Eighth World Conference on Sampling and Blending

Paper Number: 59.00

Optimizing Multivariate Variographic Analysis with Information from Multivariate Process Data Modelling (PLS regression)

Q. Dehaine¹, L.O. Filippov² and H.J. Glass³

1. Postdoctoral research associate, University of Exeter, Camborne School of Mines, Penryn, Cornwall, TR10 9FE, United Kingdom. Email: Q.Dehaine@exeter.ac.uk

2. Professor, Université de Lorraine, GeoRessources laboratory, CNRS, UMR 7359, F54518, Vandœuvre-lès-Nancy, France. Email: lev.filippov@univ-lorraine.fr

3. Professor, University of Exeter, Camborne School of Mines, Penryn, Cornwall, TR10 9FE, United Kingdom. Email: H.J.Glass@exeter.ac.uk
ABSTRACT
Whereas a classical tool from the Theory of Sampling (TOS), variographic analysis, can address practical situations with multiple variables, its application has very often been limited to one variable at a time. Recent developments have shown the benefits of using multivariate approaches for variographic characterisation of a set of variables instead of considering individual variables sequentially. Among these approaches, the multivariogram has been revealed itself to be a powerful tool when the overall time-variability of a process must be summarized in terms of a large set of properties (variables) to assess its true global variability. However, even when choosing carefully the properties of interest for the process tested to avoid unnecessary variance increase, the resulting global variance with this approach is very high. In particular, some variables which contribute to a major proportion of the global (multivariate) variability could be less important for the process performance than others having a lower variability. To address this issue, a new approach is proposed, combining the multivariogram with process modelling and multivariate data analysis methods such as Partial Least Squares (PLS) regression from chemometrics. An example from the mineral processing industry is presented, for which the process performance could be linked to key process variables (sensor data) using the PLS regression. Once introduced in the multivariogram equation, PLS model parameters (loading-weights or regression coefficients) can be used to weigh the variables according to their relevance for the process. In addition, this also permits characterisation of process performance variability with time using only the process input variables and a weighted metric according to the PLS regression model. Ultimately, this method helps to find an optimized sampling procedure in terms of frequency, sampling mode and number of increments according to the actual overall process performance. This approach has potentially many applications in the mining, food and feed, pharmaceutical or any other industry for which it can be used to reduce risks and ensure a better use and management of resources.

INTRODUCTION
Predicting process performance variability over time is crucial for the mineral processing industry and many other industrial processes. Mineral processing operations are susceptible to process variations, operating parameters changes or ore variability, which could generate significant losses of performance for the whole process. In TOS, introduced by Pierre (Gy 1998; Gy 1992; Gy 1971), process streams can be seen as elongated objects, with a very long length compared to its two dimensional cross section. These objects are considered in TOS as one-dimensional (1D) lots according to their projection in the flow direction (Gy 2004b). The preferred method for sampling such 1D lots is composite sampling (Minkkinen & Esbensen 2009), which involves taking increments of the whole stream for a fraction of the time and then merging the increments into the final sample (Gy 2004a).

In order to characterise the autocorrelation between the units (i.e., increments) of these 1D processes, TOS introduced the semi-variogram (referred to as variogram in the present text). Variograms provide critical information on the process variability over time and the magnitude of the different variability components (Gy 2004c; Esbensen & Paasch-Mortensen 2010; Esbensen et al. 2012), see FIG 1. For a set of N discrete units, representing the lot collected from a 1D process throughout a time period, the relative heterogeneity \( h_i \) associated with a single unit of mass \( M_i \), which carries a concentration \( x_i \), is defined as follows:

\[
h_i = \frac{x_i - \bar{x}_L}{\bar{M}_i}, \quad i = 1, \ldots, N
\]  

(1)

where \( x_i \) is the lot mean and \( \bar{M}_i \) is the average increment mass. This relative heterogeneity is dimensionless, and the component \( x \) can describe any intensive chemical or physical property that characterises the material being transported (e.g., density, humidity or particle size). The variogram \( v_j \) is then defined as:

\[
v_j = \frac{1}{2(N-j)} \sum_{i=1}^{N-j} (h_i - h_{i+j})^2, \quad j = 1, \ldots, N/2
\]  

(2)

where \( j \) is a dimensionless lag-parameter that defines the distance between two time increments. An extensive description of the variographic technique and its practical applications can be found in reference papers (Gy 2004c; Pitard 1993; Minkkinen 2004; Esbensen & Paasch-Mortensen 2010).

