Improved Understanding of Experimental Campaigns in Catalyst Development through machine learning

Edoardo Tamiazzo*, Alberto Biasin, Pierantonio Facco***

**CAPE-Lab – Computer Aided Process Engineering Laboratory, Padova, 35138*
Italy (Tel: +39.049.8275470; e-mail: pierantonio.facco@unipd.it)
***R&D Basic Research Department, CASALE SA, Lugano CH*

Abstract: The development of new catalysts is typically carried out by performing extended experimental campaigns of dynamic experiments through high-throughput miniature reactors in which the sequence of the experiment is often managed based on the experience of the scientists and the developers. In these systems, the sequential nature of experiments introduces complex effects that may propagate to successive experimental batches at different conditions which are difficult to model and interpret. Big amounts of data are typically collected from experimental campaigns, which provide the opportunity to develop data-driven models that extract valuable information on the system. In this study, we propose a new machine-learning methodology that allows the in-depth understanding of the experiment dynamics, associated with both the experiment batch itself and the catalyst history (namely, the sequence of multiple experiments performed in different conditions of temperature, composition, etc.). In particular, multiway multivariate latent variables techniques are used to capture the dynamic within the single experimental batch and the high auto-and cross-correlation between variables, two-dimensional dynamic modelling is used to deal with the dynamics of the catalyst history and orthogonalization is used to remove information redundancy. The methodology is validated in the case study of the development of catalyst for ammonia production. We show that the model captures the correlation between variables which describe the reaction kinetics and thermodynamics within each experimental batch, as well as the influence of catalyst history, especially in terms of feed composition. Furthermore, the model captures the contributions of both the dynamics of the single experimental batches and the catalyst history, ensuring very good predictive performance on the ammonia productivity.

Keywords: Catalyst development, Ammonia production, Machine learning, Experimental campaigns

---

1. INTRODUCTION

Discovering and optimizing new catalysts represents a significant challenge, requiring substantial investment in experimentation to collect data and analyze results. Catalyst development is typically conducted through expensive and time-consuming experimental campaigns which consist in the execution of series of multiple experimental batches in high-throughput setups (Sundermann and Gerlach, 2016). One of the main goals of the experimental campaigns is to obtain valuable information to characterize catalyst properties such as productivity and resilience to contaminants. However, understanding experimentation outcomes is not straightforward and presents several challenges (Ortega et al., 2021). The sequential nature of experimental campaigns exposes the catalyst to phenomena such as deactivation and regeneration, which highly impact the catalyst performance and cannot be easily modelled by traditional first-principles methodologies, such as kinetic models. In particular, it is difficult to characterize the dynamic features of the experimental campaigns, mainly due to: *i)* the *dynamics of a single experimental batch*, associated with the phenomena which are specific of each individual batch (e.g., the kinetics of the reaction), and *ii)* the *catalyst history* dynamics, resulting from the series of multiple stages subjected to phenomena such as poisoning, deactivation, regeneration etc.

However, these extended experimental campaigns generate large volumes of data, often redundant, that retain valuable information on the system under study. The interpretation of these data is not straightforward because of their nature: they are multivariate, collinear, and frequently affected by outliers and missing values.

In this work, we propose a machine-learning methodology to enhance the understanding of catalyst development in experimental campaigns. The proposed method includes well-known literature methods. In particular, multi-way multivariate latent-variable techniques (MPLS) are utilized to deal with data from experimental batches while handling large sets of highly correlated variables. (MacGregor et al., 1994). Furthermore, a two-dimensional (2D) dynamic model (Lu et al., 2005) is adopted to study both the dynamic effects by integrating the catalyst history in the form of the sequence of experimental batches. Orthogonalization (O) is then used to remove data redundancy of the catalyst history dynamics. (Næs et al., 2021). Accordingly, the proposed methodology is a 2D-MO-PLS (bidimensional orthogonalize multiway partial least square) and concurrently assesses the independent and joint impact of two dynamics while dealing with large volumes of correlated and redundant data. The methodology is applied to the case study of development of catalysts to produce ammonia in the laboratory scale, thereby offering a novel perspective on enhancing process understanding and decision-making.

