12th IFAC International Symposium on
Advanced Control of Chemical Processes
July 14-17, 2024. Toronto, Canada

Defining a three-zone multivariate specification region for incoming raw materials

Adéline Paris*,† Carl Duchesne Éric Poulin*****

* Department of Chemical Engineering, LOOP, Centre E4m, Université Laval, Quebec City, Canada

** Department of Chemical Engineering, LOOP, Université Laval, Quebec City, Canada

*** Department of Electrical and Computer Engineering, LOOP, Centre E4m, Université Laval, Quebec City, Canada

† *Corresponding author:* adeline.paris.1@ulaval.ca

Abstract: Multivariate specifications for the properties of incoming raw materials are used to determine if a lot could allow reaching the desired final product quality prior to its purchase. Previous works have shown that introducing optimization of process variables in the decision framework leads to accepting a wider range of materials. However, this requires to run the optimization problem each time a new lot is not deemed acceptable if the process is operated at nominal conditions. The objective of this paper is to simplify the current method by defining a more general specification region for raw materials considering three regions: accept, accept with process optimization, and reject. The concept is demonstrated using a grinding-flotation simulator where the objective is to assess a minimal profit when processing a lot of ore. For the case study, the results obtained from a test dataset show that the accuracy of the proposed method is 92%. It gives similar results compared to the decision framework considering the optimization problem.

Keywords: Multivariate specification, Process optimization, SMB-PLS, Quality control, raw material variability

1. INTRODUCTION

The use of decision-support tools to determine the acceptability of lots of raw materials prior to their purchase is an interesting solution to reduce the impact of raw material variability on the final products. Since raw material properties are often correlated [1, 2], defining multivariate specification regions are recommended. In its simplest version, it consists of defining a limit in the score space of a projection to latent structure (PLS) model by mapping final product quality in the latent variable space of the model, where the suitability of each lot of raw materials is assessed [1, 2]. This region can also be obtained using model inversion of each point of the final product quality limit [3, 4]. To increase the range of acceptability, previous works have shown that including adjustments of operating conditions is a natural solution [5, 6].

Paris et al. [5] have proposed a decision framework, called multivariate specifications with optimization (MSVO), to widen the specification regions. This tool first consists of determining if a given lot of raw materials can be transformed at nominal process operating conditions using multivariate specifications defined in the latent space of a Sequential-Multiblock PLS model. If the lot is not likely to yield the desired final product properties based on the specification region, an optimization problem is solved to determine if adjusting operating conditions could lead to the desired product quality while considering the cost of these changes.

The objective of this paper is to simplify the MSVO approach by defining a more general multivariate specification for incoming raw materials. In other words, instead of solving the optimization problem each time the second step of the approach is initiated (i.e., when a lot of raw materials falls outside the specification region), new limits in the latent space of the SMB-PLS model are defined to generate three zones as shown in Figure 1. Depending on where the new lot projects in the super score space, it could either be rejected, accepted for processing at nominal process operation conditions (POC) or accepted upon adjusting POC. The proposed concept is defined as the three-zone raw material multivariate specifications (TZ-MVS). The approach is demonstrated using a dynamic grinding-flotation simulator [7] calibrated on industrial data, which provides ground truth for quantifying its performance.

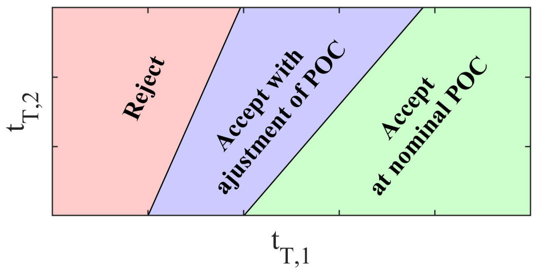

Figure 1. Conceptualization of the proposed specifications

The paper is organized as follows. Section 2 gives an overview of the methodology required to generate the TZ-MVS, which includes the development of an SMB-PLS model, the definition of the optimization problem as well as the classification procedure. Section 3 presents the case study, while section 4 focuses on the results and the discussion. The main conclusions are drawn in section 5.

