CHEMICAL ENGINEERING TRANSACTIONS
VOL. 100, 2023
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Sauro Pierucci, Flavio Manenti

ISBN 979-12-81206-03-8; **ISSN** 2283-9216
DOI: 10.3303/CET23100030

Real Time Quality Assessment of General Purpose Polystyrene (GPPS) by means of Multiblock-PLS Applied on On-line Sensors Data

Lorenzo Strani[a], Francesco Bonacini[b,*], Angelo Ferrando[b], Andrea Perolo[b], Daniele Tanzilli[a,c], Raffaele Vitale[c], Marina Cocchi[a]

[a]Department of Chemical and Geological Sciences, University of Modena and Reggio Emilia, Via 4 Campi 103, 41125 Modena, Italy
[b]Research Center, Versalis (ENI) S.p.A., Via Taliercio 14, 46100 Mantova, Italy
[c]Centre National de la Recherche Scientifique (CNRS), (LASIRE), Cité Scientifique, University Lille, F-59000 Lille, France
francesco.bonacini@versalis.eni.com

In the petrochemical industry, in order to control the final product quality over time and to detect potential plant failures, the amount of lab (off-line) analysis performed every day is very demanding in terms of resources and time. Hence, at/in-line monitoring can be an efficient solution to decrease chemical wastes and operators' efforts and to perform a fast detection of deviations from normal operative conditions. Moving toward this implementation requires both installation of analytical sensors and the development of models capable to predict in real time the quality parameters of the polymers based on both process and analytical sensors. The primary aim of the current work has been the development of real time monitoring models by advanced chemometric tools for the prediction of a General Purpose PolyStyrene (GPPS) quality property, fusing Near Infrared (NIR) and process sensors data. In the plant considered, in addition to standard process sensors, along the GPPS production line, operating in continuous, two NIR probes are installed in-line. After the arrangement of the available data in different blocks, aiming at studying the specific contribution of the two types of sensors and of the main phases of the process, Multiblock-PLS (MB-PLS) method was employed to fuse the different blocks and to assess which were the most relevant sensors and plant phases for the prediction of the two quality parameters. Good prediction performances were achieved, allowing identifying the most significant data blocks for the GPPS quality prediction. Moreover, prediction errors obtained by models computed without considering blocks of data belonging to the final stages of the process were similar to those involving all the available data blocks. Therefore, a good real time assessment of the GPPS quality can be obtained even before the production is completed, which is very promising in view of minimizing the number of off-line laboratory analyses.

1. Introduction

Among the many concerns petrochemical industries have to deal with, the high number of expensive, time-consuming and wasteful laboratory tests performed on the final products every day is one of the biggest. Nevertheless, these analyses are fundamental, since according to the results obtained it is possible to assess whether the production is maintaining the expected quality standards or not. To both reduce the daily laboratory analyses performed and improve the quality of the process monitoring, the development of chemometric models that can predict the results of laboratory measurements, based on on-line sensors data, is crucial. In this way, it is possible to obtain a real time estimation of quality parameters normally assessed off-line and, at the same time, detect possible plant faults or deviations from normal operative conditions (Bhattacharya, 2005; Macho and Larrechi, 2002; Strani et al., 2021; Zhao et al., 2006). The predictive models could be based on the data that are continuously collected by the huge number of different kinds of process sensors usually installed throughout the production line but can greatly benefit from additional "chemical" on/in-line sensors, such as spectroscopic probes. Therefore, multivariate data analysis methods that can handle data of different natures, such as infrared spectra and process sensors data, at the same time are necessary. In this context, multiblock