However, process performance often depends on several properties, represented by a set of \( p \) variables, in which case one should consider the multivariate nature of heterogeneity. This consideration is well-known in geostatistics but has not been frequently addressed in TOS. An initial solution was proposed by Oliver and
Webster, who suggested that a principal component analysis (PCA) of the data could be carried out followed by a variographic study of the first few principal components (Oliver & Webster 1989). However, only a limited number of case studies have applied this approach within the TOS community so far (Minkkinen & Esbensen 2014).

Another approach for spatial data analysis, introduced by Bourgault & Marcotte (1991), is to assume that heterogeneity is a multivariate feature. For a set of \( N \) discrete units, representing the lot collected from a 1D process throughout a time period, the global heterogeneity contribution \( H_i \) can be expressed as a vector of \( p \) individual heterogeneity contributions (Dehaine et al. 2016):

\[
H_i = [h_{i1}, \ldots, h_{ik}, \ldots, h_{ip}]^t, \quad i = 1, \ldots, N
\]

(3)

where \( h_k \) is the heterogeneity vector that is associated with the \( k \)th variable. Bourgault and Marcotte (1991) introduced the multivariate variogram defined in a similar way to that of a traditional variogram but in a \( p \)-dimensional space. This multivariate variogram (referred to as multivariogram in the present text) has been extensively used in spatial data analysis (Bourgault & Marcotte 1991; Bourgault et al. 1992; Kerry & Oliver 2003) but only recently applied to TOS (Dehaine et al. 2016; Dehaine & Filippov 2015). The multivariogram \( V_j \) is calculated by analogy to the univariate variogram\(^1\):

\[
V_j = \frac{1}{2(N-j)} \sum_{i}^{N-j} (H_i - H_{i+j})M(H_i - H_{i+j})^t, \quad j = 1, \ldots, N/2
\]

(4)

where the subscript \( t \) is the transpose operator and \( M \) is the metric (positive definite \( p \times p \) matrix) which defines the "distances" between the units i.e., the relationship between the variables. The most commonly used metrics are the Euclidean distance (ED) and the Mahalanobis distance (MD). When using ED, in which \( M \) is defined as the identity matrix, the multivariogram is simply the sum of the univariate variograms, so one should avoid this metric and rather use MD. The latter, for which \( M \) is defined as the inverse of the variance-covariance matrix of \( H \), is more appropriate as it considers the correlation in the data (De Maesschalck et al. 2000; Bourgault & Marcotte 1991).

COMBINING THE MULTIVARI OGRAM WITH THE MULTIVARIATE DATA MODELLING

Theory

In some cases that deal with a high number of variables, one must choose the variables of highest relevance for the process to avoid unnecessary variance increase. However, even with this precaution, and also combining the multivariogram with PCA to filter noise from the data, the required number of increments for example to collect for metallurgical testing is high (Dehaine et al. 2016; Dehaine & Filippov 2015). Hence, there is a legitimate need to reduce the sampling variance further. In particular, some variables which contribute to a major proportion of the global (multivariate) variability could potentially be less important for the process than other variables having a lower variability. To address this issue, it is a natural extension to combine the multivariate approach with multivariate data modelling, which is known to be able to focus on dimensionality reduced sub-spaces.

Indeed, considering the case where process performance heterogeneity could be linked to process variables (material properties, operating conditions) through the heterogeneity vector \( H_x \) by linear models such as:

\[
H_y = H_x \cdot A + r
\]

(5)

where \( H_y \) is the heterogeneity matrix of the \( q \) process responses/performance indexes (yield, recovery, etc.), \( A \) is a \( p \times q \) matrix containing the weights of each variables, which can be seen in this case as a "weights matrix" and \( r \) a residuals vector. This matrix could be obtained using various process modelling techniques or multivariate regression methods such as design of experiments (DOE), multiple linear regression (MLR) or preferentially Partial Least Squares (PLS) regression. Therefore, an estimate of \( H_y \) could be expressed as:

\[
\hat{H}_y = H_x \cdot A
\]

