Metin Türkay and Erdal Aydın (Eds.)
Proceedings of the 31st European Symposium on Computer Aided Process Engineering
June 6th to 9th, 2021, Istanbul, Turkey. © 2021 Elsevier B.V. All rights reserved.

Application of data science to study fluorine losses in the phosphate industry

Houda Ariba,[a,b] Paul Vanabelle,[c] Salah Benaly,[d] , Thomas Henry,[b] Cédric R. André,[b] Grégoire Leonard[a,*]

[a]*Université de Liège, Place du 20-Août 7, Liège 4000, Belgium*

[b]*Prayon, Rue Joseph Wauters 144, Engis 4480, Belgium*

[c] *Cetic, Avenue Jean Mermoz 28, Charleroi 6041, Belgium*

[d] *Université Mohammed VI Polytechnique, Lot 660, Hay Moulay Rachid, Ben Guerir 43150,Maroc*

g.leonard@uliege.be

Abstract

Artificial intelligence has become an attractive science for companies as it allows effective data analysis, which helps to improve the manufacturing processes. The aim of this work is to study fluorine losses in a phosphoric acid unit by applying data science methods to process data. Conductivity was used as an indirect measure of fluorine losses in each recovery cycle. After a pre-processing of the data, a Gaussian Mixture Models (GMM) clustering algorithm was applied. Two clusters were found in the data: one with limited losses, and the other with significant losses. In addition, a ratio (R) was created from measurement data to identify the level of fluorine loss compared to fluorine gain during a time step. This ratio R is used in turn to determine whether the plant generates an acceptable amount of fluorine losses.

Keywords: Clustering, Fluorine losses, Phosphoric acid unit, Data analysis

1. Introduction

In recent years, the interest in data science has risen considerably across industries worldwide (Diez-Olivan et al., 2019; Lee and Shin, 2020). Indeed, numerous sensors record real-time data at all stages of the process. When these data are accumulated over time, they can be used as an input to data science algorithms to model complex, non-linear processes. In turn, these models help process engineers to better understand how different variables affect their plant processes.

Clustering is a data science method for analysing data that groups information into clusters with similar properties or features. Clustering is achieved by many unsupervised machine learning algorithms that can be used in various way in different industries. For example, Liu et al.(2018) applied the clustering method in order to evaluate the energy consumption in different companies. Their study allowed to position companies according to their level (high or low) of energy consumption. Sancho et al. (2020) used a clustering method to classify crude oils according to their physico-chemical properties. Zhang et al. (2017) used clustering to discriminate between operational status, i.e. the distinction between operating in a transitory regime and in a permanent regime.

The present article is an application of data science in the chemical industry on a phosphate production process. More specifically, the study will focus on the fluorine losses in a concentration unit of phosphoric acid. The goal is to qualify these losses and to understand their relationship with the process variables. In parallel, a ratio (R) was defined to quantify the losses and ease the readability and discussion of raw data. This indicator will be discussed and put in relation with the clustering results. Perspectives are a better control of the loss of a product that can be valorised and a reduction of the use of caustic soda necessary to neutralize the acidity generated by the fluorine loss.

This paper is organized as follows. In section 2, the fluorine recovery process in the considered phosphoric acid concentration plant will be described. Then, in section 3, the methodology is discussed. Section 4 shows the main results that were achieved and section 5 concludes and suggests some perspectives.

2. Background

The process of fluorine recovery (Fig.1) is divided into three main steps. First, phosphoric acid is concentrated via a vacuum evaporation, leading to a liquid product that is the concentrated phosphoric acid on one hand, and to the emission of a gaseous effluent that contains fluorine on the other hand. Second, this fluorine is recovered in liquid phase in an absorption tower. Finally, the remaining amount of fluorine from the gas phase is transformed into liquid stream rejects by another recovery system. This final recovery system is itself composed of two main units: the PraySep with a spray at its entrance is used to recover the fluocilisic acid droplets contained in the gases, and a condenser is used to condense a maximum of the remaining gas. The fluorine losses are detected by a conductivity measurement in the condenser guard tank. Moreover, the amount of recovered fluorine is estimated from a density measurement in the liquid product of the absorption tower.