Paper Received: 15 January 2023; Revised: 28 March 2023; Accepted: 12 May 2023
Please cite this article as: Strani L., Bonacini F., Ferrando A., Perolo A., Tanzilli D., Vitale R., Cocchi M., 2023, Real Time Quality Assessment of General Purpose Polystyrene (GPPS) by means of Multiblock-PLS Applied on On-line Sensors Data, Chemical Engineering Transactions, 100, 175-180 DOI:10.3303/CET23100030

techniques applied to data acquired on industrial processes have been proven to be reliable and accurate both in estimating the quality parameters of the final product and in monitoring the process (Qin et al., 2001; Strani et al., 2022; Wangen and Kowalski, 1989). One of the most common multiblock methods is Multiblock-Partial Least Squares (MB-PLS) regression (Westerhuis et al., 1998), an extension of the PLS regression technique (Wold et al., 1983) taking care of ensuring equal potential contribution from each single data block, and widely used in several industrial applications. The main advantage of partitioning data in different blocks is to improve the understanding of the information contained in the data, enhancing at the same time data visualization, responses predictions and identification of the variables that most influence the model (Alinaghi et al., 2019; Biancolillo et al., 2014; Mage et al., 2019; Song et al., 2020). In this respect, the aim of the current study was to develop real time monitoring models with a chemometric multiblock approach for the prediction of a General Purpose Polystyrene (GPPS) quality property, fusing Near Infrared (NIR) and process sensors data.

2. Materials and Methods

2.1 Data collection

The GPPS production has been monitored in a production line of a facility owned by the Italian petrochemical company Versalis (ENI group). The process, which works in continuous, involves the production of 9 different grades of GPPS, each one with slightly different formulations and properties. Along the pipeline are installed 49 process sensors (PS) that measure Temperatures, flows, pressures and engine efforts in rounds per minute. The data acquired by PS has been partitioned into four different sub-datasets (blocks), one for each of the main steps of the process, namely pre-polymerization reactor (PR), main reactor (MR), devolatilizer (DV) and cut zone (CZ). Along the process line are also installed two NIR probes, one in the pipe for the recovery of condensed reagents (NIR CR) and one in the CZ area (NIR CZ). A Matrix FT-NIR spectrometer (Bruker Optics, Milan, Italy) equipped with optical fibers (length: 100 m, diameter: 600 μm) is used to collect the spectra. The optical fiber probes (HT immersion probe, Drawing-no.661.2350_1, Hellma GmbH and Co. KG, Müllheim, Germany) are directly connected to both acquisition sites positioned on the process pipe. Spectra were acquired in transmission mode in the 12500–4000 cm^{-1} spectral range, with a nominal resolution of 4 cm^{-1} (64 scans per sample). To evaluate the GPPS quality, one of the most important routine analyses performed on the final product has been considered as a reference measurement. However, because of confidential agreement restrictions with the Company, its real name cannot be revealed, thus it will be referred to as “reference property”. This property is related to the physical features of the product (expressed in grams), and assessed off-line by gathering GPPS samples two-three times per day. In this study, were used the data collected every hour from both PS and NIR probes. In particular, for PCA and real-time prediction models validation, all the available data were used, whereas for the models calibration and first validation phase, only data related to final products on which the quality was evaluated in the laboratory were considered.

2.2 Data analysis

Principal Component Analysis (PCA) (Wold et al., 1987) was performed on NIR CR, NIR MR and PS data blocks, as a first exploratory step to inspect the data and extract possible relevant information from it. Prior to PCA, NIR spectra were reduced in range (9500 – 4800 cm^{-1}) and pre-processed with Standard Normal Variate (SNV) and Mean Center, whereas on PS data was applied Autoscaling. After this step, the three different data blocks were merged into a single dataset, when merging the data, the delay time due to the positioning of the different sensors along the production line has been considered to always match, in the same data row, observation referring to the same material. In other words, each data point is related to information acquired at different times, but it is matched to the same processed material. Multiblock-Partial Least Squares (MB-PLS) regression was used to compute predictive models of the reference property and assess at the same time which data blocks are contributing more to achieve good prediction. In MB-PLS, after the individual pre-processing of the single blocks, each block is scaled to unit block variance, in order to assure fair block contribution. Data were split into a calibration set, containing acquisitions from January to July 2021 and from December 2021 to January 2022, and a first validation set, including acquisitions from August to November 2021 and from February to April 2022. To assess the number of PLS components of each model, Venetian blinds cross-validation with ten cancellation groups was used and as a criterion the best compromise among low Root Mean Square Error in Cross-Validation (RMSECV), bias and RMSEC/RMSECV ratio close to one. Model performance was determined by means of both RMSECV and Root Mean Square Error in Prediction (RMSEP). The contribution of each block, as well as the contribution of each variable, were assessed by examining the PLS regression coefficients and Variable Importance in Prediction (VIP) values. All the analyses were executed using routines and toolboxes implemented in the MATLAB environment (the Mathworks Inc., Natick, MA, USA). MB-PLS has been performed through the PLS-Toolbox version 8.9 (Eigenvector Research Inc., Wenatchee, WA, United States).

