Colucci, D.; Prats-Montalbán, JM.; Fisore, D.; Ferrer, A. (2019). Application of multivariate image analysis for on-line monitoring of a freeze-drying process for pharmaceutical products in vials. Chemometrics and Intelligent Laboratory Systems. 187:19-27. https://doi.org/10.1016/j.chemolab.2019.02.004

Submitted for publication on

Chemometrics and Intelligent Laboratory Systems

Application of Multivariate Image Analysis for on-line Monitoring of a Freeze-Drying Process for pharmaceutical products in vials

Abstract

A new Process Analytical Technology (PAT) has been developed and tested for on-line process monitoring of a vacuum freeze-drying process. The sensor uses an infrared camera to obtain thermal images of the ongoing process and multivariate image analysis (MIA) to extract the information. A reference model was built and different kind of anomalous events were simulated to test the capacity of the system to promptly identify them. Two different data structures and two different algorithms for the imputation of the missing information have been tested and compared. Results show that the MIA-based PAT system is able to efficiently detect on-line undesired events occurring during the vacuum freeze-drying process.

Keywords: *multivariate image analysis; process monitoring; infrared image; batch process.*

1. Introduction

Vacuum freeze drying (VFD) is a highly attractive process for the water removal in thermal sensitive products, mainly pharmaceutical ones, since water is removed at low temperature by sublimation. Monitoring of critical quality attributes of the product, e.g. the residual amount of ice and the product temperature, is required to guarantee a true quality-by-design manufacturing. To achieve this goal, the development of suitable Process Analytical Technologies (PAT), able to monitor/control the key variables of the process without interfering with the process dynamics, is a mandatory step, as stressed also by the Guidance for industry PAT by the American Food and Drug Administration [1].

In the past, many approaches to this problem, based on the measurement of different process variables (e.g. product temperature, sublimation rate, heat flux to the product, among the others), were proposed and tested [2], in particular at lab-scale. The measurement of the temperature of the product, possibly in a well-defined position (e.g. the bottom of the vial), was extensively and successfully applied for process monitoring and control [3]. The main drawback, up to this moment, of this approach is that the temperature measurement has to be performed using a thermocouple stuck into the product, and this does not guarantee neither the sterility requirements nor that the sensor is not interfering with the ongoing process.

In this work we used an infrared camera, instead of a thermocouple, for temperature measurement. Differently from the system proposed in the literature [4], we placed the camera inside the chamber, thus being able to monitor the vials in several positions, and not only on the top shelf of the freeze-dryer. Moreover, by this way it is possible to track vial temperature along several axial positions, and not only at the top. The spatial position of the vials inside the drying chamber has been proved to have a dramatic effect on the variability of

the product inside the single batch [5]; any monitoring algorithm, in order to be successful, has to account for this source of variability. In this work this aspect has been deeply investigated. Obviously, this system is able to monitor directly the temperature of the glass wall, and not that of the product, but several studies appeared in the literature evidenced that the temperature of the product is very close to that of the glass wall [6].

Thermal images include a lot of useless (i.e. everything that is outside the vial) information and, also the one directly related with the process, i.e. the temperature, is highly noisy, redundant and correlated. The first problem is a matter of gray-scale image segmentation, whereas the second is a frequent problem when dealing with real industrial data. Latent variables based multivariate statistical techniques can easily deal with these kinds of problems. In this framework, Kourti [7] discussed the primary role of multivariate statistical techniques in the development of PAT for the pharmaceutical industry.

Multivariate Image Analysis (MIA) is the application of multivariate statistics methods to the extraction of information from images, both spectral [8], that is directly related to the intensity of each pixel, and textural, i.e. linked to the spatial distribution of intensity gradient [9]. Prats-Montalbán et al. [10] published a complete review of MIA techniques and possible applications to problems of image segmentation, monitoring and defect detection, classification and prediction. Application to hyperspectral images was also discussed. Two different possible approaches to MIA were discussed: a global image approach and a pixel level one. The latter treats the spectral information in each pixel as a sample of the whole image, while, global image MIA is used when a set of features describing certain characteristics of the image are extracted and used for classification and/or prediction purposes. In this work, a global image approach was used.

The idea underlying the development of a latent variable multivariate monitoring

system is that only a few underlying events are driving the process, and all the measurements we obtain are just a different sight on this underlying driving force. Multivariate Statistical Process Control (MSPC) allows us to obtain a model of the process by projecting the information into a low dimensional space defined by latent variables and, in this reduced space, to build control charts able to detect any deviation from the normal operating conditions [11]. Both principal component analysis (PCA) [12] and projection to latent structures (PLS) [13] have been widely studied and applied for this purpose, being also able to successfully deal with the highly auto-correlated data typical of batch processes [14-16] such as VFD intrinsically is. Multivariate control charts for batch process monitoring were proposed by Nomikos and MacGregor [17], while Kourti [18] presented a more general discussion of MSPC of batch processes. Ramaker et al. [19] discussed the advantage of conjugating these techniques with process specific information in a so called *gray model*. Van Sprang et al. [20] presented a comparative evaluation of five different algorithms to the problem of on-line batch process MSPC. Rato et al. [21][22] recently presented a systematic methodology to compare batch process monitoring methods and compared different approaches in terms of detection strength and speed.