2. METHODOLOGY

The new TZ-MVS method is divided in six steps as shown in Figure 2. The first four correspond to the MVSO approach described in detail in Paris et al. [5]. In brief, it consists of developing two SMB-PLS models relating final product quality attributes $\mathbf{Y}$ to raw material properties $\mathbf{Z}$, process manipulated variables $\mathbf{X}_{\mathrm{MV}}$, and some process measurements $\mathbf{X}_{\mathrm{m}}$. One model is used for developing the multivariate specification region, and another for solving the optimization problem. Secondly, the multivariate specification regions are defined in the latent space of the first model, which includes only the latent variables associated with the raw material (RM) properties. After propagating the quality index of the final product (i.e. good/bad) in the super-score space of the model, a boundary is adjusted to discriminate both classes. This boundary, combined with limits on Square Prediction Errors (SPE) and Hotelling's T² (HT2), form the multivariate specifications region. Adding SPE and HT2 limits allows to accept slight deviation in the correlation structure while ensuring a correct use of the model. If one of the three limits is violated, the optimization problem is solved. The existence of a solution indicates that it is possible to accept the new lot upon adjusting process conditions, otherwise the lot is rejected. Thus, in the MVSO framework, the output of the optimization step is a binary variable that indicates the decision to purchase or not a new RM lot that could not be processed at nominal conditions.

The novelty of the TZ-MVS (i.e. steps 5 and 6 in Figure 2) consists of projecting the calibration class labels obtained from the optimization step in the latent space of the first model to define a third region. For this new classification problem, only the lots considered in the optimization problem (second step of MSVO) are used. This results in establishing a limit based on a binary classification. The combination of both classification problem results in a three-zone decision tool.

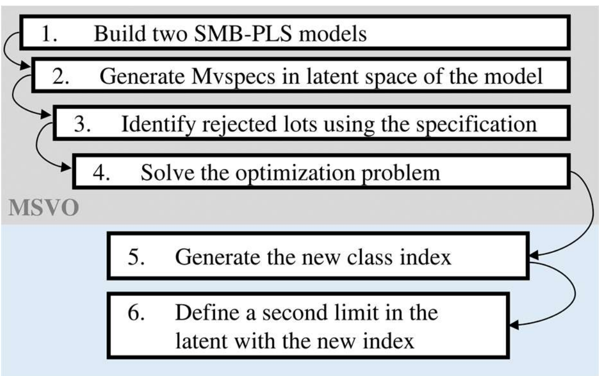

Figure 2. TZ-MVS methodology flowsheet

The next sections give an overview of the SMB-PLS model development and the formulation of the optimization problem. Brief details are provided to generate the classification procedure.

2.1 SMB-PLS modeling

To set up the TZ-MVS solution, two models are required as mentioned previously. Opting for SMB-PLS models is motivated by its sequential multi-block structure including block orthogonalization. It explains the variation in the final quality attributes $\mathbf{Y}$ using B predictor blocks $\mathbf{X}_b$ which are ordered to follow a sequential pathway. The raw material properties $\mathbf{Z} = \mathbf{X}_1$ are considered before the manipulated variables $\mathbf{X}_{\mathrm{MV}} = \mathbf{X}_2$, and the process measured variables $\mathbf{X}_{\mathrm{m}} = \mathbf{X}_3$ when making the prediction of $\mathbf{Y}$. The orthogonalization performed by the algorithm extracts the correlated variations between blocks $\hat{\mathbf{X}}_b^{\mathrm{corr}}$ and the orthogonal information in subsequent blocks $\mathbf{X}_b^{\mathrm{orth}}$. This isolate variations generated by control actions (i.e. feedback/feedfoward). The SMB-PLS regression is defined mathematically by the following equations:

$$\mathbf{X}_b = \mathbf{T}_{\mathrm{T}}\mathbf{P}_b' + \mathbf{E}_b \tag{1}$$

$$\mathbf{Y} = \mathbf{T}_{\mathrm{T}}\mathbf{C}' + \mathbf{F} \tag{2}$$

where $\mathbf{P}_b$ and $\mathbf{C}$ are the loading matrices associated respectively with $\mathbf{X}_b$ and $\mathbf{Y}$, while $\mathbf{E}_b$ and $\mathbf{F}$ are the residual matrices. For each new observation, the super-scores $\mathbf{T}_{\mathrm{T}}$ need to be calculated to make predictions:

$$\mathbf{T}_{\mathrm{T}}(:,\propto_{\mathrm{c},b}) = \left(\mathbf{X}_b - \hat{\mathbf{X}}_b^{\mathrm{corr}}\right)\mathbf{R}_{\mathrm{b}} = \mathbf{X}_b^{\mathrm{orth}}\mathbf{R}_b \tag{3}$$

using the weighted matrix $\mathbf{R}_b$ defined in Paris et al. [5]. The $\propto_{\mathrm{c},b}$ parameter provides the range of rows to consider each time equation 3 is applied, since it needs to be repeated b times to consider the deflation of the predicted correlation information:

$$\hat{\mathbf{X}}_b^{\mathrm{corr}} = \mathbf{X}_{\mathrm{b}} - \sum_{j=1}^{b-1} \mathbf{T}_{\mathrm{T}}(:,\propto_{\mathrm{c},j})\mathbf{P}_b'(:,\propto_{\mathrm{c},j}) \tag{4}$$

Compared to PLS, SMB-PLS allows choosing a different number of components A (i.e. latent variables) for each of the b blocks. The number of components is determined either by a cross-validation or using an external dataset, and depends on how the model is used. For classification (i.e. steps 2 and 6 of TZ-MSV), the number of components is determined based on classification metrics, such as accuracy. If the model is used for making predictions, like in step 4, the number of components is chosen based on its predictive ability quantified by the Q² statistic or the root mean squared errors of prediction RMSEP.

The TZ-MVS approach requires the use of limits on two statistical metrics associated with the SMB-PLS model. The squared prediction error for block b defined as:

$$\mathrm{SPE}_{\mathrm{x}_{\mathrm{b}},i} = \mathbf{e}_{\mathrm{x}_{\mathrm{b}},i}\mathbf{e}_{\mathrm{x}_{\mathrm{b}},i}' = \left(\mathbf{x}_{\mathrm{b}} - \hat{\mathbf{x}}_{\mathrm{b},i}\right)\left(\mathbf{x}_{\mathrm{b}} - \hat{\mathbf{x}}_{\mathrm{b},i}\right)' \tag{5}$$

allows to assess the consistency of the correlation structure. An upper control limit can be set using a χ^2 distribution as defined by Nomikos and MacGregor [8]. An upper control limit on the Hotelling T² (HT2) as proposed by Wierda [9] is used for detecting extrapolation and outliers. The HT2 value for an observation i is calculated as follows:

$$\mathrm{HT2}_i = \sum_{a=1}^{A} \left(\frac{\mathbf{T}_{\mathrm{T}}(i,a)}{s_a} \right)^2 \tag{6}$$

where s_a is the standard deviation of the super-score obtained in the model calibration phase.

2.2 Optimization in the SMB-PLS latent space

To set the second limit in the super-score space (i.e. step 5 of TZ-MVS approach), new class labels are required. They are obtained by determining if a solution to an optimization problem exist for each lot rejected by the multivariate specification region, considering the process operates at nominal conditions (step 3 of TZ-MSV).

The objective function of the optimization problem is defined based on plant objective such as maximizing profit under constraints. The cost of modifying the operating conditions should be considered. The optimization problem used to determine the operating conditions $\mathbf{x}_{\mathrm{MV}}^{\mathrm{new}}$ for a new lot of RM are subject to constraints divided in five parts:

$$\max_{\mathbf{x}_{\mathrm{MV}}^{\mathrm{new}}} f(\,\hat{\mathbf{y}}^{\mathrm{new}}, \mathbf{z}^{\mathrm{new}}, \mathbf{x}_{\mathrm{MV}}^{\mathrm{new}},\, \hat{\mathbf{x}}_{\mathrm{m}}^{\mathrm{new}}\,)$$