2. MATERIAL AND METHODS

The proposed machine-learning methodology (shown in Figure 1) consists of 5 steps whose details are presented in the following Subsections: (1) data extraction and pretreatment; (2) extraction and synchronization of the catalyst history; (3) multi-way modelling of the data; (4) orthogonalization of the catalyst history; (5) two-dimensional modelling and MPLS model building.

2.1 Extraction and pretreatment of data from experimental campaigns

Data from 3-years of experimental campaigns to develop a new catalyst for the production of ammonia in a high-throughput setup are considered in this study. Each campaign involves long-term experimentation over several days, during which a candidate catalyst is continuously subjected to different experimental conditions in different experimental batches. During the experimental campaigns the catalyst is exposed to different reagent gas compositions: Condition A and Condition B. The gas composition of the reagent materials affects the performance of the catalyst. For this reason, the transition between the two different feeds, separated by phases that restore the original catalyst performance, allows for the assessment of the catalyst's resilience and productivity under varying conditions. The compositions of the two feeds are not disclosed for confidentiality.

Two types of data are collected during the experimental campaigns: process data and target variable. The target variable is the concentration of outlet gaseous ammonia and is measured by a micro gas chromatograph provided by Agilent Technologies Inc.. The GC analysis is taken with a defined sampling rate. Process data come from two identical tubular micro reactors in which catalyst converts nitrogen and hydrogen into ammonia. Each reactor is placed in an electrical furnace, which guarantees the proper heat conduction and a uniform distribution of the heat. Process variables are collected at a defined time interval and include: temperature, pressure, flowrates of the fed reagents, setpoints, etc.. Note that the duration of each experiment varies. Accordingly, a different number of time points is collected for each experiment.

Both raw process and target data are then pretreated through:

a. data cleaning: some of original experiments are affected by severe problems and are discarded. Target variable profiles are filtered to identify potential GC malfunctions. Furthermore, any recorded data from ancillary operations are also removed. A total of $N = 65$ batches are considered;

b. process data and target alignment: process and target data, which are sampled at different frequencies, are aligned to the same length (i.e., same number of time points) by linear interpolation between two consecutive data points;

c. synchronization: experimental batches are synchronized based on the percentage of completion of each batch, in such a way as the total number of considered time points is $T = 233$, the average number of points for the N available batches. The actual duration of each experiment is preserved and included as an additional process variable;

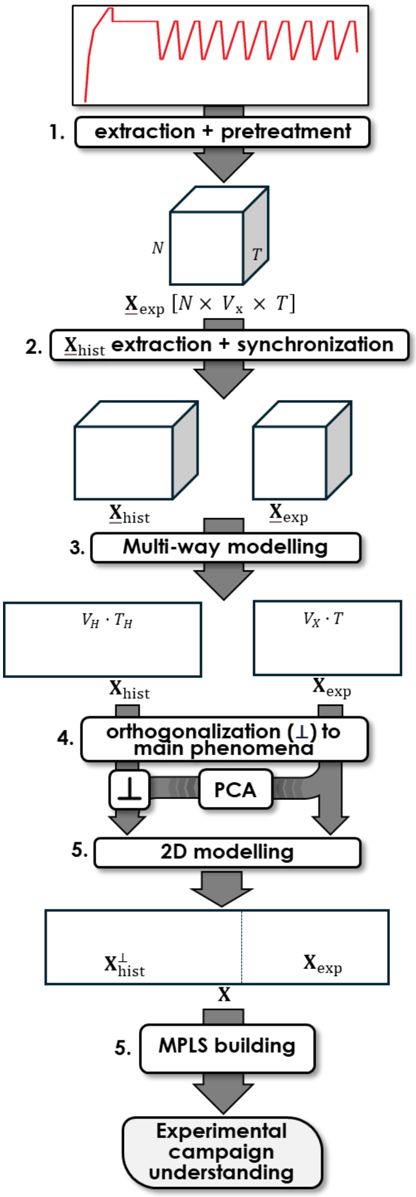


Fig 1. Proposed model schematic: 2D-MO-PLS.