In more recent years many successful applications of MSPC to real industrial problems were published; a definitely not exhaustive list of them includes industrial polymer batch process [23,24], a continuous recovery process [25], batch production of PVC [26], fed-batch fermentation [27], autobody assembly [28] and continuous slurry stripping [29]. A complete discussion of multivariate image analysis in the process industries was published by Duchesne, Liu and MacGregor [30], including different examples of image analysis application to MSPC and real-time process control and optimization [31].

This paper is thus focused on the design of a new Process Analytical Technology

(PAT) for on-line process monitoring in a VFD process. The sensor uses an infrared camera to obtain thermal images and multivariate image analysis (MIA) to extract the information after automatic detection and segmentation of the region corresponding to the product in every vial. This information allows detecting on-line undesired events eventually occurring in the batch.

The paper is organized in six sections: section 2 introduces the experimental work and some useful nomenclature, section 3 deals with image preprocessing and feature extraction, section 4 presents the MSPC scheme and the different approaches to two of the major issues related to batch process MSPC (data unfolding and missing data estimation) tested, section 5 presents the main results, and section 6 the general conclusions of this work.

2. Experimental study and nomenclature

Drying experiments were carried out using a lab scale equipment LyoBeta 25™ freeze-dryer (Telstar, Spain). In all tests ten vials (ISO 8362-1 10R) were placed at 30 cm from the camera, and a new image was acquired every five minutes for 50h corresponding to almost 600 image acquisitions. The actual elapsed time from the beginning of the process was used to report the results of continuous variables “sampled” by the sensor, while the progressive number of the image acquisition was preferred to refer to the results of the calculation performed by the algorithm on the single images.

The sensor, together with the infrared camera (FLIR A35), includes a HDTV rgb (upper or lower case letters) camera. In this work, only the thermal images were used and will be discussed. The data are stored into a microprocessor that can be accessed via wi-fi and

monitored in real time through a graphical user interface. A case in plastic material was designed (IMC Services s.r.l., (upper or lower case letters) Italy) to resist and protect the electronics from the low pressure, low temperature and high moisture level typical of the drying chamber during the freeze-drying process [32].

The normal operating conditions (NOC) set was obtained processing 5 ml of a solution 10% b.w. of sucrose (Sigma Aldrich, 99.5%) at -20°C and 20 Pa. Five batches were processed in the same operating conditions, thus obtaining a total of 50 vials. Each vial in the batches was assigned with a number referring to the position on the shelf as Figure 1A shows, progressively increasing with the number of the batches. Batch 6 was intended to be another NOC batch but, due to the vibrations of the equipment, the vial in position number 7 felt down and was, for this reason, regarded as a fault, while the remaining nine vials were considered successfully dried.

The detection ability of the system was evaluated in four additional batches. In batch 7 a breakage in the vacuum system have been simulated and, after 5 hours of drying, chamber pressure was raised to 50 Pa. In batch 8 the shelf temperature was set to -10°C while in batch 10 a solution 5% b.w. of sucrose was used. Batch 9 aimed to prove the ability of the model to detect faults affecting the single vials and, while shelf temperature and chamber pressure were set to the NOC values, only four vials (corresponding to vials 81, 88, 89 and 90) were filled with a 10% solution. For the remaining six vials the configuration was the following: a piece of glass was inserted into two of them; another one was filled with pure water; one more with the same 5% b.w. solution used for batch 10, and the remaining two with, respectively, 2.5 and 7.5 ml of solution. Batches 11 and 12 are NOC batches included into the test set and used only to assess the monitoring performance of the algorithm. Given the high amount of time required to obtain a new batch these were assembled randomly by selecting

vials from the five batches that constitutes the training set (batches 1 to 5). The only restriction imposed during selection was to preserve the position of the vial on the shelf (e.g. vial 111, first vial in batch 12 was selected between the vials 1, 11, 21, 31 and 41), as it is one of the variables studied in this work. Should we stress that this "simulated" NOC test data set would give overoptimistic results? Table 1 resumes the number of the vials belonging to each batch, the operating conditions tested and whether it was used to train the model or to test it.

The ratio between the pressure measurements obtained from a conductive Pirani gauge and a capacitive Baratron manometer was measured on-line in every batch. This ratio is greater than one when the ice is sublimating, while approaches the unitary value at end of the primary drying [33]. For this reason it has been used to determine when the drying was completed and as a comparison for the features extracted (see section 3) from the thermal images.