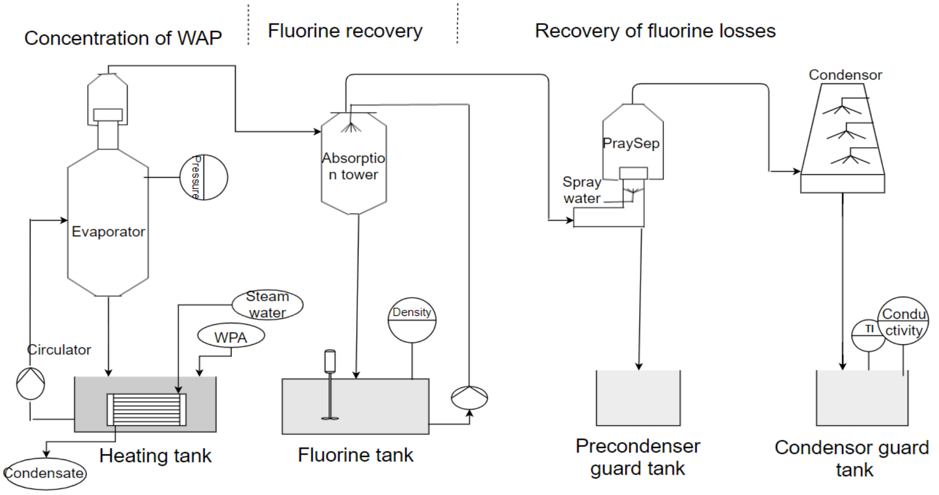

Figure 1: Diagram of the phosphoric acid concentration process.

3. Methods

The data used in this study contain all 2019 data, we focus on the most important variables. Initial data visualization was conducted in JMP® software (SAS Institute) (Yee et al., 2000). The most important signals, which are density and conductivity, are plotted over several cycles in Figure 2. As the fluorine recovery happens in cycles of varying duration (an average of 70 min), the trends of these variables in each cycle was of interest. During a typical cycle, density in the fluorine tank increases due to fluorine recovery, while at the process exit, conductivity in the condenser guard tank increases due to fluorine losses.

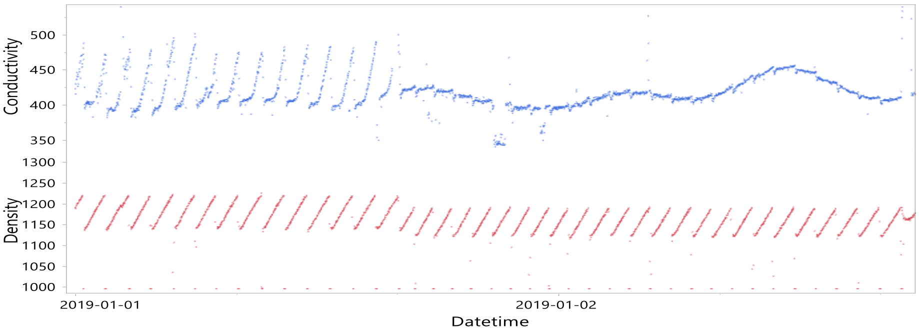

Figure 2: Evolution of the conductivity and the density in the fluorine tank as a function of time.

The figure 2 shows that the absolute values given by the conductivity sensor present a very high variability. Beyond short-term cycles, long-term variations also appear that may be due to incorrect calibration of the conductivity meter, or to other parameters which may affect the conductivity of the mixture in the condensate tank but that are not measured. For instance, the quality of the sprayed water before the PraySep, or even the water used to clean the vapour condensation could cause this long-term variation. To get rid of the long-term trend the conductivity signal has been normalised per cycle. For each conductivity point of a cycle, the value of the conductivity at the start of that cycle was subtracted. Consequently, a new standardised signal was constructed, making losses easier to analyse.

In order to study the behaviour of this conductivity in a more systematic fashion, a clustering method was applied. Gaussian Mixture Models (GMM) were used to automatically classify and label the conductivity cycles according to parameters qualifying their evolution. It was assumed that cycles with few losses, with medium losses and with many losses all follow a normal distribution. GMM allow us to identify these normal distributions (mean, variance). Note that, at other times, the conductivity signal is noisier or does not present a regular pattern. From these observations, we decided to fix the number of clusters to 3 in the initialisation parameters, plus a specific cluster for outliers. As the slopes of the conductivity characterise the fluorine losses well, we have used parameters derived from the slopes as inputs to this algorithm. Specifically, conductivity slopes have been calculated over 20-minute ranges at the start, the middle and at the end of each cycle, as well as the slope of the total conductivity cycle.