(6)

\(^1\) Univariate functions are noted in lowercase, whereas multivariate equivalents are noted in UPPERCASE.
The multivariogram of $Y$, noted $V_{Y_j}$, is expressed as:

$$V_{Y_j} = \frac{1}{2(N-j)} \sum_{i=1}^{N-j} (H_{Y_i} - H_{Y_{i+j}}) M (H_{Y_i} - H_{Y_{i+j}})^t, \quad j = 1, \ldots, N/2$$

(7)

where $M$ is the MD defined as the inverse of the variance-covariance matrix of $H_y$. Using the relationship defined in EQN (5), it follows:

$$V_{Y_j} = \frac{1}{2(N-j)} \sum_{i=1}^{N-j} (H_{X_i} - H_{X_{i+j}}) A M A^t (H_{X_i} - H_{X_{i+j}})^t, \quad j = 1, \ldots, N/2$$

(8)

At this stage, it is still not possible to calculate $V_{Y_j}$ as $H_y$ is, a priori, unknown and so is its variance-covariance matrix, and hence $M$ in the above equation. However, one could estimate $M$ using EQN (6), therefore allowing to calculate an estimate of $V_{Y_j}$:

$$\tilde{V}_{Y_j} = \frac{1}{2(N-j)} \sum_{i=1}^{N-j} (H_{X_i} - H_{X_{i+j}}) \tilde{M} A A^t (H_{X_i} - H_{X_{i+j}})^t, \quad j = 1, \ldots, N/2$$

(9)

where

$$\tilde{M} = [\text{Cov}(\tilde{H}_y)]^{-1} = [\text{Cov}(H_X \cdot A)]^{-1} = [A^t \text{Cov}(H_X) A]^{-1}$$

(10)

leading to:

$$\tilde{V}_{Y_j} = \frac{1}{2(N-j)} \sum_{i=1}^{N-j} (H_{X_i} - H_{X_{i+j}}) A [A^t \text{Cov}(H_X) A]^{-1} A^t (H_{X_i} - H_{X_{i+j}})^t, \quad j = 1, \ldots, N/2$$

(11)

$$\tilde{V}_{Y_j} = \frac{1}{2(N-j)} \sum_{i=1}^{N-j} (H_{X_i} - H_{X_{i+j}}) M' (H_{X_i} - H_{X_{i+j}})^t, \quad j = 1, \ldots, N/2$$

This last expression shows that the multivariogram $\tilde{V}_{Y_j}$ actually correspond to the multivariogram of $X$ with a weighted metric $M'$ defined as:

$$M' = A [A^t \text{Cov}(H_X) A]^{-1} A^t$$

(12)

Hence, by using this new metric in the calculation of $V_3$, one could assess an estimate of the multivariogram of $Y$ using only predictor variables data. In addition, depending on the accuracy of the prediction model, it could be used to characterise the process performance variability based on its input variables.

**PLS Regression**

PLS regression (PLS-R) is a statistical method that allows to model correlations between two sets of data, the multivariate $X$-data, called predictors, and the dependent $Y$-data, called responses, by regression. It replaces the classical MLR and allows direct correlations to be modelled between one, or more, response(s) $Y$ (PLS1 and PLS2) and a multivariate $X$ matrix containing the process data. PLS compensates for the troublesome colinearity between $X$-variables, usually the bane of MLR (Esbensen et al. 2002). PLS models can be viewed as interrelated PCA scores of the predictors, $t$, and the responses, $u$, maximizing $\text{cov}(t,u)$ (Hoskuldsson 1996). PLS reduces the $X$ dimensionality by extracting latent variables (linear combinations of all $X$-variables: PLS-components), which are both correlated to the responses while simultaneously capturing the largest possible amount of variation in the predictors. PLS $t$-scores can be viewed as "tilted PCA scores, rotated in $X$-space so that the covariance between input and output variables, $\text{cov}(t,u)$, is maximized. Full details on the statistical and data analytical features of PLS regression can be found in the chemometric literature (Hoskuldsson 1996; Martens & Naes 1992; Wold et al. 2001; Esbensen et al. 2002).
Therefore, PLS regression defines which of the predictor variables have the highest weight in predicting the Y responses for future data, based on a rigid calibration/validation framework. Using these PLS loading-weights (loading weights) in the new metric, defined in EON (12), to weight the multivariogram could potentially help to optimize the sampling procedure as PLS only extracts the systematic features of the process variations and descriptor variables. It thus dampens the effect of those variables that are less important in predicting the future responses according to their significance for the process responses (Y). Strictly speaking the method does not reduce the sampling variance but helps to find an optimized sampling frequency that is sufficient in predicting the overall process performance. Furthermore, if the PLS regression models display satisfactory validation results in predicting the Y responses, the regression coefficients could be used instead of the PLS weights to predict the overall process performance variability.

