Process/model mismatch diagnosis by latent variable modeling

N. Meneghetti, P. Facco, F. Bezzo, M. Barolo

UNIVERSITÀ DEGLI STUDI DI PADOVA

CAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering
University of Padova (Italy)

1. MOTIVATION AND OBJECTIVE

- **The problem**: when a mismatch is found by comparing an available dataset to the predictions of a **first principles model (FP)** it may not be trivial to identify the cause for the **process/model mismatch (PMM).**
- **Objective**: **diagnosis of a PMM** using **historical data** and a data-based **DB model.**
- **Challenges**: uncertainties on several parameters of the models; limited plant data available; high non-linear correlations between the variables involved.
- **Strategy**: a **DB model** (Principal Component Analysis, **PCA** [1]) is used to assess the consistency between the correlation structure of a historical operation dataset and that of a similar dataset generated using the FP model.

2. CASE STUDY : MILLING PROCESS

A simulated **milling process** for the size reduction of a granular polymer is used as a case study. The FP model includes mass and population balances of the solid distributed phase. The population balance equation on mass basis for phase *p* is [2]:

$$\frac{\partial M_p(y,t)}{\partial y} = \int_0^{y_{max}} P_{B,p}(z) b_p(y,z) M_p(z,t)dz - P_{B,p}(y) M_p(y,t)$$

The empirical formulations suggested by Vogel and Peukert [2] for the grinding rate selection function $P_{B,p}$ and the breakage function B_p have been used:

$$B_p = \left(\frac{z}{y}\right)^q \frac{1}{2}\left(1+\tanh\left(\frac{y-y'}{y'}\right)\right), \qquad \frac{\partial B_p(z,y)}{\partial y} = b_p(z,y)$$

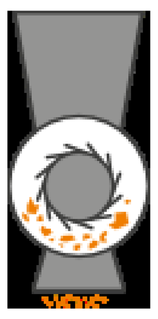
$$P_{B,p} = 1 - \exp\left(-f_{\mathrm{Mat}} zk\left(W_{m,\mathrm{kin}} - W_{m,\mathrm{min}}\right)\right)$$

where $P_{B,p}$ and B_p depend on several parameters (f_{Mat}, $W_{m,kin}$, $W_{m,min}$, q) specific of the type of material involved. The modelling package gSOLIDS® 3.0 [3] was used as a simulation tool to obtain the historical dataset.

The historical and simulated datasets have been calculated, considering **15** different steady states.

INPUTS

- **W** (mass flowrate)
- $\mathbf{D_{in}}$ (mean particles diameter of PSD_{in})
- σ_{in} (standard deviation of PSD_{in})
- ρ_{bulk} (bulk density)
- v (rotational velocity)
- 4 polymers (f_{Mat}, $W_{m,kin}$, $W_{m,min}$, q)



FP MODEL

HISTORICAL OUTPUTS: PSD_{out}

SIMULATED OUTPUTS: PSD_{out}

3. DEVELOPMENT OF THE DATA-BASED MODEL

In order to simulate a PMM, erroneous values of the **parameter f_{Mat}** (related to the strength of the material) was introduced in the FP model. The methodology proposed includes 4 steps:

1. **Auxiliary data designation.** For each sample, inputs, outputs and parameters are combined to obtain 2 sets of **V auxiliary variables** concatenated to form a model matrix and a process matrix.

INPUTS + HISTORICAL OUTPUTS →

$$x_1 = f_1(P_{B,p}) \quad x_5 = f_5(B_p, P_{B,p}, PSD_{in}, \rho)$$
$$x_2 = f_2(PSD_{in}) \quad x_6 = f_6(P_{B,p}, PSD_{in}, \rho)$$
$$x_3 = f_3(B_p) \quad x_7 = f_7(PSD_{out})$$
$$x_4 = f_4(B_p, P_{B,p})$$

→ $\mathbf{X_M}$ [15×V] PROCESS MATRIX

INPUTS + SIMULATED OUTPUTS →

$$x_1 = f_1(P_{B,p}) \quad x_5 = f_5(B_p, P_{B,p}, PSD_{in}, \rho)$$
$$x_2 = f_2(PSD_{in}) \quad x_6 = f_6(P_{B,p}, PSD_{in}, \rho)$$
$$x_3 = f_3(B_p) \quad x_7 = f_7(PSD_{out})$$
$$x_4 = f_4(B_p, P_{B,p})$$

→ $\mathbf{X_\Pi}$ [15×V] MODEL MATRIX

- The presence of the population balance requires the discretization of integral term of the mass balance. The size range considered, has been partitioned into 40 bins (corresponding to a specific particle size). The mass balance must to be solved for each bin, consequently, $\mathbf{X_M}$ and $\mathbf{X_\Pi}$ become 3-D matrices [15×7×40].