The way in which the density evolves during a cycle gives an idea of the productivity and the efficiency of the fluorine recovery. At the same time, the conductivity behaviour during the cycle reflects the Fluorine losses that are not recovered by the absorption tower and the PraySep. Therefore, to accurately assess the performance of the fluorine recovery, the density gain and the conductivity loss must be considered in combination. To solve this problem an indicator that measures this recovery has been constructed. It is a ratio between the average of the normalised conductivity and the average of the density gained during a given time. Note that the density signal measures the cumulative density of the fluorine tank mixture which increases during the fluorine recovery cycle, while the conductivity gives values considered as instantaneous since the condenser guard tank is quite small, consequently the residence time of the mixture in this tank is very low.

$$\mathrm{R} = \frac{\overline{\sigma}}{\overline{\rho_a} - \overline{\rho_i}} \tag{1}$$

Where: $\overline{\sigma}$: Average of normalized conductivity over a period Δt; $\overline{\rho_i}$: The density of the first point of this Δt; $\overline{\rho_a}$: Average of density over this Δt

This ratio measures how many conductivity points are generated for density points gained over a given period of time during the fluorine recovery cycle. Theoretically, this ratio increases as a function of time during the recovery cycle. At the beginning of the cycle, the fluorine tank is filled with fresh water, as a result its recovery capacity is high. Subsequently, the density of the mixture in the fluorine tank increases over time, consequently its capacity to recover fluorine decreases, which generates more fluorine losses. As a result, the values of both the numerator and the denominator increase with time. However, the losses (numerator) increase faster than the gains (denominator), so that this ratio can be used as an indicator of when the cycle starts being inefficient in terms of fluorine recovery.

4. Results and discussion

The GMM algorithm was applied with the conductivity slopes as features. The model groups the data into 4 families, the cluster 0 contains the cycles whose probability of belonging to the 3 other groups is negligible, therefore, they can be considered as outliers. As a result of this clustering, the cluster 1 contains 53% of the cycles which corresponds to stable signals with low slopes (low fluorine losses). The cluster 2 includes disturbed signals with very high slopes and large standard deviations between the ratios, this group has over 12% of the conductivity cycles. The cluster 3 includes 33% of the cycles, it gathers signals that do not include too much noise and with high slopes compared to those of the first cluster. In the rest of this study, we will focus on the 1st and the 3rd clusters to quantify the losses of each of them, and to analyse the distributions of the production parameters. The 2nd cluster has been discarded because the signals in this group are too noisy to be interpreted.

Following this, we studied the evolution of the ratio R during the fluorine recovery cycles. To do so, each cycle has been divided into 3 parts of 20 minutes. Then, the R value is calculated at the start, the middle, and the end of the 20-minute periods to produce the R_20_Start, R_20_Middle, and R_20_End values, respectively. The figure 3 shows the distribution of these ratios for the two clusters.

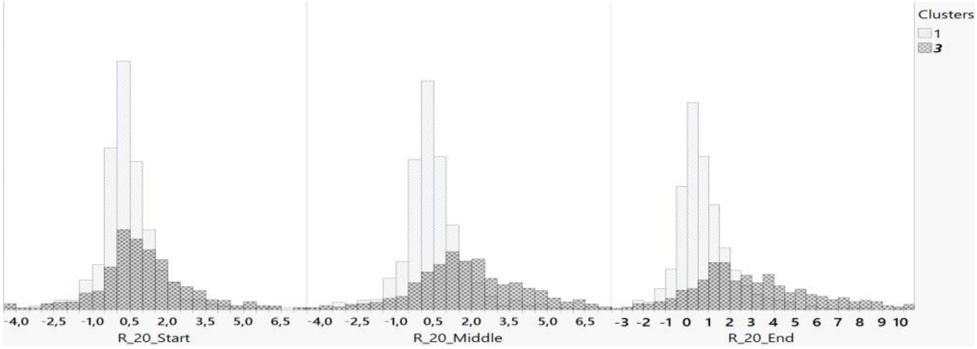

Figure 3: The histogram distribution of the fluorine loss ratios for the two clusters 1 and 3

The figure 3 illustrates that there is a progressive shift in the histograms of these ratios for the two clusters. In fact, the difference in R ratios between the two clusters starts at the beginning of the cycle, although this difference is relatively small in absolute value (0.5 on average for the first cluster and 0.94 for the third cluster). One can also observe that the standard deviation of cluster 3 is much larger than the one of cluster 1. Over time the difference becomes more important until the end of the cycle where the two clusters become clearly separated with a small overlapping section. At the end of the cycles, the ratio for the first cluster remains almost stable (1.08 on average), while the ratio for the third cluster considerably increases, from 0.94 at the start to 3.75 on average at the end. Also, we notice that 90% of the data of the first cluster have an R_20_End less than 3. On the other hand, for cluster 3, 50% of the data are higher than this value. So, the value R_20_End=3 can be used as a limit value which indicates high fluorine losses.