Application

Case study
In this work, we consider an industrial example from the mining industry. The industrial mineral processing plant considered is producing a clay minerals concentrate with a given moisture content (pulp density) for the paper, ceramics, paints, plastics and rubber industries. The clay resources are classified into distinct ore grades (denoted G1, G2, G3, G4 and G5 hereafter) as a function of mineralogical and physical properties. The final clay product is obtained after the refining of the clay matrix (i.e., ore) through a complex flowsheet comprising sorting, crushing, screening operations and cyclone refining loops while the waste is rejected as stones, gravels and sands (FIG 2). Some operating parameters such as streams density, cyclones pressure, conveyors product weight or flowrates are measured every 5 mins by sensors (flowmeters, pressure gauges, conveyor weightometers) located at distinct locations of the processing route (TABLE 1 an FIG 2). These sensors provide real-time data which is used to monitor process behaviour and make adjustments.

In terms of performance, the only way to assess the effectiveness of the process is to apply a metallurgical balance to evaluate the clay recovery. In addition, further analyses are performed to assess the final product quality, including moisture content measurements. These operations could be laborious and costly. Besides, they could only provide a posteriori information on the process performance, which is not suitable for a continuous process. This study is based on multivariate descriptor relationships with the specific objective of direct process performance variability prediction, in terms of clay recovery and product moisture content, using all instantaneously available process data (i.e. sensor data) according to their relevance as given by the PLS regression.

Method
Validation of the PLS models in the present study was performed using full cross-validation, also known as leave out-one (LOO) method. It consists in making as many sub-models as there are units, each time leaving out just one unit and only using it for testing; subsequently the unit is put back in X. The squared difference between the predicted and the known Y-value for each omitted unit is summed over all units and averaged, giving the usual validation Y-variance. The degree of prediction strength is evaluated using conventional modelling indices i.e. trend line slope and coefficient of determination (R²), pertaining to a fitted linear regression model between predicted (y) versus reference (x) values which must both be as close to 1 as possible. It should be noted that there is a fierce debate as to the merit of cross validation within chemometrics – and beyond. While we are using cross-validation here, strong arguments have been presented to use test set validation instead (Esbensen & Geladi 2010; Esbensen et al. 2002). The present demonstration of PLS-modelling in connection with applying appropriate variable weights in the multivariogram can be used with either validation approach without loss of general insight; test set validation can be implemented by authors who follow the more recent chemometric recommendations, ibid.

Results
In order to assess the influence of the multivariate sensors on process performance, PLS regression is applied to a set of sensor data collected during one week of production and to the corresponding process performance data (clay recovery, product pulp density). The number of factors to keep in the PLS regression model is given by the evolution of the explained Y-variance and the Predictive Residual Sum of Squares (PRESS) according to the number of factors in the model (FIG 3-A). For this example, a 7-component PLS model on the full variables set predicts both clay recovery and final product density with satisfactory validation results (slope close to 1 with R²= 0.80 and R²=0.94 respectively) as shown in the prediction versus observed responses plots in FIG 3-C&D. The performance prediction models are primarily carried by the belt weight data (SB-W15, SB-W, GB-W), the plant matrix feed flow (CM-F) and several other sensor variables also have minor, but significant influence. The developed models seem to apply to all ore grades suggesting
that all ore grades display similar X-Y data structures (covariance relationships). Note that these results refer to a PLS-R applied to raw X-Y data with the objective of defining regression models for plant performance prediction. Additionally, PLS-R has also been applied to the X-Y heterogeneity data in order to retrieve the weights to be used in the metric defined in EQN (12). The results are very similar to the PLS-R applied to raw X-Y data and hence, are not presented here. Outputs of both PLS-R are used thereafter depending on the objective, ie process performance or process variability prediction.