d. data management: data are organized in three-dimensional arrays. $\underline{\mathbf{X}}_{\text{exp}}\ [N \times V_X \times T]$ contains information of $V_X = 33$ process variables for N experiments in T time points, while $\underline{\mathbf{Y}}\ [N \times 1 \times T]$ collects the time profiles of ammonia concentration.

2.2 *Catalyst history extraction and synchronization*

The catalyst history is extracted from the process data $\underline{\mathbf{X}}_{\text{exp}}$ considering the smallest integer number of batches that covers at least 18 hours of the experimental campaign before the current batch begins. We determined heuristically that considering the last 18 hours of catalyst history is the most appropriate choice. This value provides a tradeoff between the capability of capturing the dynamics of the catalyst history and the amount of data that is considered.

This information is collected in a matrix $\underline{\mathbf{X}}_{\text{hist}}$ $[N \times V_H \times T_H]$ of $V_H = 8$ variables, a subset of V_x that mostly capture the propagation of the drifts along the experiments synchronized in $T_H = 500$ time points. This subset includes the V_x variables that are the most relevant to catalyst history dynamics (e.g., the effect of temperature on catalyst activity). This preliminary selection reduces redundancy by limiting the number of variables. However, since more than one experimental batch is included to study catalyst history dynamics and the entire time trajectory of the variables is considered, further redundancy removal by orthogonalization is needed.

2.3 *Multi-way modelling*

The multi-way arrays $\underline{\mathbf{X}}_{\text{exp}}$, $\underline{\mathbf{X}}_{\text{hist}}$ and $\underline{\mathbf{Y}}$ are batch-wise unfolded (Nomikos, 1996) by horizontally concatenating data at different time points to generate the two-dimensional matrix that account for data dynamic nature and correlation structure among variables over different time points. The resulting matrixes $\mathbf{X}_{\text{exp}}[N \times V_X \cdot T]$, $\mathbf{X}_{\text{hist}}$ $[N \times V_H \cdot T_H]$ and $\mathbf{Y}$ $[N \times 1 \cdot T]$ are then auto-scaled, namely, centered to zero mean and scaled to unit variance. Note that, since the dynamic information for multiple past batches is retained along T_H, $\mathbf{X}_{\text{hist}}$ accounts for the dynamic correlation structure over different past batches.

2.4 Orthogonalization

Given the matrixes $\mathbf{X}_{\text{exp}}$ and $\mathbf{X}_{\text{hist}}$, a PCA model (Wold et al., 1987) is calibrated on $\mathbf{X}_{\text{exp}}$. PCA is used to project experiments onto a reduced space that retains only the main driving forces of the experiments, namely those which are associated to reaction kinetics, thermodynamics and mass balances:

$$\mathbf{X}_{\text{exp}} = \mathbf{S}\mathbf{C}^{\intercal} + \mathbf{R} \qquad (1)$$

The principal phenomena occurring in the experimental batches $\mathbf{X}_{\text{exp}}$ are summarized by the scores $\mathbf{S}$ $[N \times A]$ of the PCA model in a reduced space of A orthogonal principal components (PCs), whose director cosines are $\mathbf{C}$ $[A \times V_X]$, where subscript $\intercal$ denotes the transpose, and $\mathbf{R}$ $[N \times V_X]$ are the residuals.