subject to

$$\begin{aligned}
&\mathrm{I.}\begin{cases} \mathbf{t}_{\mathrm{t}}^{\mathrm{new}} = \mathrm{eq.}\,3 = f(\mathbf{z}_{\mathrm{n}}^{\mathrm{new}}, \mathbf{x}_{\mathrm{MV,n}}^{\mathrm{new}}, \mathbf{x}_{\mathrm{m,n}}^{\mathrm{new}})\ \forall\, b = 1,2,3 \\ \hat{\mathbf{y}}_{\mathrm{n}}^{\mathrm{new}} = \mathbf{t}_{\mathrm{t}}^{\mathrm{new}}\mathbf{C}' \\ \hat{\mathbf{x}}_{\mathrm{MV,n}}^{\mathrm{new}} = \mathbf{t}_{\mathrm{t}}^{\mathrm{new}}\mathbf{P}_{\mathrm{MV}}' \\ \hat{\mathbf{x}}_{\mathrm{m,n}}^{\mathrm{new}} = \mathbf{t}_{\mathrm{t}}^{\mathrm{new}}\mathbf{P}_{\mathrm{m}}' \end{cases} \\
&\mathrm{II.}\begin{cases} \mathbf{x}_{\mathrm{MV,n}}^{\mathrm{new}} = (\mathbf{x}_{\mathrm{MV}}^{\mathrm{new}} - \mathbf{x}_{\mathrm{MV,mean}})/\mathbf{x}_{\mathrm{MV,std}}/\sqrt{J} \\ \hat{\mathbf{x}}_{\mathrm{m}}^{\mathrm{new}} = \hat{\mathbf{x}}_{\mathrm{m,n}}^{\mathrm{new}} \times \mathbf{x}_{\mathrm{m,std}} \times \sqrt{N} + \mathbf{x}_{\mathrm{m,mean}} \\ \hat{\mathbf{y}}^{\mathrm{new}} = \hat{\mathbf{y}}_{\mathrm{n}}^{\mathrm{new}} \times \mathbf{y}_{\mathrm{std}} + \mathbf{y}_{\mathrm{mean}} \end{cases} \\
&\mathrm{III.}\begin{cases} \mathrm{SPE}_{\mathbf{x}_{\mathrm{MV}}} = (\hat{\mathbf{x}}_{\mathrm{MV}}^{\mathrm{new}} - \mathbf{x}_{\mathrm{MV}}^{\mathrm{new}})^2 \le \mathrm{SPE}_{\mathrm{UCL}} \\ \mathrm{HT2} = \sum_{a=1}^{A_{\max}} \left(\frac{\mathbf{t}_{\mathrm{t}}^{\mathrm{new}}(a)}{s_a} \right)^2 \le \mathrm{HT2}_{\mathrm{UCL}} \end{cases} \\
&\mathrm{IV.}\begin{cases} \boldsymbol{\rho}_{\min} \le \hat{\mathbf{y}}^{\mathrm{new}} \le \boldsymbol{\rho}_{\max} \\ f(\hat{\mathbf{y}}^{\mathrm{new}}, \mathbf{x}_{\mathrm{MV}}^{\mathrm{new}}, \hat{\mathbf{x}}_{\mathrm{m}}^{\mathrm{new}}, \mathbf{z}^{\mathrm{new}}) \le \aleph \end{cases} \\
&\mathrm{V.}\{\ \Phi_{\min} \le \mathbf{x}_{\mathrm{MV}}^{\mathrm{new}} \le \Phi_{\max}
\end{aligned} \tag{7}$$

The equality constraints (I, II) define the model, and includes the normalization and block scaling. The inequality constraints (III) ensure the solution is consistent with the model structure. Part IV defines the zone of acceptability of the final product. It could either be limits on final properties or a function that take into account the economics. The last part (V) is required to set the bounds on the optimized manipulated variables. More details on the optimization problem are available in [5].

2.3 Classification procedure

Steps 2 and 6 of the TZ-MVS approach require to establish limits to discriminate between observations that meet the final quality criteria and those that do not, based on predetermined class label. These limits are defined in the super-score space of an SMB-PLS model, which only retains the components associated with the **Z** block. The choice of the optimal limit is made in two steps: first, the shape of limit is determined followed by the selection of the optimal number of components.

To achieve this, as many models as the number of Z-variables are calibrated. For each of them, a supervised learning approach is used to solve the binary classification problem. To facilitate the use of the resulting limit (i.e. a direct inequality), discriminant analysis is considered. Applying this technique requires to determine the best shape of discriminant between linear, quadratic and their alternatives (i.e. pseudo or diagonal). Using a grid search approach and a 10-fold cross-validation, the best discriminant is determined for each model. If the classes are unbalanced, a misclassification cost should be considered to determine the most appropriate discriminant.

For each discriminant retained, the classification performance is calculated using the validation dataset to determine the ideal number of components. The choice is based on the plant's specific objectives, e.g. maximizing accuracy or minimizing the number of false positives. For the multivariate specification region considering nominal process conditions (step 2), the performance should be calculated using the whole model structure, including the discriminant as well as the limits on SPEz and HT2. For step 6, only the discriminant is considered since the class labels were determined using these limits.