3. Image segmentation and features acquisition

The thermal images are 256x320 pixels. The camera is equipped with a 63°x50° lens which leads to a slight optical distortion, known as barrel effect. This second order deviation from the ideal rectilinear projection can be compensated by remapping the pixels according to the following equation:

$$r_{new} = r_{old} + f \cdot r_{old}^2 \quad (1)$$

where r is the distance from the center of the image of a generical pixel and f a correction

factor (negative in this kind of optical aberration) depending on the distance between the camera and the object [34]. Since in all our tests the same distance was used, this factor is approximately constant and equal to -1.5.

After optical correction, the Hough transform [35] was used to detect the position of the vials in the images, as shown in Figure 1B. Being known the diameter of the vials bottom and the length of the line detected by the Hough transform we could infer the width of a single pixel and, thus, as we know also the height of the vial, the height of the region to be segmented, Figure 1C.

The whole portion of the image corresponding to the product into every vial was segmented and, to study the evolution over time of the temperature distribution, the values of mean, standard deviation (std), skewness and kurtosis of the temperature in this region were calculated from the measured values. The results were collected into a three-dimensional data structure using two approaches. In the first approach, in the following referred as vial-wise or *VW*, each vial has been considered as a single observation, thus $\underline{\mathbf{X}}_1$ is an $I \times J \times K$ data structure where I is the number of vials (fifty considering 10 vials in each one of the batches included in the training set), J is the number of variables measured (mean, std, skewness and kurtosis), and K is the number of time instants (six hundred). In the second approach, referred as batch-wise or *BW*, also the position on the shelf was included among the modeled variables, thus $\underline{\mathbf{X}}_2$ is a $B \times J^* \times K$ data structure, where B is the number of batches (i.e. five in the training set), K is again the number of time instants while the variables (J^*) are forty, corresponding to mean, std, skewness and kurtosis for each one of the ten positions that a vial could occupy on the shelf.

4. Batch process monitoring

In both approaches the data structures were unfolded putting all the features extracted for a single observation beside each other in order of time acquisition. This kind of unfolding preserves the information on the single observation, beside capturing the cross-correlation and the auto-correlation along time [16]. The final matrix obtained for the *BW* approach $\mathbf{X}_2(B\times J^*K)$ has ten times more columns and ten times less rows of $\mathbf{X}_1(I\times JK)$, the unfolded matrix obtained with the *VW* approach. After mean centering and scaling the unfolded data sets $\mathbf{X}_1$ and $\mathbf{X}_2$, a PCA model with, respectively, A_1 and A_2 principal components was built using only the batches of the training set, that is, the 50 successfully dried vials included in batches 1 to 5. The general structure of a PCA model is the following:

$$\mathbf{X} = \mathbf{T}\cdot\mathbf{P}' + \mathbf{E} \quad (2)$$

where **T** is the $I\times A_1$ (or $B\times A_2$) *score* matrix, **P** is the $A_1\times JK$ (or $A_2\times J^*K$) *loading* matrix and **E** is the residual matrix, having the same dimension of the original matrix **X**.

Only the batches included into the training set were used to build up the PCA model; batches 6 to 12 were used for validation purposes [36]. Once the latent variable subspace is known, unusual behaviors can be detected using two multivariate control charts buit on the following two statistics: Hotelling T^2 (T^2) and the squared prediction error (*SPE*), defined by Equations 3 and 4, respectively. For each observation:

$$T^2 = \sum_{a=1}^{A} \frac{t_a^2}{\lambda_a} \quad (3)$$

$$SPE = \sum_{c=1}^{KJ} e_c^2 \quad (4)$$

where t_a is the a-th score, λ_a its corresponding variance and e_c is the error obtained after predicting the measurement of variable c for a certain observation.

In on-line monitoring the *SPE* is computed only on the information measured at instant k, and for this reason it is called *instant SPE* (*SPEI*):

$$SPEI = \sum_{c=1+(k-1)J}^{JK} e_c^2 \quad (5)$$

where *JK* becomes *J*K* in the batch-wise approach. Nomikos and MacGregor [17] proposed to use the errors on a moving window of five instant measurements to compute the upper control limit (UCL) for this statistic. First guess UCLs for these charts were computed both empirically, that is taking the 99.5 % percentile of the actual values of both statistics obtained from the training set, and using their theoretical approximations, following the approach of Nomikos and MacGregor [17].

The percentage of time instants that a single statistic overtakes the UCL in NOC batches, is called Overall type I (OTI) risk, and should be close to the imposed significance level (ISL = 0.5%):

$$\mathrm{OTI} = 100 \times \frac{Nf}{I_{NOC} \times K} \% \quad (6)$$

where *Nf* is the number of time instants that a single statistic overtakes the UCL for the

overall training set [38]. Notice that the number of vials *I* should be replaced with *B*, the number of batches in the training set, in the *BW* approach.