3. Results and discussion

3.1 PCA results

The first PCA model, computed on PS data, showed a sample separation mainly based on the different grades of GPPS. However, PCA, when performed on NIR datasets, highlighted additional information not extractable from the previous model. As an example, considering NIR CR dataset, it is possible to observe an increasing trend in the scores of the first Principal Component (PC) up to end of November 2021 after which all values become negatives (Figure 1a). In particular, in the first half of the time range inspected, it is observed a slow but constant increase of the scores over time, followed by a plateau. Then, in the first half of December, it is supposed that an event occurred responsible for the sudden change toward negative scores values. By studying the NIR absorption bands responsible for this change, as reported in the loadings plot in Figure 1b, and considering the date in which the sudden change took place, this trend was induced by the substitution of both source and molecular sieves of the NIR spectrometer. Hence, the slow but constantly increasing trend of the scores is due to the rise in humidity over time, a trend that disappears after the molecular sieves were changed. In the same way, the big change in December spectra is caused by the instrument source substitution. However, this first analysis emphasized the power of PCA, able to detect and highlight even the slightest changes in the spectra over time.

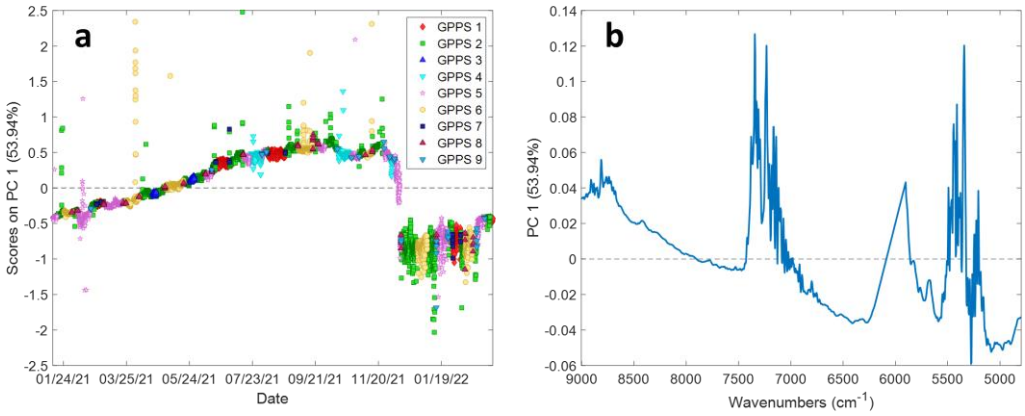

Figure 1: Scores as a function of time (A) and loadings (B) on PC1 for the NIR CR dataset.

More interesting information were extracted by the PCA analysis performed on NIR CZ, where PC3 showed a constant increase of the scores throughout the time range (Figure 2a), not visible in the previous analyses. Looking at the loadings plot (Figure 2b) it can be observed that the NIR band mainly responsible for this particular trend is the one at 7100 cm^{-1}. Since this band can be ascribed to the water O-H bond, it could indicate a humidity accumulation in the last step of the process. PS related to the final area of the process were not able to detect this feature, as sensors that measure humidity are not present.