These regression models are applied to a set of 57 consecutive sensors measurements recorded over a night shift that will be used to conduct a variographic analysis. During this night shift, two ore grades where consecutively fed to the plant. Plotting together the observed and predicted responses (clay recovery and product density) suggests that PLS regression models are able to characterise the process performance (FIG 4 A-B). The comparison results for this dataset are shown in FIG 4 C-D. The results for the final product density (slope close to 1 and R²=0.96) are good whereas the results for the clay recovery model (slope of 0.8 and R²=0.77) are lower, but still acceptable for direct on-site process performance prediction based on real-time process data.

The matrix of PLS loadings-weights (W) as well as the matrix of PLS regression coefficients of all responses (C) are presented in TABLE 2 and TABLE 3, respectively. These outputs obtained from PLS regression applied to heterogeneity data can both be used as weighting matrices and used to define the weighted metric for the calculation of the multivariogram of X as shown in EQN (12). FIG 5-A shows a comparison between the multivariogram of X calculated from all the X-variables (sensor data) with a non-weighted metric (raw multivariogram, M = [Cov(H)]⁻¹), a weighted metric using PLS loading-weights (M' = W[WᵀCov(H)W]⁻¹Wᵀ) or a weighted metrics using the coefficient of the PLS regression models (M'' = C[ᵢCᵢCov(H)Cᵢ⁻¹Cᵢ⁻¹] or the reference multivariogram of Y calculated using experimental data (M = [Cov(H)]:=σ²). It can be seen that the raw multivariogram of X displays a large nugget effect of 9 with a sill around 16. This means that the estimated global sampling variance with this approach is very high. In terms of a sampling protocol, this implies that a large number of increments must be collected to obtain a reasonable sampling variation (Dehaine et al. 2016). The X-multivariogram weighted with the loading-weights matrix (W) display a lower but still relatively important nugget effect and a sill compared to the reference Y-multivariogram. However, the X-multivariogram weighted with the PLS regression coefficients matrix (C) is relatively close to the actual Y-multivariogram as highlighted in FIG 5-B. It can be seen that the predicted X-multivariogram displays the same characteristics, i.e. nugget effect, range and sill than the experimental Y-multivariogram. The global relative standard deviations of the sampling error (SDSE) display a similar trend (FIG 5-C). For a given number of units (increments) collected to make the final composite sample (Nu), the SDSE of the raw X-multivariogram is higher than that of the W-weighted X-multivariogram which itself is higher than the one of the C-weighted X-multivariogram:

$$SDSE[V_X(M)] > SDSE[V_X(M')] > SDSE[V_X(M'')]$$

The difference between these SDSE tends to decrease as the number increment collected increases, but the SDSE predicted by the PLS regression coefficients-weighted X-multivariogram stay the closest to the actual SDSE displayed by the Y-multivariogram. Hence, using the PLS model coefficients as weights in the new metric defined in EQN (12), to weight the multivariogram, can help to optimize the sampling procedure in terms of frequency, sampling mode and number of increments according to the actual global process performance by using the available real-time process data.

One other benefit of optimizing multivariate variographic analysis with process modelling is to use PLS regression models to decompose the variogram of each process performance response/index, therefore predicting process performance variability. This could be done separately for each process performance response by changing the metric in the calculation of the X-multivariogram, as done before. Considering the ith process performance response/index yi, an estimate of this response could be assessed using the PLS regression model:

$$\hat{y}_i = X.C_i$$

where Cᵢ refer to the regression coefficient vector of the ith process performance response y (see TABLE 4). Therefore, using these Cᵢ to weight the X-multivariogram allows to decompose the variogram of the corresponding yi. FIG 6 shows a comparison between the predicted and observed clay recovery and product density variograms. It can be seen that the predicted variograms are close to the experimental ones, especially for the product density. The predicted variograms display the same characteristics, i.e. nugget effect, range and sill than the experimental ones. These results are encouraging and suggest that this approach could potentially be used to predict process performance variability with time using real-time process data.
CONCLUSIONS