Due to the limited number of batches available, a *one-batch-out cross-validation* approach was used. It consists in removing in turn each batch (in case of the *VW* approach, all the ten vials obtained in the same batch) from the training data set, build a PCA model using the remaining training batches (vials), measure the actual OTI for the deleted batch (vials), measure the average OTI after all iterations, modify, if needed, the UCLs by multiplying this by a coefficient greater or lower than 1, and repeat the procedure, until the desired OTI is achieved [37]. To evaluate the actual performance of the algorithm, the UCLs have been manually retuned, i.e. the original ones (both theoretical or empiric) were multiplied by a positive factor greater or lower than 1. To what extent this is already stated in the highlighted text in yellow??

The main issue when dealing with on-line batch multivariate SPC is that at time *k* ($k<K$) the future part of the trajectory of each variable *j* (or *j**) is missing and has to be "filled in" [17]. Arteaga and Ferrer showed that among the different scores estimation methods for future multivariate incomplete observations from an existing PCA model, the most statistical efficient ones are those that estimate the scores for the new incomplete observation as the prediction from a regression model: the so-called regression-based method. Out of these methods, two are recommended: the *Trimmed Score Regression* (*TSR*) method and the *Known Data Regression (KDR)* [39], [40]. These have been tested and compared in this work.

Given a reference matrix of observations $\mathbf{X}$ and its PCA decomposition $\mathbf{X} = \mathbf{T}\cdot\mathbf{P'} + \mathbf{E}$, when a new partially unknown observation $\mathbf{z}$ is available at time *k*, it can be written as $\mathbf{z}^{\mathbf{T}} = [\,\mathbf{z}^{*\mathbf{T}}\ \mathbf{z}^{\#\mathbf{T}}]$, where $\mathbf{z}^{*}$ includes the first *Jk* (or *J***k*) known values of $\mathbf{z}$ and $\mathbf{z}^{\#}$ contains all the

values still unknown. This partition induces a partition also into $\mathbf{P^T} = [\mathbf{P^{*T}}\ \mathbf{P^{\#T}}]$ and $\mathbf{X} = [\mathbf{X^*}\ \mathbf{X^\#}]$, see Figure 2. Ferrer and Arteaga proved that $\mathbf{z^\#}$ can be estimated by the general formula:

$$\hat{z}^{\#} = \mathbf{S}^{\#*} \times \mathbf{L} \times \left(\mathbf{L}^T \times \mathbf{S}^{**} \times \mathbf{L} \right)^{-1} \times \mathbf{L}^T \times z^* \tag{7}$$

where $\boldsymbol{L}$ is key matrix, different for each method of imputation used,,equal to the identity matrix $\mathbf{I}$ (K-k×K-k) for the *KDR* method, and to the partial loading matrix $\mathbf{P^*}$ (A×K-k) when the *TSR* method is used. Given the full covariance matrix $\boldsymbol{S}$ obtained from the know data set $\mathbf{X}$, $\boldsymbol{S}^{\#*}$ and $\boldsymbol{S}^{**}$ are the partition induced by the separation in **z [**39**]**:

$$\mathbf{S} = \frac{\mathbf{X}^T\mathbf{X}}{n-1} = \frac{1}{n-1}\begin{pmatrix} \mathbf{X}^{*T}\mathbf{X}^* & \mathbf{X}^{*T}\mathbf{X}^{\#} \\ \mathbf{X}^{\#T}\mathbf{X}^* & \mathbf{X}^{\#T}\mathbf{X}^{\#} \end{pmatrix} = \frac{1}{n-1}\begin{pmatrix} \mathbf{S}^{**} & \mathbf{S}^{*\#} \\ \mathbf{S}^{\#*} & \mathbf{S}^{\#\#} \end{pmatrix} \tag{8}$$

The two regression methods were compared in terms of accuracy of the score estimation, accuracy of the prediction of the future observation and, indeed, the fault detection ability following the procedure used by Garcia-Muñoz et al. [41]. Three fundamental properties of a good predicted score matrix were checked: *orthogonality*, *coherence* and *stability.* The score predicted for each principal component must be orthogonal, thus the covariance matrix should be diagonal with the terms on the diagonal, in order to be coherent, arranged in a decreasing order. *Stability* means that the estimation of the score must be constant in time and equal to the true value, i.e. the one obtained at the end of the process when all the variables are known, also at the beginning of the batch when most of the matrix is missing. The future prediction sum of squares (*FPRESS)* and the future prediction mean square error (*FPMSE*), Equations 10 and 11, introduced by Garcia-Muñoz et al. [41] were used as a measurement of

the quality of the forecast of the unknow part of the trajectory of variable *j*, in observation *i* made at time instant *k*:

$$FPRESS_k^{ij} = \sum_{l=1}^{K-k} \left(e_k^{ij}\big|_l \right)^2 \quad (9)$$

$$FPMSE_k^{ij} = \frac{\sum_{l=1}^{K-k} \frac{1}{l} \left(e_k^{ij}\big|_l \right)^2}{\sum_{l=1}^{K-k} \frac{1}{l}} \quad (10)$$

Quizá habría que indicar que FPRESS se calcula sumando los FPRESS^j para cada variable *j*.

where $e_k^{ij}\big|_l$ is the error made for each observation *i* at time instant *k* when forecasting the future part of the unknown trajectory of variable *j*, that is, the values corresponding to the data to be acquired from *k+1* to *K* or, equivalently, for *l* from 1 to *K-k*. *FPRESS* is the equivalent of a *SPE* calculated on the predicted part of the observation and represents a measure of global forecast accuracy. In the *FPMSE* the error at each instant of time is weighted by the inverse of the distance to the current time sample giving back a measure of the local forecast accuracy at specific time instant *k*.