3. CASE STUDY

The TZ-MVS concept is demonstrated using a grinding-flotation plant simulator calibrated on industrial data [7]. The process shown in Figure 3 consists first of reducing the particle size of crushed galena using two grinding stages: a high-pressure grind rolls (HPGR) and a ball mill (BM), respectively, in closed loop with a wet screen, and an hydrocyclone cluster. The grinding product coming from the cyclone overflow feeds the flotation circuit, which separates and concentrates the lead (i.e. the valuable mineral). The process is well instrumented and proportional-integral-derivative (PID) control loops are implemented to maintain the process around the nominal operation point.

The aim is to define a TZ-MVS that allows to determine whether or not an ore deposit with specific properties should be mined, given a minimum profit of \$5175/h after processing. Since a RM specification region needs to be defined, variations between ore lots are simulated by modifying three ore properties. A total of 300 lots are processed at nominal conditions (i.e. the setpoints in gray in Figure 3). The lots are split in three groups of 100 observations to form a calibration, a validation and a test dataset. Each of them maintains the same class ratio of 2:1. The grade of the ore (z_1) varies between 12.5% and 13.5% while the average grain size (z_2) ranges between 40 and 50 µm. A positive correlation between these two variables is set at 90%. The third property defined the ease with which rock breaks. It is the hardness factor (z_3), which varies between 0.85 and 1.15, and is negatively

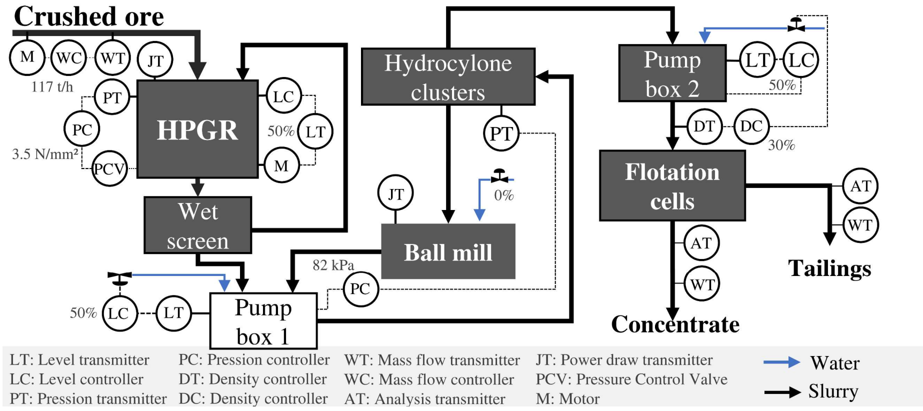

Figure 3. Grinding-flotation plant simulator

correlated at 80% with ore grade. These three properties form the **Z**-matrix.

Two manipulated variables are considered: the feed flow rate w_f as well as the force applied to the grinding rolls of the HPGR, since they influence the profit of the plant, and the final production. At nominal conditions, they are fixed at 117t/h and 3.5N/mm². Modifying the force over 3.5N/mm² should be performed only if required because mechanical drawbacks, such as increased stress and wear might occur [10]. Since excitation of these variables is required to build the predictive model, a 3² factorial design (i.e. [98, 117, 136] t/h and [2.6 3.8 5] N/mm²) is performed to supplement the calibration dataset. To model non-linear relations existing between RM properties and manipulated variables, the DOE is repeated with three groups of ore lots having different properties. For each lot processed, the specific force and the flow rate are stored in the $\mathbf{X}_{\mathrm{MV}}$-matrix. It also includes the interaction $z_3 w_f$ and the quadratic term w_f^2.

The block $\mathbf{X}_{\mathrm{m}}$ contains the measurements of the power drawn by each mill (q_{HPGR}, q_{BM}) while block **Y** contains the two final properties: the concentrate grade g_{c} and the concentrate flowrate w_{c}. Both are required to calculate the economic criterion J used in the objective function, and to define the class labels. It consists of the net value of flotation (i.e. the net smelter return where the operational costs are removed) with a penalty for ore losses in tailings. For the process illustrated in Figure 3, the economic criterion is:

$$\mathcal{J} = 4533 g_c w_c - 150.05 w_c - 46.65 w_f - 2165 w_f z_1 \\ -0.5946 q_{\mathrm{BM}} - 0.3 q_{\mathrm{HPGR}} - 4533 \quad (8)$$

where the numerical values of the coefficients were obtained from the economic parameters available in Thivierge et al. [7].