The $\mathbf{S}$ scores are then used to discard redundant information in $\mathbf{X}_{\text{hist}}$ with respect to $\mathbf{X}_{\text{exp}}$ through orthogonalization:

$$\mathbf{X}^{\perp}_{\text{hist}} = \mathbf{X}_{\text{hist}} - \mathbf{X}_{\text{hist}}\,\mathbf{S}(\,\mathbf{S}^{\intercal}\mathbf{S})^{-\mathbf{1}}\mathbf{S}^{\intercal}\mathbf{X}^{\intercal}_{\text{hist}} \qquad (2)$$

Indeed, the catalyst history is constituted by a series of experiments whose variables display analogous correlation structures to those of the $\mathbf{X}_{\text{exp}}$ variables due to kinetics, thermodynamics and mass balance. The exclusion of the part of information in $\mathbf{X}_{\text{hist}}$ which is collinear to $\mathbf{S}$ improves the robustness of the model, avoiding at the same time data redundancy. In this case study PCA provides an orthogonalization which is very similar to the one provided by a PLS (usually the adopted approach for orthogonalization; Næs et al., 2021), but permits identifying explicitly and removes the variability in $\mathbf{X}_{\text{hist}}$ which is already present in $\mathbf{X}_{\text{exp}}$ independently from the prediction of $\mathbf{Y}$.

2.5 Two-dimensional (2D) modelling and MPLS model building

A predictive model to estimate the time profile of the ammonia concentration at the reactor outlet is built considering both the experimental batches data and the previous history of the catalyst. To this purpose, $\mathbf{X}_{\text{exp}}$, which contains dynamic information for each experiment, and $\mathbf{X}^{\perp}_{\text{hist}}$, which contains catalyst history not included in $\mathbf{X}_{\text{exp}}$, are used as predictors in MPLS (Nomikos, 1996). In particular, $\mathbf{X} = \left[\mathbf{X}^{\perp}_{\text{hist}} \middle| \mathbf{X}_{\text{exp}}\right]$ is the predictor matrix, while $\mathbf{Y}$ is the predicted one. The resulting model captures the auto- and cross-correlations between in experimental batches and catalyst history dynamics, predicting the response time profile through:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\intercal} + \mathbf{E} \qquad (3)$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^{\intercal} + \mathbf{F} \qquad (4)$$

$$\mathbf{T} = \mathbf{X}\mathbf{W}(\mathbf{P}^{\intercal}\mathbf{W})^{-\mathbf{1}} \qquad (5)$$

where $\mathbf{T}$ $[N \times L]$ is the score matrix, $\mathbf{P}$ $[(V_H \cdot T_H + V_X \cdot T) \times L]$ and $\mathbf{Q}$ $[T \times L]$ are the loading matrices, $\mathbf{E}$ $[N \times [(V_H \cdot T_H + V_X \cdot T)]$ and $\mathbf{F}$ $[T \times V_Y]$ are the residual matrices, and $\mathbf{W}$ $[[(V_H \cdot T_H + V_X \cdot T) \times L]$ is the weight matrix. L is the number of orthogonal latent variables (LVs) retained in the model. In 2D-MO-PLS weights reveals the correlations between the time profiles of process variables, catalyst history and the ammonia profiles.

3. RESULTS AND DISCUSSION

3.1 2D-MO-PLS model calibration

In this work both PCA and MPLS are built by nonlinear iterative partial least squares (NIPALS; Geladi and Kowalski, 1986).

The PCA model calibrated on $\mathbf{X}_{\text{exp}}$ retains 2 PCs that explain 53% of $\mathbf{X}_{\text{exp}}$ variability, thus obtaining the $\mathbf{S}$ scores used to orthogonalize $\mathbf{X}_{\text{hist}}$.

The MPLS model is built between $\mathbf{X}$ and $\mathbf{Y}$ and retains 2 LVs that explain 43.5% of $\mathbf{X}$ and 59.4% of $\mathbf{Y}$ variability.

The number of PCs in PCA and LVs in PLS are jointly selected by a Monte Carlo cross-validation on 5000-iteration which consists in a random partitioning of the available datasets in calibration and validation (90% and 10%, respectively) to maximize model prediction performance (Wold, 1978) in terms of determination coefficient in validation Q^2.