Finally, after the PCA model was created and the control limits tuned, the monitoring performances of both T^2 and *SPEI* control charts were compared by projecting the observations of the test set onto the reference model. At every time step, only the information known up to that time instant was used, the missing part of the observation was forecasted, with either the *KDR* or the *TSR* algorithm, thus simulating an on-line acquisition system. The occurrence of false positives and false negatives were investigated, together with the amount of time needed to perform the calculation on the whole data set of images.

5. Results

Figure 3 shows the Pirani-Baratron pressure ratio trajectory (Figure 3*a*) compared with the 10 trajectories, one for each vial, described by the four variables: mean, std, skewness and kurtosis of the temperature of the pixels corresponding to the product into the vials (Figure 3*b*, 3*c*, 3*d* and 3*e*, respectively) measured during the drying of one of the reference batches. Average temperature shows a change of slope around 9 hours after the onset of the primary drying stage, and an asymptotic behavior up to the thermal equilibrium. The standard deviation (std), after a sudden decrease, grows up until a maximum reached at almost 9 hours; then, it slowly decreases again until reaching an almost constant value at 36 hours. Both skewness and kurtosis show a maximum, followed by a local minimum around 9 hours. An almost constant value is kept from 36 hours to the end. The local maxima (or minima, as well as the change of slope in the mean temperature) seems to correspond to the first slope change of the Pirani/Baratron pressure ratio. The constant values at the end indicates that the thermal equilibrium has been reached, i.e. there is no more sublimation, thus the primary drying is over. Significant differences in the thermal trajectories obtained in different tests may reveal an abnormal heat transfer, that is an anomalous drying kinetic and a lower product quality. The features extracted from the thermal images, although based on simple first order statistics, contain some relevant information about the process required to perform a process monitoring. Multivariate statistical techniques are, nevertheless, mandatory to deal with such amount of noisy and redundant data.

In the *VW* approach , the reference model was created extracting 10 PC corresponding to 93.6% of variance explained. In the *BW* approach, two or more principal components, give back basically the same results. Thus, the reference model was built extracting two principal

components. Regarding the UCL for T^2, there is a remarkable difference between the theoretical and the empirical values, being the former always lower than the latter. The UCL for *SPEI* computed with the theoretical distribution and the one obtained taking the percentile of 99.5% are always very similar. After tuning the control limits, the obtained OTIs for the *VW* approach with *TSR* and *KDR* were, respectively, 0.47% and 0.48% for *SPEI*, and 0.48% and 0% for T^2; in the case of a *BW* data approach we obtained 0.47% and 0.47% for *SPEI*, and 0.43% and 0% for T^2. It was impossible to achieve a higher OTI in the case of T^2 simulated with *KDR* because the scores are straight and as the limits were relaxed the error soon exceeded the ISL.

5.1 Missing information algorithms comparison

Figure 4 shows the evolution along time of the prediction of the score of the first principal component for both *TSR* and *KDR* algorithms and both data unfolding approaches for the training data set: 50 vials for the *VW* approach corresponding to 5 batches for the *BW* approach. In both cases the scores predicted with the *KDR* algorithm are more stable during the process and, except for the first few images of the *VW* approach, they are perfectly constant and equal to the true values (i.e. the values obtained at the end of the batch). The score predicted with the *TSR* algorithm asymptotically moves toward the true values, but without completely reaching them. In both cases the coherence of the score covariance matrix was respected, that is the variance explained by the scores of the first component is greater than that of the second component and so on, but the full orthogonality of the scores was obtained only applying the *KDR*. A diagonal matrix was obtained from the first instant of time in case of a *BW* approach and after the first 15 images in the *VW* approach.

After the scores were computed, an estimation of the original data matrix **X** could be

obtained by multiplying the score matrix and the transpose loading matrix; subtraction of this estimation from the original data matrix gives back the forecast error of the future (unknown) part of the trajectory of all the variables in observation *i* made at each time instant *k*. From the final *K-k* columns, the values of *FPRESS* and *FPMSE*, shown in Figure 5 and 6, were obtained. As expected, being a global estimation of the prediction error, *FPRESS* always decreases with time, while FPMSE, that accounts for the local forecast accuracy, could increase, decrease or, as in this case, keep an almost constant (except for generalized increment towards the end) value. To ease the comparison among the different algorithms, a darker line with symbols, representing the mean value of the trajectories, has been reported. In Figure 5, corresponding to *VW* approach, the mean *FPRESS* shows a maximum of 26 at *k* = 3 for *TSR*, while the maximum for *KDR* is 135 and is located at the fifteenth data acquisition. This five-time difference is kept through all the process and the mean value for *TSR* also goes faster to zero. A slighter difference can be noticed also in the *FPMSE*, with the *TSR* always behaving moderately better. Same conclusions can be stated from the analysis of Figure 6, corresponding to *BW* approach. In this case FPRESS is one order of magnitude greater than in the *VW* approach while *FPMSE* is basically constant.