4. RESULTS AND DISCUSSION

This section focuses on presenting the result by illustrating each step of the TZ-MVS approach defined in Figure 2. Prior to defining the multivariate specification regions at nominal process conditions, the calibration dataset containing 100 observations is combined with the DOE data to generate the different models. A total of three SMB-PLS models are calibrated (A = 1,2,3), since the maximal number of components that can be retained is equal to the number of raw material properties considered in this case study. As the TZ-MVS is defined to ensure a minimal profit of \$5175/h, class labels are generated using this threshold. As mentioned in section 2.3, the optimal shape of discriminant for each model needs to be determined. A misclassification cost is considered to counterbalance the 2:1 class ratio. For the three models, the diagonal quadratic discriminant gives the best cross-validation performance. This type of discriminant considers a diagonal covariance matrix that may vary among classes. For each model, the resulting discriminant is coupled with the 95% limits on the SPEz and HT2 to generate the specification region. Table 1 provides the performance calculated with the validation dataset to determine the optimal number of components. The discriminant alone leads to similar performance, i.e. 12 false negatives. Adding the limits on SPEz and HT2 explains the performance difference. Even if a decrease in performance occurs, the limits on SPEz and HT2 are required to ensure a correct use of the model. It should be noted that when using the specification, if it suggests rejecting a lot solely on the basis of not meeting the limit on SPEz or HT2, a case-by-case decision could be taken. If a certain level of risk can be accepted, processing this lot can be used to enhance the model and the specification regions (i.e., a designed experiment). For the case study a conservative (i.e. no risk) approach is considered. Opting for one component leads to maximizing the accuracy with no false positive (i.e. predict an unprofitable lot as profitable). Thus, the nominal POC specification region relies on a diagonal quadratic discriminant developed in the latent space of a one-component SMB-PLS model.

Table 1. Classification performance assessed on a validation dataset – Specification at nominal conditions [Specification region/Discriminant alone]

Metrics \ Comp.	1	2	3
Accuracy [%]	86	82	83
# FP [obs.]	0/0	0/0	0/0
# FN [obs.]	14/12	18/12	17/12

After establishing the specifications at nominal conditions, the optimization problem needs to be set up to generate the new class labels (i.e. TZ-MSV step 5). Using the same calibration dataset that includes the DOE, a second model is calibrated, where components associated with block $\mathbf{X}_{MV}$ and $\mathbf{X}_m$ are considered. Cross-validation results show that the highest Q²Y value is obtained using $A_z = 3$ and $A_{X_{MV}} = 4$. Adding components associated with $\mathbf{X}_m$ do not improve the predictive ability of the model (i.e., $A_{X_m} = 0$). The 99% limits on SPE_{xm} and HT2 from the resulting SMB-PLS model are calculated to set the bounds in the third section in equation 7. Since the goal is to achieve at least \$5175/h, the fourth section includes an inequality based on the economic criterion (equation 8). The same equation is used as the objective function.

The new-class labels are obtained by solving the optimization problem for the 39 observations in the calibration dataset that were rejected based on the multivariate specifications at nominal conditions. The ones that did not respect the limits on SPE and/or HT2 are not considered since they are inconsistent with the model structure. The optimization problem suggests that 18 lots should lead to a profit of at least \$5175/h while 21 should be rejected. The same approach is used for the validation dataset. From the 45 observations that fail to meet the threshold based on the multivariate specifications at nominal conditions, a total of 22 lots could be accepted if POC are adjusted.

Using the 39 observations from the calibration dataset, and their new class labels, a discriminant is calibrated in the super-score space of each SMB-PLS model built with only *Az* components, as developed previously (i.e. during the multivariate specification at nominal condition step). Cross-validation results reveal that the optimal shape of discriminant depends on the model. Quadratic discriminant performs better with the model with one or two components, while linear discriminant gives better results with three components. The classification performance of these three discriminants is compared to determine the optimal one using the validation dataset and the new class labels. Table 2 shows that two discriminants give the same accuracy, but that their misclassification errors are different. The quadratic discriminant defined in the subspace of the one-component SMB-PLS model allows reducing the number of false positives compared with the linear discriminant. However, this is at the expense of adding a false negative. In this case, the decision should be based on plant objectives: accepting the risk of processing a lot that won't reach the threshold (i.e. accepting FP) or opting for a more conservative decision, and reject a lot that can potentially yield the desired profit (i.e. accepting more FN). For the case study, conservative decisions were considered. Thus, the quadratic discriminant defined in the latent space of the one-component SMB-PLS model is used to define the second limit.