The multivariate variographic approach is a powerful tool when the overall time-variability of process streams must be summarized in terms of a large set of properties. However, it leads to an overestimation of the global sampling variance with regards to the process objective. Some of the variables accounting for a high proportion of the global variability may not be significant for the process performance. A method has been proposed, consisting in weighting the variables according to their significance for the process by combining the multivariogram with PLS regression to find which of the predictor variables have the highest weight in predicting the responses for future data, based on a rigid calibration/validation framework. Weighting the multivariogram using the PLS loading weights have helped to optimize the sampling frequency as PLS only extracts the systematic features of the process variations and descriptor variables. Using the PLS regression coefficient even allows accurate prediction of the overall process performance variability. Strictly speaking the method does not reduce the sampling variance but helps to find an optimized sampling procedure in terms of frequency, sampling mode and number of increments according to the actual overall process performance by using the available real-time process data. This approach has been successfully applied to a mineral processing case study but it has potentially many applications in the mining, feed and food, pharmaceutical or any other industry. It is simple and inexpensive and has great potential to reduce risk and to ensure better use and management of resources.

ACKNOWLEDGEMENTS

This work has been supported by the European FP7 project “Sustainable Technologies for Calcined Industrial Minerals in Europe” (STOICISM), grant NMP2-LA-2012-310645.

REFERENCES


Gy, P., 2004c. Sampling of discrete materials III. Quantitative approach—sampling of one-dimensional objects. Chemometrics and Intelligent Laboratory Systems, 74(1), pp.39–47. Available at:


ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOE</td>
<td>Design Of Experiments</td>
</tr>
<tr>
<td>ED</td>
<td>Euclidean Distance</td>
</tr>
<tr>
<td>LOO</td>
<td>Leave Out-One</td>
</tr>
<tr>
<td>MD</td>
<td>Mahalanobis Distance</td>
</tr>
<tr>
<td>MLR</td>
<td>Multiple Linear Regression</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Components Analysis</td>
</tr>
<tr>
<td>PLS</td>
<td>Partial Least Squares projection to latent structures</td>
</tr>
<tr>
<td>PLS-R</td>
<td>PLS-Regression</td>
</tr>
<tr>
<td>PRESS</td>
<td>Predictive Residual Sum of Squares</td>
</tr>
<tr>
<td>SDSE</td>
<td>Standard Deviation of Sampling Error</td>
</tr>
</tbody>
</table>
FIGURES

FIG 1 – Typical positively increasing variogram. The nugget effect is estimated by backward extrapolation of the variogram towards the origin, which provides an estimate of $V_0$. The range is the lag distance beyond which no further autocorrelation occurs, and the total sill, which is denoted as process variance $c h$, is the flat part of the variogram.

FIG 2 – Simplified flowsheet of the clay refining process. It can be seen as a size-classification process which removes the coarsest fractions of the clay matrix which are not valuable for clay recovery. Numbers in circles refers to the sensors location (see TABLE 1).
FIG 3 – PLS Regression models applied on $X$ and $Y$ data for plant performance prediction. A: Plot of $Y$-variance and PRESS values illustrating the cumulative contribution of the PLS components. B: Corresponding loadings-weights plot with proportions of total data variance modelled shown as (X%, Y%). C: Prediction vs. reference plot for final product density, using a 7-component model. D: Prediction vs. reference plot for clay recovery, using a 7-component model. Colours represent different types of feed ore.
FIG 4 – Process performance with time during the variographic experiment conducted over two shifts with change in feed ore grade. A: Predicted clay recovery based on the PLS model compared with reference clay recovery (calculated). B: Predicted product density based on the PLS model compared with reference product density (online measurements). C: Predicted vs observed clay recovery with trend line. D: Predicted vs observed product density with trend line.
FIG 5 – A: Comparison between the multivariogram of $X$ calculated from all the $X$-variables (sensor data) with a non-weighted metric ($M = [\text{Cov}(H_X)]^{-1}$), a weighted metric using PLS loading-weights ($M = W[W^T\text{Cov}(H_X)W]^{-1}W^T$) and a weighted metrics using the regression coefficient of the PLS models ($M = C[C^T\text{Cov}(H_X)C]^{-1}C^T$) and the true multivariogram of $Y$. B: Close-up of the multivariogram values from 0 to 4. C: Comparison between the corresponding relative standard deviation of the sampling error (SDSE) as a function of the number of units that are collected to make the final composite sample for a systematic sampling mode. D: Close-up of the section of 3 to 57 units in the final sample.
FIG 6 – Comparison between process performance variograms for both $y$-variables (observed and predicted) using the multivariogram with the metric defined by the PLS models. Note that these variograms have been calculated from the raw variables and not their corresponding heterogeneity.