KDR seems to give back a more stationary prediction of the scores, while the *TSR* algorithm better forecasts the original value of the observations. The larger the number of columns included in the unfolded data set, the higher the prediction errors are.

5.2 Classification performance - VW approach

Once a PCA model of the process has been fitted using the observations included in the NOC batches and the UCL for both *SPEI* and T^2 have been tuned in order to have an OTI closed to the imposed *ISL* of 0.5%, we can evaluate the ability of this model to discriminate a fault

377 from a successful drying. The classification performance of this monitoring system has been
378 evaluated by projecting the whole test set on the obtained model. At each time step, only the
379 measurements available up to that instant were used while missing part of the observations
380 was forecasted with the algorithms just discussed, thus simulating a real on-line monitoring
381 system. Tuned empirical limits performed slightly better than the theoretical ones and have
382 been used. This could be due to the fact that taking the percentile of the actual distribution of
383 the statistics helps to better follow the instantaneous variation of the distribution itself and
384 better describe the little misbehavior that could occur.
385 The control chart for *SPEI* is almost the same in both cases while the T^2 control charts are
386 quite different especially at the beginning. Using the *trimmed score regression* algorithm,
387 while tuning UCLs, the control charts for *SPEI* detected 8 false positives vials in the training
388 set (8, 9, 15, 16, 19, 31, 40 and 50). Comparable results have been obtained with the *known*
389 *data regression* method; *SPEI* detected 7 false positives (vials 1, 6, 9, 15, 41, 50, 54) in the
390 training data set. Looking at the vials that appeared as false positives in *SPEI* (they are 1, 6, 8,
391 9, 15, 16, 40, 41, 50 and 54) we can notice a certain periodicity in the results. Position 1 and
392 10 in every batch corresponds to the external vials, directly radiated by the chamber walls. In
393 the first two batches a thermocouple was located inside the vials in position 5 and 6, see
394 Figure 1. This slight difference into the data structure of vials 6, 15, 16 might be due to the
395 influence of the thermocouple on the drying kinetics.
396 These observations mildly overtake the control limits on a limited number of time instants.
397 If we accept these spurious errors as part of the unavoidable statistical error rate, that is, we
398 assume that the phenomena responsible for these instantaneous faults cannot jeopardize the
399 quality of the resulting product, the fault detection performance could be further optimized.
400 This new relaxation of the control limits was achieved by considering faults only the vials

401 whose *SPEI* crossed the control limits in more than 5% of the time instants. In this way all
402 the false positives in *SPEI* where properly classified as successful tests.
403 The validation of the obtained model was performed by projecting all the vials of the test
404 set onto the obtained model step by step simulating a real-time acquisition. Figure 8
405 compares the control charts obtained with the different algorithms for three vials:
406 - number 105, a NOC vial, always below the UCL;
407 - number 51, a vial expected to be in control but reported as a fault in the *SPEI* control
408 charts;
409 - number 75, dried at a higher shelf temperature and lies over the control limits in all cases
410 almost all the time.
411 Vial 57, as well as all the vials of the anomalous batches 7, 8 and 10 were detected as
412 faults. In batch 9 six vials were tampered and all of them have been correctly
413 discriminated. The T^2 control chart reported nine false negatives and two false positives
414 (vials 52 and 56). Only one of the four vials of batch 9 dried with the original 10% sucrose
415 solution (vial 81) has been correctly found to be a successful drying test. Vials 88 and 89
416 and 90 have been highlighted as faults by the T^2 control chart. This results are from TSR?
417 Using the *KDR* algorithm four false positives where highlighted into batch 9 together
418 with vials number 51, 52, 56, 58, 60. The T^2 control charts detected vials 51 to 60 (the whole
419 batch number 6) and 81, 88 and 90 as false positives but no false negatives.
420
421 The anomalous behavior of vials 51, 60, 81 and 90 appears to prove what has been
422 stated about the effect of the radiation from the surrounding. The appearance of vials 56
423 could be due to either the effect of the thermocouple used also in this batch, or to a greater
424 amount of radiation from the surrounding due to the absence of vial 57 after it felt down. A

425 possible explanation of the anomalous behavior of some vials in position 8 and 9 the
426 presence of the cold led, required for the illumination of the rgb camera field of view, right in
427 front of them.