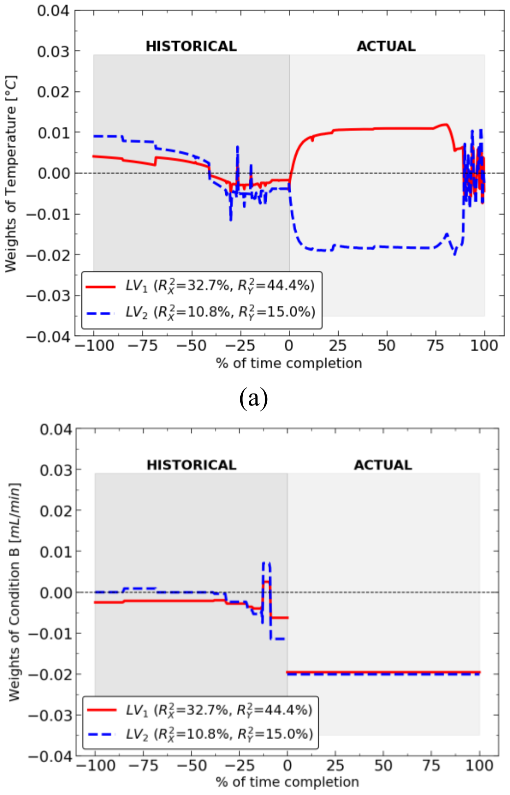
3.2. Study of the correlation of temperature and poisoning in the experimental batches and catalyst history

The weight plot of the 2D-MO-PLS model reveals the correlations between the ammonia time profile which, is positively correlated with higher values of the weights along both LV_1 and LV_2 for all time points, and experimental batch dynamics and catalyst history. This indicates that higher weight values are associated with an increase in ammonia concentration.

Temperature and Condition B are of particular relevance in this case study, as they are recognized as factors that can substantially impact experimental outcomes. Temperature controls the thermodynamic and kinetic phenomena occurring during the reaction, while Condition B gas composition is monitored to test catalyst resilience.

Temperature (Figure 2a) is positively correlated with ammonia along LV_1. The strength of this correlation is more pronounced in the central part of actual experiment, where higher temperatures enhance the kinetics. Nevertheless, this correlation is weakened at the end of the experiment, as the system approaches the thermodynamic limit. The influence of past temperature readings is low; however, temperature from the previous experiment appears to promote catalyst restoration, while that from the most recent experiment seems to contribute to catalyst deactivation. Condition B in the current experiment (Figure 2b) is negatively correlated with both temperature and ammonia concentration along LV_1 as its presence reduces ammonia productivity. The contribution of historical weights remains predominantly negative. The strength of this negative correlation increases in the most recent part of the catalyst history, thereby highlighting the considerable influence of Condition B in previous experiment on the outcome of the current experiment.

A joint reading of LV_2 weights of temperature and Condition B allows understanding that this LV captures the combined effect gas composition with temperature. Indeed, for the experiment's dynamics both weights of temperature and Condition B are always negative, indicating a negative correlation with ammonia concentration. An analogous pattern is also observed in the latter phase of the catalyst history. This correlation indicates that experiments performed with Condition B gas composition in previous experimental batches, without the restoration of original catalyst activities, has a significant negative impact on the performance of the current experiment. The model ability to capture this phenomenon highlights its effectiveness in the experimental campaign understanding.



(a)

(b)

Fig 2. Weights of 2D-MO-PLS for: (a) reactor temperature; and (b) gas composition

3.3. Experimental campaign improved understanding through 2D-MO-PLS

The 2D-MO-PLS scores encompass the total variable correlation, including both variation in time and variability between batches. In conjunction with weights, they enable the understanding of the relationships between the behavior of each experimental batch (i.e., the conditions in which it is carried out), the history of the catalyst, and the ammonia time profile.