TABLES

TABLE 1
Process variables measured by the on-line sensors used in the PLS regression (X-matrix). For sensors location see Figure 2.

<table>
<thead>
<tr>
<th>N°</th>
<th>Variables</th>
<th>Code</th>
<th>Units</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Calculated Matrix Feed Flow</td>
<td>CM-F</td>
<td>Tons/h</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Stone Belt Weigher</td>
<td>STB-W</td>
<td>Tons/h</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Gravel Belt Weigher</td>
<td>GB-W</td>
<td>Tons/h</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Sand Belt Weigher</td>
<td>SB-W</td>
<td>Tons/h</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Secondary Cyclone Product Density</td>
<td>2CP-D</td>
<td>kg/m$^3$</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>Secondary Product Flow</td>
<td>2P-F</td>
<td>m$^3$/h</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>Secondary Residue Flow</td>
<td>2R-F</td>
<td>m$^3$/h</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>Stone Belt Weigher - 15 minutes average</td>
<td>STB-W15A</td>
<td>Tons/h</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>Gravel Belt Weigher - 15 minutes average</td>
<td>GB-W15A</td>
<td>Tons/h</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>Sand Belt Weigher - 15 minutes average</td>
<td>SB-W15A</td>
<td>Tons/h</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>Cavex Cyclones Pressure</td>
<td>CC-P</td>
<td>Bar</td>
<td>7</td>
</tr>
<tr>
<td>12</td>
<td>LP Water Flow Rate</td>
<td>LPW-F</td>
<td>m$^3$/h</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>HP Water Flow</td>
<td>HPW-F</td>
<td>m$^3$/h</td>
<td>9</td>
</tr>
<tr>
<td>14</td>
<td>Primary Cyclones Feed Pressure</td>
<td>PCF-P</td>
<td>Bar</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>Secondary Cyclones Feed Pressure</td>
<td>SCF-P</td>
<td>Bar</td>
<td>11</td>
</tr>
</tbody>
</table>
TABLE 2
Loading weights obtained from the 7-component PLS-R on X-Y heterogeneity data which can be used as a 15 x 7 weight matrix (W).