Table 2. Validation classification performance – Specification at modified POC

Metrics \ Type: Comp.	Quadratic 1	Quadratic 2	Linear 3
Accuracy [%]	93.3	91.1	93.3
# FP [obs.]	2	2	3
# FN [obs.]	1	2	0

The resulting TZ-MVS for this case study is defined as:

$$\begin{cases} \left(\dfrac{t_T}{1.721}\right)^2 \leq 3.9472 \\ \displaystyle\sum_{i=1}^{3}\left(z_i - t_T p_{z,i}\right)^2 \leq 0.2308 \\ \text{And/or} \begin{bmatrix} t_T \leq 0.41 \\ t_T \leq 0.96 \rightarrow \text{requires adjusting POC} \end{bmatrix} \end{cases} \tag{9}$$

where p_i represents the ith element of the loading vector of block $\mathbf{Z}$ ($\mathbf{p}_z = [0.46 \quad -0.48 \quad -0.47]$). If the first three inequalities are respected, the lot is accepted directly. In case the third limit is violated, but the fourth is respected, the lot is accepted under the condition that POC are adjusted.

A test dataset is used to establish the final classification performance of the TZ-MVS represented by equation 9. This ensures that the defined limits work well with new data never seen in calibration or decision-making. Table 3 presents the performance of each step while Figure 4 shows the final TZ-MVS for the test dataset. The three background colors in Figure 4 correspond to the different decisions: i.e., accept (green), reject (red) or accept with POC modification (blue). The green squares and the red triangles represent the true class labels associated respectively with accept (class 0) and reject (class 1). For each lot that is rejected based on the specification limits based on nominal process conditions, a grid search approach is performed to identify all the lots that meet the threshold to generate the true label for class 2 "accepted with POC modifications". They are represented by the blue dots. The five black stars indicate the lots that were rejected due to violation of the SPE and/or HT2 limits. Of these, four are among the 12 false negatives specified in the Table 3. The eight others are represented by the gray asterisks. The red triangles in the upper part of the blue zone show the four false positives. The global accuracy for the proposed method is 92%.

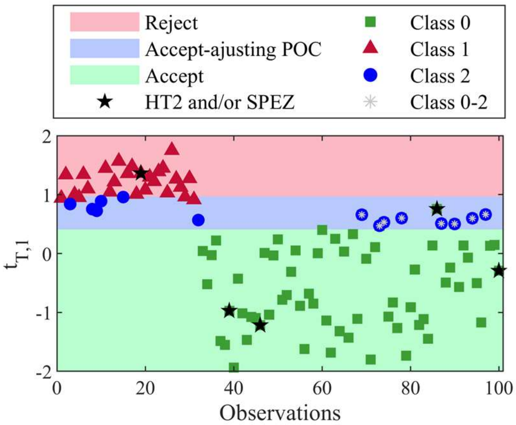

Figure 4. Representation of the TZ-MSV for the test dataset

For the case study, the TZ-MVS performs well introducing the third zone instead of running the optimization for all lots rejected by the specification limits leads to a small decrease in the performance (i.e. 4%) compared to the MVSO as shown in Table 3. This is caused by the four false positives. All in all, from the 39 lots that were rejected on the basis of the multivariate specification at nominal process conditions, a total of 14 are accepted upon adjusting operating conditions. This represents an increase of 18% compared with operating at fixed nominal conditions. The use of limits instead of solving each time the optimization problem facilitates the use of the decision tool when a new lot of RM is available from the suppliers. It should be noted that the size of the region associated with accept – modification of POC (i.e. blue area in Figure 4) depends on the flexibility of the process to accommodate a wide range of RM properties, and the impact of these modifications on the economic criterion. If the range of variations in operating conditions is very small or if the effect on the economic criterion is insignificant, this may result in a null blue zone may result.

