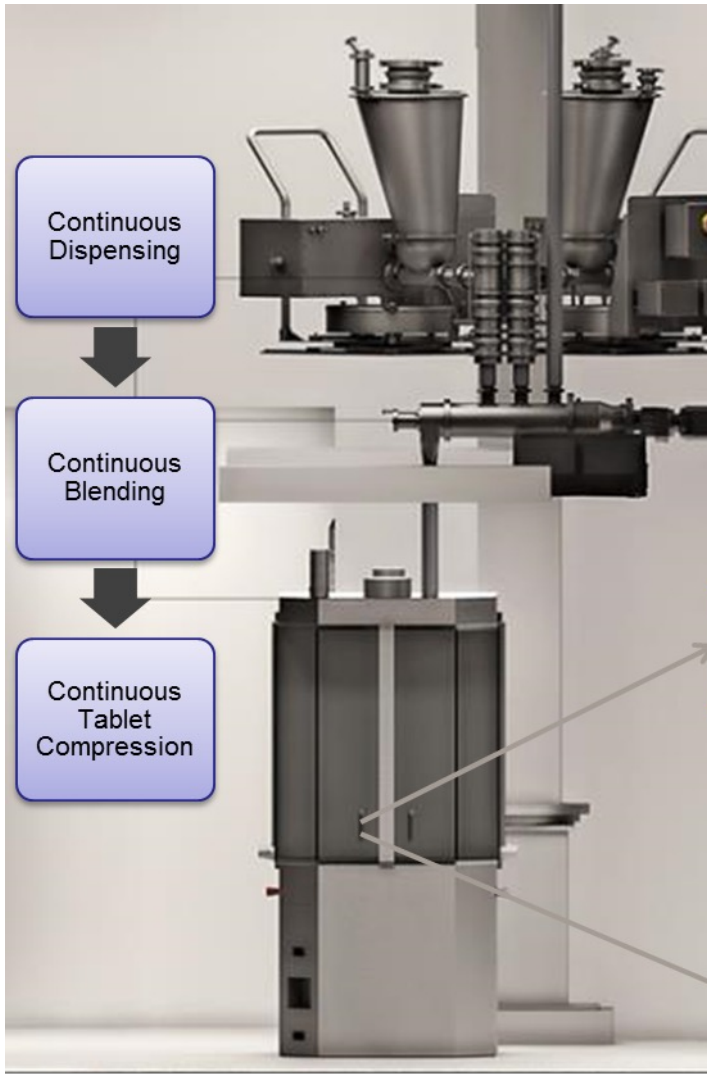


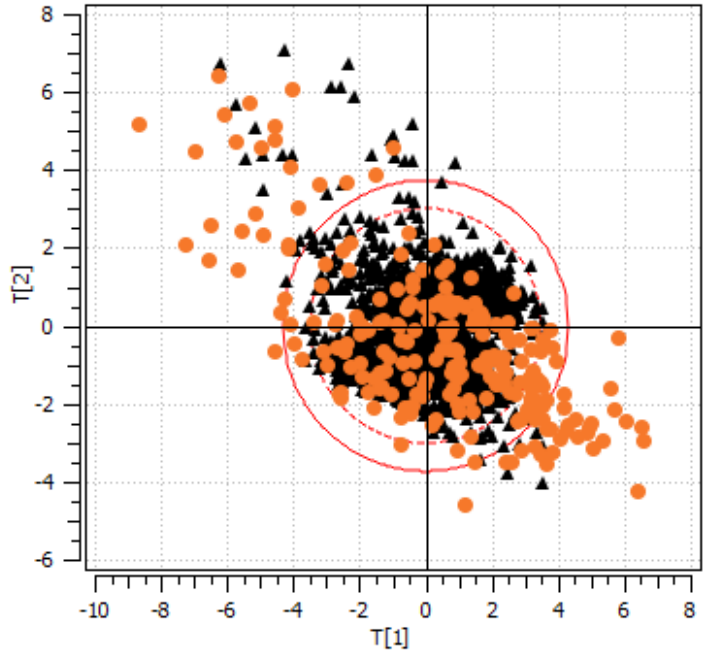

Monitoring the Startup of the Feedeers

Model built with data from 6 different molecules



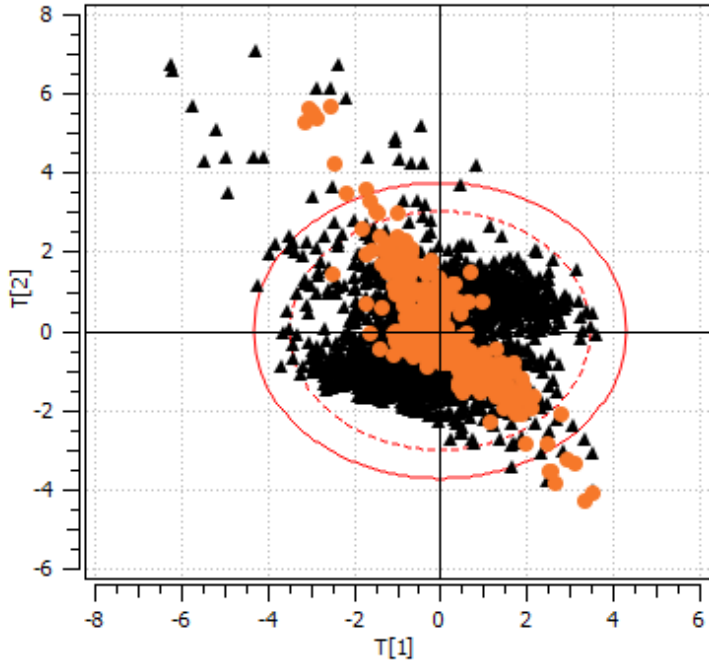
# Process Monitoring - MSPC




Same API
different formula

Model built with data from 6 different molecules




Different API and different formula

# Process Monitoring - MSPC

- Details that need close attention.
  - Management of lags
    - Variables are sampled throughout the train and the interactions are not instantaneous (hence RTD)
      - A lagged model runs at the slowest dynamics
        - No time to react
      - A non-lagged model requires more latent variables

# Process Monitoring - MSPC

- Very interesting new method:

Journal of Process Control 67 (2018) 1–11

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

## A novel dynamic PCA algorithm for dynamic data modeling and process monitoring

Yining Dong [a,c], S. Joe Qin [a,b,c,*]