<table>
<thead>
<tr>
<th>Variable</th>
<th>X Weights 1</th>
<th>X Weights 2</th>
<th>X Weights 3</th>
<th>X Weights 4</th>
<th>X Weights 5</th>
<th>X Weights 6</th>
<th>X Weights 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM-F</td>
<td>0.6278</td>
<td>0.0145</td>
<td>0.6283</td>
<td>0.5642</td>
<td>0.3355</td>
<td>0.3268</td>
<td>-0.1997</td>
</tr>
<tr>
<td>STB-W</td>
<td>-0.0962</td>
<td>0.5247</td>
<td>0.2318</td>
<td>0.1638</td>
<td>-0.3890</td>
<td>-0.5560</td>
<td>0.1061</td>
</tr>
<tr>
<td>GB-W</td>
<td>-0.0235</td>
<td>0.5752</td>
<td>-0.3314</td>
<td>0.0101</td>
<td>0.1654</td>
<td>0.4468</td>
<td>-0.2280</td>
</tr>
<tr>
<td>SB-W</td>
<td>0.5579</td>
<td>0.0878</td>
<td>-0.4891</td>
<td>-0.5200</td>
<td>-0.4532</td>
<td>-0.1249</td>
<td>0.0968</td>
</tr>
<tr>
<td>2CP-D</td>
<td>-0.0469</td>
<td>0.0741</td>
<td>-0.3295</td>
<td>0.5316</td>
<td>0.0894</td>
<td>0.1460</td>
<td>0.6833</td>
</tr>
<tr>
<td>2P-F</td>
<td>-0.3576</td>
<td>0.1518</td>
<td>0.6761</td>
<td>-0.1965</td>
<td>-0.0818</td>
<td>0.0730</td>
<td>0.2642</td>
</tr>
<tr>
<td>2R-F</td>
<td>-0.0409</td>
<td>-0.0672</td>
<td>0.0491</td>
<td>-0.1196</td>
<td>-0.0622</td>
<td>0.4429</td>
<td>0.3709</td>
</tr>
<tr>
<td>STB-W15A</td>
<td>0.1719</td>
<td>0.3538</td>
<td>0.3272</td>
<td>0.2458</td>
<td>-0.3141</td>
<td>-0.3071</td>
<td>0.0663</td>
</tr>
<tr>
<td>GB-W15A</td>
<td>0.1100</td>
<td>0.4710</td>
<td>-0.2929</td>
<td>0.0166</td>
<td>0.1003</td>
<td>0.3630</td>
<td>-0.0306</td>
</tr>
<tr>
<td>SB-W15A</td>
<td>0.6936</td>
<td>-0.1188</td>
<td>0.0777</td>
<td>-0.0472</td>
<td>-0.1455</td>
<td>-0.0717</td>
<td>0.2932</td>
</tr>
<tr>
<td>CC-P</td>
<td>0.0845</td>
<td>0.1769</td>
<td>0.2355</td>
<td>-0.5138</td>
<td>0.2510</td>
<td>0.1441</td>
<td>-0.1546</td>
</tr>
<tr>
<td>LPW-F</td>
<td>0.0407</td>
<td>0.1014</td>
<td>-0.2668</td>
<td>0.0503</td>
<td>0.9376</td>
<td>-0.4883</td>
<td>0.2566</td>
</tr>
<tr>
<td>HPW-F</td>
<td>0.1206</td>
<td>0.0902</td>
<td>0.1700</td>
<td>-0.4399</td>
<td>0.4695</td>
<td>-0.3892</td>
<td>-0.0651</td>
</tr>
<tr>
<td>PCF-P</td>
<td>0.0690</td>
<td>0.0315</td>
<td>0.2628</td>
<td>-0.3307</td>
<td>0.1462</td>
<td>-0.0333</td>
<td>0.3269</td>
</tr>
<tr>
<td>SCF-P</td>
<td>-0.0178</td>
<td>0.0853</td>
<td>0.2470</td>
<td>-0.4261</td>
<td>0.0669</td>
<td>0.4477</td>
<td>0.3140</td>
</tr>
</tbody>
</table>

TABLE 3
Regression coefficients obtained from the PLS-R on X-Y heterogeneity data which can be used as a 15 x 2 weight matrix (C).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Density</td>
<td>0.12</td>
<td>-0.05</td>
<td>&lt;0.01</td>
<td>-0.04</td>
<td>-5.95</td>
<td>&lt;0.01</td>
<td>0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>&lt;0.01</td>
<td>-0.01</td>
<td>-0.02</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Clay Recovery</td>
<td>2.48</td>
<td>-1.17</td>
<td>-0.11</td>
<td>-0.95</td>
<td>-385.67</td>
<td>0.14</td>
<td>0.21</td>
<td>-0.06</td>
<td>-0.11</td>
<td>0.28</td>
<td>0.06</td>
<td>0.14</td>
<td>-0.21</td>
<td>-0.38</td>
<td>-0.09</td>
</tr>
</tbody>
</table>

TABLE 4
Regression coefficients obtained from the PLS-R on raw X-Y data which can be used separately as 15 x 1 weight vectors (C).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Density</td>
<td>0.29</td>
<td>-0.28</td>
<td>-0.20</td>
<td>-0.35</td>
<td>-7.48</td>
<td>-0.01</td>
<td>0.12</td>
<td>-0.20</td>
<td>0.16</td>
<td>-1.11</td>
<td>0.03</td>
<td>-0.06</td>
<td>-19.42</td>
<td>-4.62</td>
<td>-3.39</td>
</tr>
<tr>
<td>Clay Recovery</td>
<td>0.05</td>
<td>-0.06</td>
<td>-0.04</td>
<td>-0.07</td>
<td>-3.81</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.04</td>
<td>0.02</td>
<td>0.59</td>
<td>0.01</td>
<td>-0.01</td>
<td>-3.39</td>
<td>-4.62</td>
<td>-3.39</td>
</tr>
</tbody>
</table>