2. **DB model development.** Both matrices are autoscaled on the mean and standard deviation of $\mathbf{X_M}$. A multi-way PCA (MPCA,[4]) model is built from $\mathbf{X_M}$ and the residuals matrix $\mathbf{E}_M$ is calculated from:

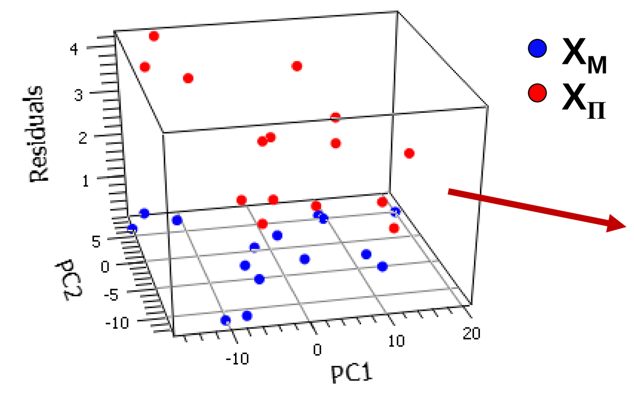
$$\hat{\mathbf{X}}_M = \mathbf{T}_M \mathbf{P}_M^T \qquad \mathbf{X}_M - \hat{\mathbf{X}}_M = \mathbf{E}_M$$

- The MPCA is equivalent to performing a PCA on a large two-dimensional matrix, formed by unfolding the three-way array X in such a way as to put each of its vertical slices, corresponding to a specific bin, side by side, resulting in a two dimensional matrix [15×280].

4. ANALYSIS AND RESULTS / 1

3. **Process matrix projection.** $\mathbf{X_\Pi}$ is projected onto the MPCA model space and the residual matrix $\mathbf{E_\Pi}$ is estimated.

$$\mathbf{T}_\Pi = \mathbf{X}_\Pi \mathbf{P}_M$$
$$\hat{\mathbf{X}}_\Pi = \mathbf{T}_\Pi \mathbf{P}_M^T$$
$$\mathbf{X}_\Pi - \hat{\mathbf{X}}_\Pi = \mathbf{E}_\Pi$$

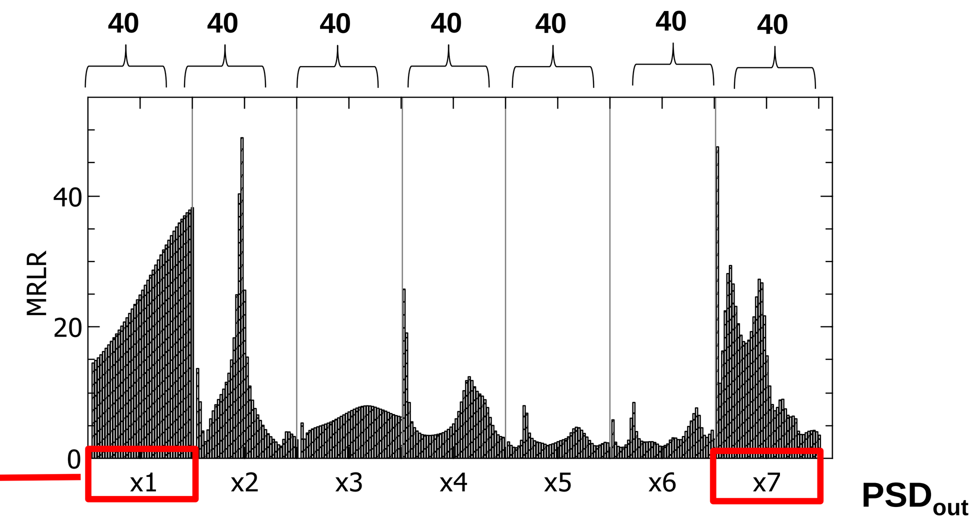


Large values of residuals, confirm that the correlation structure of $\mathbf{X_\Pi}$ is not well represented by the PCA model on $\mathbf{X_M}$

4. **Residuals analysis.** The two residuals matrices, $\mathbf{E_\Pi}$ and $\mathbf{E}_M$, are compared using the MRLR index to identify the auxiliary variables that are most responsible for the inconsistency in the correlation structures of $\mathbf{X_\Pi}$:

$$MRLR_v = \frac{\sum_{n=1}^{N}\left(\frac{e_{\Pi,n,v}}{CL_{95\%,e_{M,v}}}\right)}{N}$$

The MRLR index permits one to account for the contribution due to the PMM only, weighting the contribution related to the un-modeled variability of $\mathbf{E}_M$.