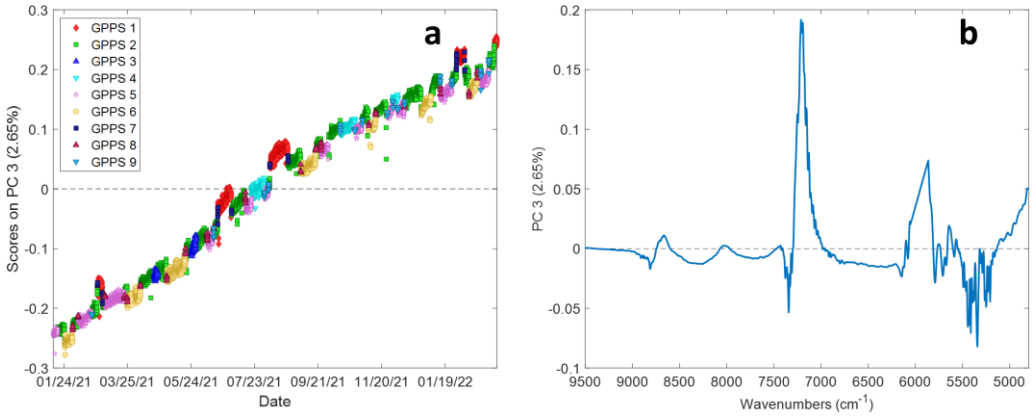

Figure 2: Scores as a function of time (A) and loadings (B) on PC3 for the NIR CZ dataset.

Another interesting information can be retrieved from the PC1 scores plot as a function of time, coloring the samples according to the quality of the final product, assessed by the reference method (Figure not shown for the sake of brevity). The clear separation between samples is explained by the presence (negative scores values) or absence (positive scores values) of a certain type of oil in the formulation. In fact, in the loadings the band at 5800 cm^{-1} is linked to this compound. However, the most interesting feature is the fact that the samples colored in red, corresponding to the final product showing non-optimal quality values, are mostly present during the transition between the formulations with and without oil.

3.2 Predictive models results

MB-PLS regression was used to compute models for the prediction in real time of the reference property, thus giving information about the quality of GPPS. Six different MB-PLS models were calculated, considering every time different block combinations, with the aim to evaluate if a reasonable prediction of the reference property could be achieved before the production is complete. Therefore, one model was calculated without using data collected in CZ stage and another one without the DV and CZ data. Moreover, with the same logic, models built with different combinations of PS blocks, i.e. PR, MR, DV and CZ, without considering the NIR data, were also calculated. In the latter case, we were aiming at assessing if the NIR contribution was or not critical. The results obtained are reported in Table 1. As expected, models that involve data acquired in the final stages of the process provide the best results. However, the RMSEP is just 0.5 points lower than the one obtained by models without the end-process blocks, meaning that it is possible to obtain a fair good estimation of the reference property when the product is halfway through the process. Furthermore, it is interesting to observe how NIR data does not significantly improve the prediction of quality, suggesting that, for this property (more linked to physical than chemical variability), PS data are sufficient to obtain a good prediction performance. The results also suggest that DV block is not very important for the prediction, as its presence/absence does not decrease/increase the prediction error significantly.

Table 1: Results of MB-PLS regression

Blocks used for model computation	LVs	RMSEC (g)	RMSECV(g)	RMSEP (g)
PR, NIR CR, MR, DV, NIR CZ, CZ	6	1.56	1.68	1.92
PR, NIR CR, MR, DV	10	2.01	2.22	2.6
PR, NIR CR, MR	9	2	2.13	2.5
PR, MR, DV, CZ	11	1.46	1.56	1.86
PR, MR, DV	9	1.94	2.03	2.52
PR, MR	7	2.1	2.16	2.68

In Figure 4a is shown the plot of predicted vs. measured values resulting from the model that provides the lowest RMSEP, namely the one in which were used all the PS blocks but none of the NIR blocks. The different colors indicate the different GPPS grades, each one with a different range of reference property. It can be noticed that the lower the reference property value, the lower the prediction error. In fact, for instance, products 5 and 6, the ones that have high values of reference property show more scattered samples, around the 1:1 line, than products 1 to 4. This is consistent with the results of the reference property analysis, which also present lower accuracy for high reference property values. Note that GPPS number 7, 8 and 9 are produced only when there is a change in the formulation (is evident in Figures 1a and 2a), so high quality is not expected for the related final product. However, they have been included in the model, both in calibration and validation sets, in order to better span the range of the reference property values. Figure 4b illustrates the regression coefficients associated with the four different blocks considered, i.e. PR, MR, DV and CZ. The higher, in absolute value, the regression coefficient of a variable (whose names have not been revealed because of confidential agreement restrictions with the company), the higher its influence on the reference property prediction. Therefore, through this plot it is possible to assess which are the most important PS (or the spectral regions in the case of models that involve NIR data) to strictly control during the production. In this case, a sensor installed in PR stage, related to the introduction of a certain compound in the pipeline, resulted in the most important variable for the prediction, followed by other sensors present in MR and CZ stages. Moreover, it is confirmed how sensors installed in the DV stage have a low influence on the model's performance prediction.