Once these observations have been made, relationships with process variables can be analysed. The important variables that can influence the fluorine recovery are: weak phosphoric acid (WPA) feed rate, circulator amperage, steam flow, vacuum, and flow rate of the spray water before the PraySep. The behaviours of these variables have been studied but they do not show a very clear difference depending on whether a data point belongs to one cluster or to the other. This could be explained by the continuous nature of the slopes, by the noise on the conductivity signal and by the overlapping nature of the clusters. Therefore, this prevents a clear distinction between the two clusters in their respective process operating conditions. Nevertheless, it has been remarked that the flow rate of spray water did not change between the two clusters, whereas there is a difference for the steam flow rate. It is possible that the spray water flow rate is not adapted to the productivity of the unit.

To retrieve more information about the influence of process variables, it was decided to work on a sample that encompasses the two extreme types of conductivity cycles. Cycles with intermediate conductivity slope were manually removed from the data set in order to strengthen the differences between the two clusters. The resulting sample is composed of 253 cycles, of which 35% belong to the 1st cluster and 65% to the 3rd cluster. The distribution of R ratios for this sample is similar to the results in Figure 3. We notice that in this case all R_20_End values for the first cluster are lower than 3 and more than 70% of the cluster 3 R_20_End values are higher than 3. This further confirms the value of R_20_End=3 as a threshold indicator of high fluorine losses.

As a second step, the process operating conditions have been studied for this sample, in the same way as it was done for the whole database. As results, the difference between the two clusters appears more clearly when doing the study on this sample. The distribution of the WPA flow, the steam flow, the vacuum pressure, and pump amperage show that cycles belonging to the 1st cluster correspond to low values of these process variables. In contrast, cycles belonging to the 3rd cluster have higher parameter values. This confirms the important effect of the process load on the conductivity behavior during a recovery cycle. Indeed, when the flow rate of phosphoric acid to be treated is increased (WPA flow), operating conditions are being pushed towards the process limits (for instance, more steam is required for the evaporation), but this seems to be insufficient to reach similar Fluor recovery performances. Nevertheless, all these variables present a large zone where the two clusters overlap, which confirms that the process load alone does not explain all the fluorine losses.

5. Conclusion

To conclude, this study was conducted to study fluorine recovery in a phosphoric acid concentration plant. In order to better characterize the losses during the recovery cycles, a ratio R has been defined which considers the gain (density increase in the recovery vessel) and the losses (conductivity increase in the liquid effluent waste). Since conductivity evolves in different ways, a clustering method using GMM algorithm has been used to classify the data according to low or high fluorine recovery performance. Following this methodology, it has been found that from a value of R=3 there are high fluorine losses. So, this ratio can be used as an indicator of fluorine losses.

Finally, it should be mentioned that the same procedure discussed previously was applied to another fluorine recovery unit located in the same phosphate plant. As a result, the same conclusions were reached. To go beyond this study, a monitoring system that triggers an alert for high fluorine losses when the ratio R is higher than a given value (in this case when R>3) will be studied. In addition, and to clearly identify the causes that influence the behaviour of fluorine losses, a detailed experiment on the flow rate and quality of the spray water injected to the PraySep unit will be carried out.

References

Diez-Olivan, A., Del Ser, J., Galar, D., Sierra, B., 2019. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. Inf. Fusion 50, 92–111.

Lee, I., Shin, Y.J., 2020. Machine learning for enterprises: Applications, algorithm selection, and challenges. Bus. Horiz., artificial intelligence and machine learning 63, 157–170.

Liu, G., Yang, J., Hao, Y., Zhang, Y., 2018. Big data-informed energy efficiency assessment of China industry sectors based on K-means clustering. J. Clean. Prod. 183, 304–314.

Sancho, A., Ribeiro, J.C., Reis, M.S., Martins, F.G., 2020. Cluster Analysis of Crude Oils based on Physicochemical Properties, in: Pierucci, S., Manenti, F., Bozzano, G.L., Manca, D. (Eds.), Computer Aided Chemical Engineering, 30 European Symposium on Computer Aided Process Engineering. Elsevier, pp. 541–546.

Zhang, Y., Bingham, C., Martínez-García, M., Cox, D., 2017. Detection of Emerging Faults on Industrial Gas Turbines Using Extended Gaussian Mixture Models. International Journal of Rotating Machinery.

Yee, Christine, et al., 2000. Design of experiment and data analysis by JMP®(SAS institute) in analytical method validation Journal of pharmaceutical and biomedical analysis,pp 581-589