Processing Analytical Control (PAT) in Green leaf threshed processing line with hyperspectral images.

Marcos A. T. Volochen (PQ)*,Larissa R. Terra (PQ), Jose Marcelino Netto (PQ), Jorge A. A. Quintero(PQ), Samuel Kaiser (PQ).

British American Tobacco (BAT) – Rod. 471 s/n. Santa Cruz do Sul, RS.
Email: marcos_antonio_teixeira_volochen1_external@bat.com

keywords: HSI, Class-modeling, Flue Cured Virginia, Air Cured Burley

INTRODUCTION

In the context of Industry 4.0, all data generated by equipment are leveraged to benefit the company. This drives the development of a robust control chain to ensure the quality, reliability, and effective utilization of data throughout its lifecycle. In this study, the data source is a hyperspectral camera installed on a processing line. However, mapping all potential factors influencing the captured images is unfeasible. Therefore, filtering unreliable data from the pipeline is crucial to avoid misinterpretation and poor decision-making. Chemometric analyses can be employed to ensure only reliable results are released.

RESULTS AND DISCUSSION

To evaluate whether a spectrum is an outlier, three methods were applied: a DD-SIMCA model developed for Flue-cured tobacco, and two statistics — Q-residual reduced and Hotelling's T^2 reduced — obtained from the latent space of a PLS regression model targeting total alkaloids. Both models were constructed using the same dataset, which included 3,647 spectra of Flue-cured tobacco. The preprocessing steps were identical: application of the first derivative using the Savitzky-Golay method, followed by standard normal variate (SNV) correction, spectral alignment using the icoshift algorithm, and mean-centering. For the DD-SIMCA model, a significance level of $\alpha = 0.01$ and 18 principal components were used.

The validation set was collected during the 2024 crop season, when outlier predictions diverged from results previously observed for the same tobacco grades. A total of 599 spectra were acquired. Using the DD-SIMCA approach, 65% of the validation dataset were correctly classified as not belonging to the model group. In comparison, the Q-residual and Hotelling's T^2 statistics diagnosed 80% of the same spectra as outliers when applied to the latent space of the PLS model.

Subsequently, these methods were applied to the full dataset of images collected during the 2024 crop, comprising 197,925 Flue-cured samples. The DD-SIMCA approach identified 572 outlier images, representing 0.3% of the samples. The Q-residual and Hotelling's T^2 statistics flagged over 7,046 samples, accounting for 3.5% of the dataset. These included not only true outliers but also samples exhibiting variability not well captured by the model.

CONCLUSION

While DD-SIMCA was effective for excluding process outliers, the combination of Q-residual and Hotelling's T^2 performed comparably and offered additional insights. These statistics proved to be valuable tools for identifying new sources of variability and served as indicators of when the model might require retraining. As such, the two approaches are complementary: DD-SIMCA ensures robustness against known deviations, while Q and T^2 help monitor and adapt to evolving process conditions.

ACKNOWLEDGMENTS

We are grateful to BAT GLT and BAT GLAD departments for all their support.

[1] Rodionova, O. Ye.; Titova, A. V.; Pomerantsev, A. L. Trends in Analytical Chemistry, v. 78 (2016), p. 17–22.
[2] Vitale, R. et al. Analytica chimica acta , v. 1270 (2023), p. 341304–341304.