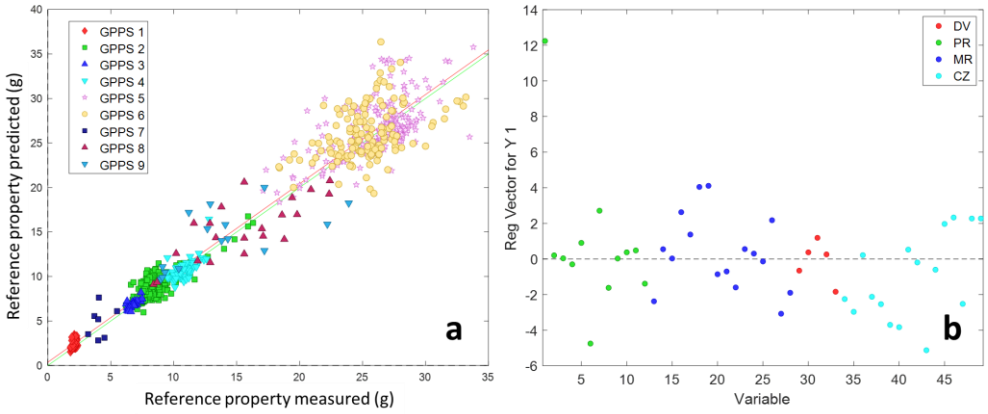

Figure 3: MB-PLS results for reference property prediction resulting from the model calculated with all the PS blocks but none of the NIR blocks. Predicted vs. measured value plot (a); regression coefficient (b).

Lastly, Figure 5a shows the predicted values of reference property obtained by applying the above mentioned best model (comprising all PS data blocks) for the time points for which the reference analysis measurements were not acquired off-line (laboratory analysis). It can be observed how the predictions follow the trend of the reference analysis for all the products. Specifically, looking at Figure 5b, which is a zoom of the right part of Figure 5a, the model makes very low prediction errors, such as for GPPS number 1 and 2, whereas, for other products, especially those in which oil is present in the formulation, the model makes higher errors. Despite this, it can be seen that the two trends are comparable. The presence of a systematic error shows how the model can follow the changes of reference property over time, but for certain products, it underestimates or overestimates its values systematically.

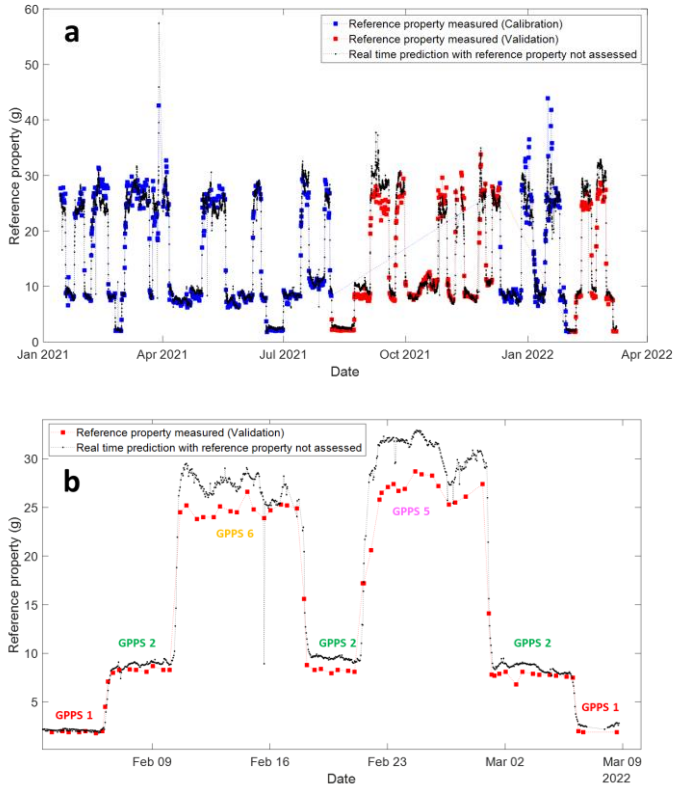

Figure 4: Real time predictions of reference property (time evolution of the measured and predicted values) obtained by the model calculated with all the PS blocks but none of the NIR blocks. Whole time range (a) and zoom on last inspected period (b).