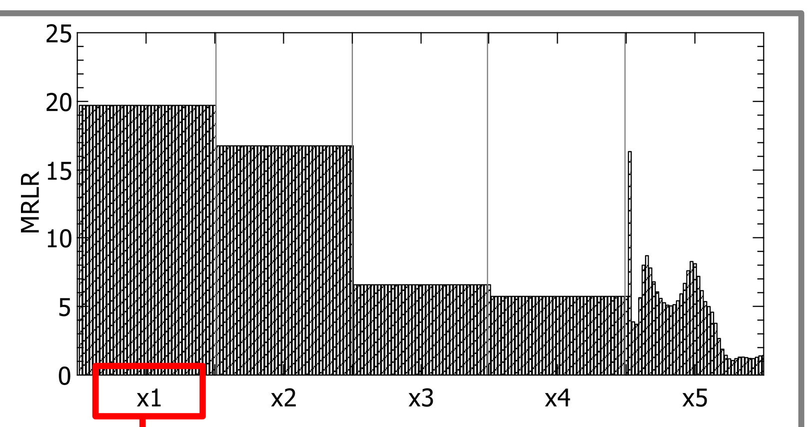


$P_{B,p}$, includes 3 parameters: f_{Mat}, $W_{m,kin}$, $W_{m,min}$

Which parameter is the cause of the mismatch?

The analysis was **repeated** considering 5 different auxiliary variables**:**

$$x_1 = f_{\mathrm{Mat}} \quad x_3 = W_{m,\mathrm{min}} \quad x_5 = f(PSD_{out})$$
$$x_2 = W_{m,\mathrm{kin}} \quad x_4 = q_{\mathrm{M}}$$

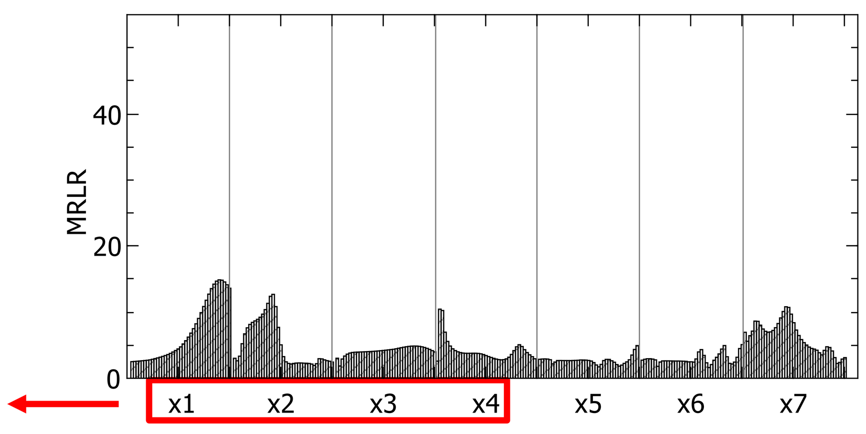


The parameter f_{Mat} is correctly identified as the reason of the PMM

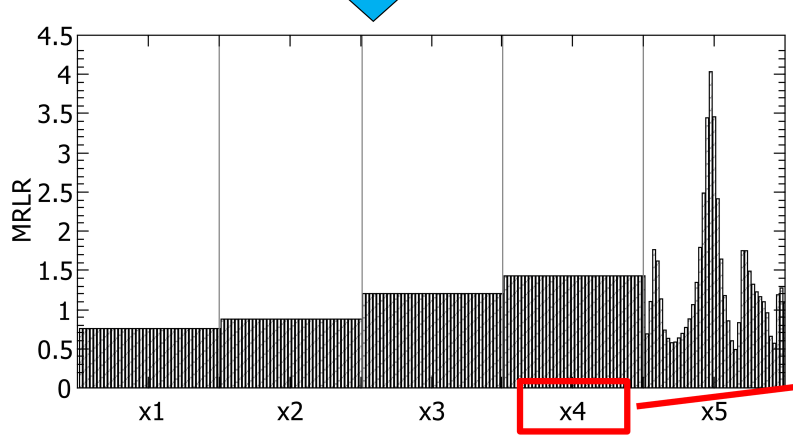
5. ANALYSIS AND RESULTS / 2

Modified values of **parameter *q***, related with the rotational velocity, were purposely introduced in the FP model in order to cause a PMM.

The procedure is again applied considering 7 auxiliary variables, in order to identify which section of the model should be improved.



Variables related both with $P_{B,p}$ and B_p present high values of MRLR



A further analysis is needed in order to identify which parameter is the most responsible of the PMM.

The parameter ***q***, has been correctly identified as the reason of the PMM.

6. CONCLUSIONS

A methodology has been proposed to diagnose the causes for PMM and thus to support model enhancement. The methodology exploits the information embedded in the historical available data (no further experiments are required) using the same FP model and DB model. The idea is to provide the modeler with a tool for detecting which sections of the FP model are not consistent with the data, thus targeting subsequent theoretical and experimental efforts.

REFERENCES

[1] Jackson J.E., 1991, John Wiley & Sons, New York, NJ.

[2] Vogel L., W. Peukert, 2005, Chem. Eng. Sci., 60, 5164-6176.

[3] gSOLIDS, 2013, (version 3.0), Process Systems Enterprise Ltd., London, UK.

[4] Nomikos P. and J.F. MacGregor, 1995b, Technometrics, 37, 41-59.

Contacts:
Prof. Massimiliano Barolo E-mail: max.barolo@unipd.it