428 In general, we could state that the *trimmed score regression* algorithm gives back a
429 better *SPEI* control chart, probably because of the lower error in the estimation of the
430 variables, while the *known data regression* algorithm gave back better T^2 control charts,
431 which could be a direct consequence of the better score estimate we discussed. The *KDR*
432 requires 13 times the computational time required for a simulation using the *TSR*.

433

434 *5.3 Classification performance - BW approach*

435 Since we segmented and extracted features from each one of the different vials in the images,
436 the most natural way to treat the data was the vial-wise approach. The idea to organize the
437 data including the potential effect of the vial position on the shelf was conceived when we
438 noticed the periodicity in the results just discussed in the previous subsection. There is also a
439 matter of variables and matrix dimension, i.e. in the *BW* approach we have only 5 batches,
440 but for each image we obtain forty new columns and the resulting matrix is ten times wider
441 than the *VW* one. The computational time required was 28 times greater when *TSR* is used
442 and lasted almost 10 days (instead of a few minutes) with the *KDR*. In this last case (use of
443 the *KDR* algorithm for a matrix organized taking the single batch as an observation) the time
444 required for the analysis of a single image is greater than the 5 minutes required for each data
445 acquisition, thus jeopardizing the possibility of a real time application of the algorithm. After
446 matching the desired OTI, both *TSR* and *KDR* presented three false positives (batches 3, 4
447 and 5) in the SPEI control chart; the three batches of the training set that had no
448 thermocouples inside the central vials. Again, a 5% threshold was set to correctly

449 discriminate the false positives in the *SPEI*, although just raising the control limits would
450 have fit the purpose.
451 As for the *VW* approach, we tested the behavior of the monitoring scheme by
452 projecting all the batches of the test set on the plane defined by the two principal components.
453 Also in this case tuned empirical limits performed better and were preferred.
454 Figure 8 shows the control charts of both *SPEI* and T^2 for three of the seven batches
455 that constitute the test set. Batch 12 was always perfectly classified by both algorithms, batch
456 8 was always clearly signaled as an outlier, and batch 10 was detected only by the T^2 control
457 charts.
458 Neither false positive nor false negative were detected with both the *TSR* and the *KDR*
459 by the *SPEI* control charts. No false positive where detected in the T^2 control charts when
460 using either the *TSR* or the *KDR* algorithm. In both cases only batches 6 and 8 were detected
461 as faults. Even after the limits in the T^2 control chart were retuned, it was impossible to set
462 apart the faulted batch 7, 9 and 10 from the NOC data set.
463 These three batches have a common characteristic: they all simulated faults mainly
464 concerning the mass transfer. In the VFD process mass and heat transfer are intimately
465 coupled that is, any deviation in the mass transfer affects also the evolution of the
466 temperature profiles in the products, but it will always be an indirect and weaker effect.
467 Being a weaker effect, also the breakage into the correlation structure of the data will be less
468 pronounced. Moreover, the glass of the vials is almost opaque to infrared radiation
469 (emissivity 0.9) thus, the temperature we measure is that of the vial wall. The two values
470 were proved to be quite similar, but it might partially mask some slight variation in the
471 thermal history of the product. The effects on the heat transfer are strong enough to be read
472 by the *SPEI* chart, but not to raise an alarm into the T^2 chart. On the other side, in batch 6 a

vial felt down, and the camera measured the temperature of the front door of the dryer, which is warmer than a normal vial and almost constant during the whole process. This single vial misbehavior was strong enough to completely compromise the data structure. In batch 8 the whole shelf was set at a higher temperature. Thus, also the glass of the vial, which is always in contact with the shelf, is directly heated and both the average temperature as well as the whole temperature distribution change. This direct effect is strong enough to be detected also by the less responsive of the control charts.

This lower detection ability of the T^2 control charts in batch processes monitoring has already been reported in the literature and is basically due to the strong auto-correlation in the data [38].

As the position of the vial on the batch is, in this approach, part of the model, any harmful effect of the position should be highlighted into the contribution plot. In Figure 9 we reported the contribution plots of batch 6 (a), 8 (b) and 11 (c), respectively, for *SPEI* when *TSR* is used, after 6.7 hours of drying. Batch 11 is the reference, a good batch correctly discriminated (note the small value of the contributions). In case of batch 6 the bars corresponding to the variables in position 7 (variables 25 to 28) are three orders of magnitude greater than the others as well as those of batch 11 (I can´t see it!), denoting that something in that vial is going wrong. In batch 8 all the forty variables are higher than expected, betokening an unconventional processing.

6. Conclusions

In this work a PAT for MIA based real time monitoring of vacuum freeze-drying has been

497 developed and tested. The sensor uses an infrared camera to get thermal images of the
498 ongoing process. These images are segmented, after optical aberration correction, and global
499 features of these regions are extracted and used to detected unusual behavior in the new
500 observations.