4. Conclusions

In this work has been observed how the development of predictive models through multiblock regression methods allowed obtaining good prediction performance of a GPPS reference property, highlighting at the same time the most influent plant areas and sensors. In particular, an explorative data analysis on NIR data, performed by PCA, showed trends, not visible by standard PS, that highlighted possible future problems with humidity in the pipeline. Then, MB-PLS has been used to compute different models for the prediction in real time of the reference property. The results obtained showed, in addition to good prediction performances, which are the most important data blocks, related to different process stages. Finally, it has been observed that a good estimation of the reference property can be obtained also by models built without considering data from the final steps of the process. These results are of great industrial interest, as GPPS quality could be assessed before its production is concluded.

In summary, multivariate tools and specifically multiblock approaches can help the petrochemical industry to assess the quality of the final product in real time, reducing the number of time-consuming and wasteful off-line laboratory analyses and facilitating the plant operators in detecting faults and deviations from normal operative conditions.

References

Alinaghi M., Bertram H.C., Brunse A., Smilde A.K., Westerhuis J.A., 2020, Common and distinct variation in data fusion of designed experimental data, Metabolomics, 16, 1-11.

Bhattacharya T, 2005, Prediction of silicon content in blast furnace hot metal using partial least squares (PLS), ISIJ international, 45(12), 1943-1945.

Biancolillo A., Bucci R., Magrì A.L., Magrì A.D., Marini F., 2014, Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication, Analytica chimica acta, 820, 23-31.

Macho S., Larrechi M.S. 2002, Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. TrAC Trends in Analytical Chemistry, 21(12), 799-806.

Måge I., Smilde A.K., van der Kloet F.M., 2019, Performance of methods that separate common and distinct variation in multiple data blocks, Journal of Chemometrics, 33(1), e3085.

Qin S. J., Valle S., Piovoso M. J., 2001, On unifying multiblock analysis with application to decentralized process monitoring, Journal of Chemometrics: A Journal of the Chemometrics Society, 15(9), 715-742.

Song Y., Westerhuis J.A., Smilde A.K., 2020, Separating common (global and local) and distinct variation in multiple mixed types data sets, Journal of Chemometrics, 34(1), e3197.

Strani L., Mantovani E., Bonacini F., Marini F., Cocchi M., 2021, Fusing NIR and Process Sensors Data for Polymer Production Monitoring. Frontiers in Chemistry, 785.

Strani L., Vitale R., Tanzilli D., Bonacini F., Perolo A., Mantovani E., Ferrando A., Cocchi M., 2022, A Multiblock Approach to Fuse Process and Near-Infrared Sensors for On-Line Prediction of Polymer Properties, Sensors, 22(4), 1436.

Wangen L.E., Kowalski B.R., 1989, A multiblock partial least squares algorithm for investigating complex chemical systems, Journal of chemometrics, 3(1), 3-20.

Westerhuis J.A., Gurden S.P., Smilde A.K., 2000, Generalized Contribution Plots in Multivariate Statistical Process Monitoring, Chemometrics and Intelligent Laboratory Systems, 51 (1), 95–114.

Wold S., Esbensen K.H., Geladi P., 1987, Principal component analysis. Chemometrics and Intelligent Laboratory Systems, 2, 37-52.

Wold S., Martens H., Wold H., (Ed) 1983, The multivariate calibration problem in chemistry solved by the PLS method, Chapter in Matrix Pencils (Ed), Springer, Berlin, Heidelberg, 286-293.

Zhao S. J., Zhang J., Xu Y.M. 2006, Performance monitoring of processes with multiple operating modes through multiple PLS models, Journal of process Control, 16(7), 763-772.