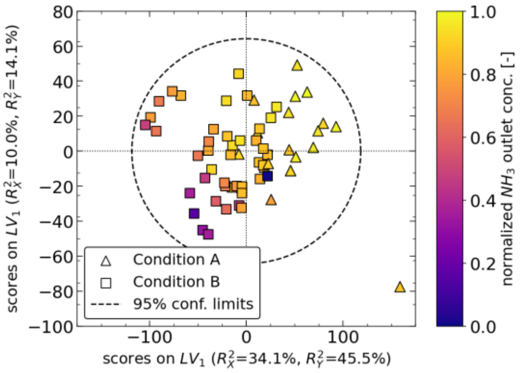


Fig 3. Experimental campaign improved understanding through 2D-MO-PLS score space.

The score space maps similar experiments as close points identifying batches that exhibit analogous structures of dynamic correlation in the process variables of the experimental batch and the history of its catalyst.

In Figure 3, experiments performed with Condition B (squares) and with Condition A (triangles) are projected onto the 2D-MO-PLS score space. Experiments are colored depending on the value of the normalized ammonia productivity. The score plot shows that the productivity variability is described along the diagonal of the first and third quadrant. Indeed, high values of LV_1 and LV_2 correlate with increased ammonia production.

These experiments are separated along LV_1 based on the distinction between Condition A and B, with experiments characterized by Condition B located in the left space of the score space, thus exhibiting lower productivity. Some experiments performed with Condition A are located on the left part of the score space and overlap with those performed with Condition B. This suggests that their performances are comparable to those observed under different reagent gas compositions, indicating that even if the current experiment is performed with Condition A composition, conditions exist which led to reduced productivity.

In the score plot, both batch dynamics and catalyst history are visualized at the same time, making the determination of whether the position of each experiment is primarily driven by the impact of the current experiment or influenced by its history a challenge. To address this problem, the components of scores $\mathbf{T}$ are separated in the part concerning the dynamics of the experiment $\mathbf{T}_{\text{exp}}$ and part concerning the catalyst history $\mathbf{T}_{\text{hist}}$. This are calculated inverting equation (3) using $\mathbf{P}_{\text{hist}}^{\top}$ and $\mathbf{P}_{\text{exp}}^{\top}$ derived from loadings $\mathbf{P}^{\top} = \left[\mathbf{P}_{\text{hist}}^{\top}\middle|\mathbf{P}_{\text{exp}}^{\top}\right]$.

In Figure 4, the aliquots of the two different types of dynamics are shown separately to provide a clear overview of the behavior of the experimental batch depending on the catalyst history. The y-axis represents the contribution of the score of the current experimental batch $\mathbf{t}_{\text{exp}}$ along the first latent variable (namely, the first column of $\mathbf{T}_{\text{exp}}$), which primarily captures the dynamics within a single experiment. The x-axis is associated with the catalyst history along the second latent variable $\mathbf{t}_{\text{hist}}$ (namely, the second column of $\mathbf{T}_{\text{hist}}$), which mainly describes the effect of reagent gas composition. Each experimental batch is represented as a point in this space, with the color of the point indicating the final ammonia productivity (yellow indicating high productivity, while purple low productivity).

The vertical direction differentiates experiments based on the different reagent gas composition, with high scores corresponding to experiment performed with Condition A. The horizontal direction separates experiments based on the impact of their historical contribution; low x-values are associated with experiments whose productivity is reduced due to the presence of Condition B in previous experiments. In this region we also find experiments performed under Condition A regime with reduced productivity, influenced by the history of the catalyst which was previously exposed to experiment performed under Condition B regime. This explains the presence of the overlapping region observed in Figure 3. Finally, a few experiments with Condition B composition appear on the right side of the score space. These experiments have high productivity (indicated by lighter colors) compared to other samples with similar y-values. This difference is attributed to their history which included effective phases that restore the original catalyst performances regeneration, mitigating accordingly the effects of experiment performed with Condition B in catalyst history.