Table 3. Classification performance for the test dataset

Metrics \ Step	Spec	Opt	TZ-MVS	MSVO
# Observations	100	39	100	100
Accuracy [%]	88	89.7	92	96
# FP [obs.]	0	4	4	0
# FN [obs.]	12	0	4	4

5. CONCLUSIONS

The objective of this paper was to demonstrate that the utilization of the decision scheme proposed by Paris et al. [5] which considers multivariate specification regions for incoming raw materials for a process operating at nominal, and solving an optimization problem to adjust process operation when needed can be simplified. The proposed solution consists of using a second limit in the super score space of an SMB-PLS model instead of solving an optimization problem to assess if the lot of RM can be accepted upon using modified process operating conditions. A grinding-flotation plant simulator is used to illustrate the development of this tool, which determines if processing a lot of ore is economically viable. The resulting three-zone multivariate specification region is defined using a one-component SMB-PLS model that relates the three ore properties, the two manipulated variables (i.e. force applied on the grinding rolls and the feed flowrate) and the energy consumption of both mills to the final quality attributes known as the concentrate grade and flowrate. The region consists of a diagonal quadratic discriminant adjusted in the latent space of the model, combined with 95% limits on SPEz and HT2, to determine the acceptability of a lot using nominal process conditions. In addition, a quadratic discriminant is defined in the same subspace to generate the zone where a lot can be accepted upon adjusting process operating conditions. For the proposed case study, the accuracy calculated on a test dataset is 92%. Compared with the original method, the number of false positives increase slightly, but the performance remains similar.

Future works will focus on testing the proposed approach using an industrial case study. Before developing this decision tool in a new context, it would be necessary to ensure that the level of excitation in the raw material properties and manipulated variables is sufficient to define specifications and allows optimization of operating conditions. The uncertainty propagation will also be investigated to see if the performance can be improved.

ACKNOWLEDGMENTS

The authors would like to thank Alex Thivierge for sharing the simulator. The financial support provides by the Natural Sciences and Engineering Research Council (NSERC) [RGPIN-2018-04800 and ESD3-546425-2020], the Fonds de Recherche du Québec – Nature et Technologies (FRQNT) [Scholarship 287725] and Rio Tinto [Scholarship] is also gratefully acknowledged.

REFERENCES

1. De Smet, J., *Development of multivariate specification limits using partial least squares regression*. 1993, McMaster University.
2. Duchesne, C. and J.F. MacGregor, *Establishing multivariate specification regions for incoming materials.* Journal of Quality Technology, 2004. **36**(1): p. 78-94.
3. García-Muñoz, S., S. Dolph, and H.W. Ward, *Handling uncertainty in the establishment of a design space for the manufacture of a pharmaceutical product.* Computers & Chemical Engineering, 2010. **34**(7): p. 1098-1107.
4. Paris, A., C. Duchesne, and É. Poulin, *Establishing Multivariate Specification Regions for Incoming Raw Materials Using Projection to Latent Structure Models: Comparison Between Direct Mapping and Model Inversion.* Frontiers in Analytical Science, 2021. **1**.
5. Paris, A., C. Duchesne, and É. Poulin, *Adjusting plant operating conditions to widen multivariate specification regions for incoming raw materials – An optimization framework* Chemometrics and Intelligent Laboratory Systems, 2023: p. 104991.
6. Borràs-Ferrís, J., C. Duchesne, and A. Ferrer, *Defining multivariate raw material specifications via SMB-PLS.* Chemometrics and Intelligent Laboratory Systems, 2023. **240**.
7. Thivierge, A., J. Bouchard, and A. Desbiens, *Comparing economic model predictive control to basic and advanced regulatory control on a simulated high-pressure grinding rolls, ball mill, and flotation circuit.* Journal of Process Control, 2023. **122**: p. 159-171.
8. Nomikos, P. and J.F. MacGregor, *Multivariate SPC charts for monitoring batch processes.* Technometrics, 1995. **37**(1): p. 41-59.
9. Wierda, S.J., *Multivariate statistical process control—recent results and directions for future research.* Statistica neerlandica, 1994. **48**(2): p. 147-168.
10. Zervas, G. *The Metcalf concentrator HRC™ 3000: Performance at variable specific force*. in *Proceedings of an International conference on autogenous grinding, semi-autogenous grinding and high pressure grinding roll technology*. 2019.