[a] Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA
[b] Mork Family Department of Chemical Engineering and Material Science, University of Southern California, Los Angeles, CA 90089, USA
[c] School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, 2001 Longxiang Blvd., Longgang, Shenzhen, Guangdong, China

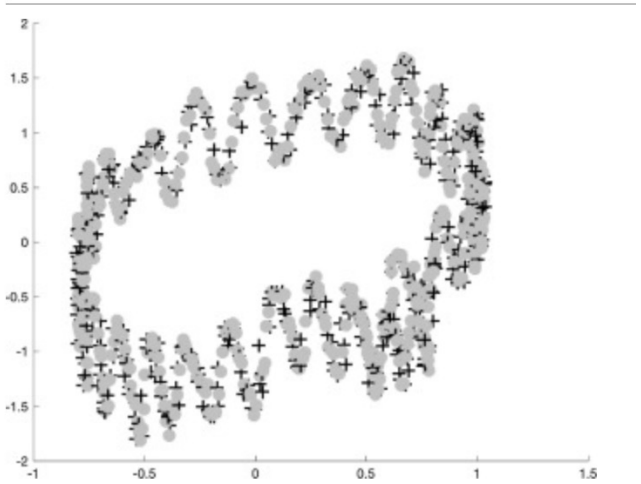
Computers & Chemical Engineering
Volume 114, 9 June 2018, Pages 69-80

## Dynamic latent variable analytics for process operations and control

Yining Dong [a b], S. Joe Qin [a b c]

# Process Monitoring

- DiPCA aims to capture latent-spaces that are auto-regressive.
  - **Explicit modeling of lags**
- Model semi-oscillatory behavior



# Final Remarks

- The systems engineering community has and continues to develop useful computational approaches that can be exploited to accelerate product development and de-risk process operations.
- Continuous manufacturing is very amenable to the implementation of these tools.

# Thank you

- Let’s talk!

sal.garcia@lilly.com