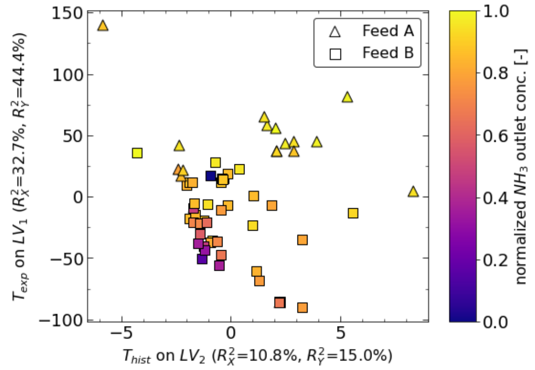


Fig 4. Experimental campaigns understanding by separation of the dynamic effects within an experimental batch and the history of its catalyst from the scores of 2D-MO-PLS.

3.4 Estimation of outlet ammonia profile

The proposed 2D-MO-PLS is compared to a MPLS built between $\mathbf{X}_{\text{exp}}$ and $\mathbf{Y}$, calibrated with 2 LVs. Table 1 shows the estimation accuracy of the proposed methodology in terms of: determination coefficients in calibration R_Y^2 and in validation Q^2 and average absolute estimation error $\bar{\varepsilon} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{V}\sum_{i=1}^{V}\left|y_{T,i} - \hat{y}_{T,i}\right|\right)$ where $V = 5$ are validation batches sampled from N and $y_{t,i}$ and $\hat{y}_{t,i}$ are model predicted and real ammonia measurements taken at time sample t. Furthermore, the ratio between the error $\bar{\varepsilon}$ and the variability of the calibration data $\sigma_{Y_{T,cal}}$ is calculated $\bar{\varepsilon}/\sigma_{Y_{T,cal}}$.

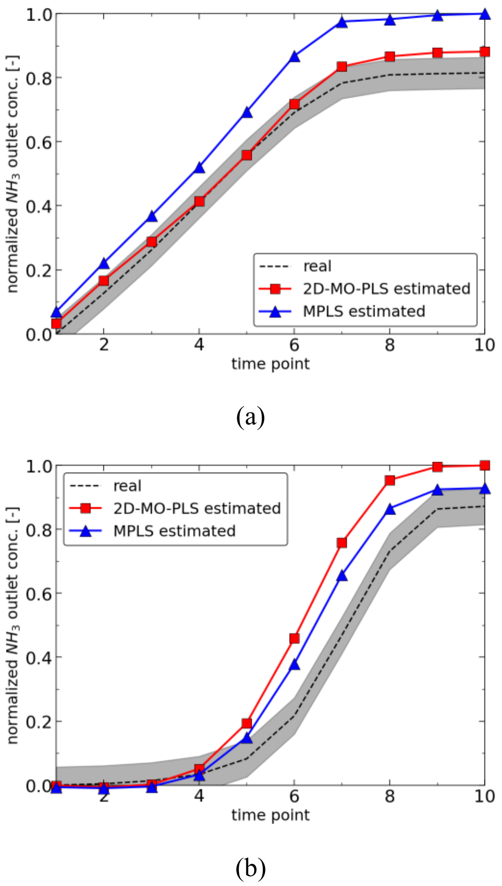
In general, the 2D-MO-PLS model does not demonstrate inferior predictive performance compared to a standard MPLS model. R_Y^2 is comparable between the two models, both in calibration and in validation. Nevertheless, the 2D-MO-PLS exhibits slightly superior predictive performance with lower error.

Table 1. Model performances of 2D-MO-PLS method compared to MPLS method.

Model	R_Y^2 [%]	Q^2 [%]	$\bar{\varepsilon}$ [vol%]	$\bar{\varepsilon}/\sigma_{Y_{T,cal}}$ [%]
2D-MO-PLS	59.4	81.2	1.57	33.3
MPLS	58.9	80.2	1.59	33.8

The estimated profile (Figure 5) for two experimental batch reveals that for some experiments the predictions of the 2D-MO-PLS are more accurate than those of the MPLS ($Q^2_{2D-MO-PLS}$ 25% higher than Q^2_{MPLS}) (Figure 5a). In this experiment the catalyst history reveals that previous experiments on the catalyst were performed with Condition B, resulting in a significant influence of the catalyst history. Conversely, Figure 5b demonstrates that, for an experiment previously exposed to Condition A, 2D-MO-PLS exhibits slightly low accuracy ($Q^2_{2D-MO-PLS}$ 12% lower than Q^2_{MPLS})

This demonstrates the ability of the 2D-MO-PLS in correcting predictions accounting for the influence of catalyst history.



(a)

(b)

Fig 5. Estimation of ammonia productivity time profile using 2D-MO-PLS (red line with squares) and MPLS (blue line with triangles) compared to the measurement (black dashed line). Experiment (a) is known to be affected by its catalyst history dynamics compared to experiment (b). The shaded area around the measured value identifies the measurement uncertainty (95%). For confidentiality reasons the y-axis scale is normalized between 0 and 1.

However, the incorporation of catalyst history decreases the prediction accuracy with respect to MPLS when the history does not play a role in the actual batch. This is also due to the fact that 2D-MO-PLS introduces a significant increase in complexity (namely, ~ 1,000 extra predictors). Nevertheless, the predictive performance remains consistent, indicating that the model effectively captures valuable information from both experimental batch dynamics and catalyst history.

4. CONCLUDING REMARKS

In this paper, we proposed a 2D-MO-PLS model to improve the understanding of not-designed experimental campaigns, made of sequence of experimental batches, in the high-throughput development of new catalysts. The method integrates 2D and multiway latent variables methodologies to capture two dynamic effects that influence the system: the dynamics of the experimental batch and the dynamics of the history of the catalyst in previous experiments. Moreover, the application of orthogonalization assists in the separation of these dynamic effects, while addressing the issue of information redundancy. Results demonstrate that the model captures very well both the dynamic effects of individual experimental batches and the catalyst history (namely, the influence from previous experiments). Although the model predictive performance does not significantly improve those of a standard MPLS, it successfully accounts for the propagation of historical phenomena, such as different inlet gas composition, providing deeper insights into process dynamics, especially in cases where the sequence of experiments plays a critical role.

Future work will focus on enhancing the model performance through variable selections techniques, reducing model complexity by improving the orthogonalization methodology, and transfer of the methodology to different catalysts.

REFERENCES

Farrusseng, D., 2008. High-throughput heterogeneous catalysis. Surface Science Reports 63, 487–513.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta 185,

MacGregor, J.F., Jaeckle, C., Kiparissides, C., Koutoudi, M., 1994. Process monitoring and diagnosis by multiblock PLS methods. AIChE Journal 40, 826–

Næs, T., Romano, R., Tomic, O., Måge, I., Smilde, A., Liland, K.H., 2021. Sequential and orthogonalized PLS (SO-PLS) regression for path analysis: Order of blocks and relations between effects. Journal of Chemometrics 35, e3243.

Nomikos, P., 1996. Detection and diagnosis of abnormal batch operations based on multi-way principal component analysis World Batch Forum, Toronto, May 1996. ISA Transactions 35, 259–266.

Ortega, C., Otyuskaya, D., Ras, E.-J., Virla, L.D., Patience, G.S., Dathe, H., 2021. Experimental methods in chemical engineering: High throughput catalyst testing — HTCT. The Canadian Journal of Chemical Engineering

Sundermann, A., Gerlach, O., 2016. High-Throughput Screening as a Supplemental Tool for the Development of Advanced Emission Control Catalysts: Methodological Approaches and Data Processing. Catalysts 6, 23.

Wold, S., 1978. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. Technometrics 20, 397–405.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemometrics and Intelligent Laboratory Systems, Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists 2, 37–52.