501 Two different approaches of data unfolding and missing data estimation have been
502 tested and compared, with the aim to obtain the best combination for addressing the problem
503 at hand. The *TSR* algorithm appears to better forecast the missing values and this gives back a
504 slightly more responsive *SPEI* control chart. On the other side, the *KDR* algorithm better
505 estimates the scores of the future observation that ensures better, although non-optimal,
506 performances of the T^2 control chart. In any case the nature of batch data makes the T^2
507 statistics not reliable. The main drawback of the *known data regression* algorithm is the time
508 required for the data analysis, basically due to the need to invert the partial covariance matrix
509 of the training data set ($\boldsymbol{S}^{**}$), whose dimensions increase in time and is normally very ill-
510 conditioned. For matrix with a great number of columns, the long time required jeopardizes
511 the possibility to apply this algorithm on-line.

512 Both modeling approaches guarantee a fine fault detection. In a *BW* approach,
513 detection of a fault related to a single vial is deputed to the analysis of the contribution plots,
514 and thus less immediate. Using the single vials as an observation, the algorithm is more prone
515 to type II (false negative) errors since the effects related to the spatial position of the vial on
516 the shelf are not included in the model and could mask a deviation of the same amplitude in
517 the control charts; although we did not face this kind of problem. Given the lower number of
518 columns, the computational time is dramatically lower in this second case.

519 This PAT could be used to assess whether the variation of freeze-drying process is
520 only due to common causes, that is the process is in statistical control, or some special cause

might affect the product quality. Since this information is available on-line it might strongly reduce the failure rate of the process, the waste production, the laboratory tests to be performed at the end of the batch, and the time required from the end of the process to the release of the batch. Anyway, before considering any industrial application, the algorithm developed should be, indeed, validated on larger industrial data sets.

The performance of this algorithm could be further improved including in the data set other variables available and currently measured during the process, especially those directly related with the mass transfer inside the chamber (e.g. chamber pressure, vapor flow, etc.). Future works will aim to prove the possibility to apply other multivariate techniques and the infrared imaging technology for process optimization and control.

References

[1] U. S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Veterinary, Medicine (CVM), Office of Regulatory Affairs (ORA), Pharmaceutical CGMPs. September 2004. *Guidance for Industry, PAT A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*. https://www.fda.gov/downloads/drugs/guidances/ucm070305.pdf (last access date: June 2018).

[2] D. Fissore, R. Pisano, A.A. Barresi, Process analytical technology for monitoring pharmaceuticals freeze-drying – A comprehensive review, Drying Technology, In press. doi: 10.1080/07373937.2018.1440590.

[3] D. Fissore, R. Pisano, A.A. Barresi, On the use of temperature measurement to monitor a freeze-drying process for pharmaceuticals, Proceedings of I2MTC – 2017. IEEE International Instrumentation and Measurement Technology Conference, Torino, Italy, May 22-25, 2017, pp. 1276-1281.

[4] H. Emteborg, R. Zeleny, J. Charoud-Got, G. Martos, J. Luddeke, H. Schellin, K. Teipel, Infrared thermography for monitoring of freeze-drying processes: instrumental developments and preliminary results, Journal of Pharmaceutical Sciences 103 (2014) 2088-2097.

[5] A.A. Barresi, R. Pisano, V. Rasetto, D. Fissore, D.L. Marchisio, Model-based monitoring and control of industrial freeze-drying processes: effect of batch nonuniformity, Drying Technology 28 (2010) 577-590.

[6] S.A. Velardi, A.A. Barresi, Development of simplified models for the freeze-drying

process and investigation of the optimal operating conditions, Chemical Engineering Research and Design 86 (2008) 9-22.

[7] T. Kourti, Process analytical technology beyond real-time analyzers: the role of multivariate analysis, Critical Reviews in Analytical Chemistry 36 (2006) 257–278.

[8] P. Geladi, H. Grahn, 1996. *Multivariate Image Analysis*. Ed. John Wiley & Sons Ltd. Chichester, England., 1996.

[9] M.H. Bharati, J.F. MacGregor, Texture analysis of images using Principal Component Analysis, Proceeding of SPIE/Photonics Conference on Process Imaging for Automatic Control, Boston, 2000, pp. 27-37.

[10] J.M. Prats-Montalbán, A. De Juan, A. Ferrer, Multivariate image analysis: a review with applications, Chemometrics and Intelligent Laboratory Systems 107 (2011) 1-23.

[11] T. Kourti, J.M. MacGregor, Multivariate SPC methods for process and product monitoring, Journal of Quality Technology 28 (1996) 409-428.

[12] P. Nomikos, J.M. MacGregor, Monitoring batch processes using multiway principal component analysis, AIChE Journal 40 (1994) 1361-1375.

[13] P. Nomikos, J.M. MacGregor, Multi-way partial least squares in monitoring of batch processes, Chemometrics and Intelligent Laboratory Systems 30 (1995) 97-108.

[14] P. Nomikos, *Statistical process control of batch processes*. PhD Thesis, McMaster University, Hamilton, Ontario, 1995.

[15] P. Nomikos, Detection and diagnosis of abnormal batch operations based on multi-way principal component analysis, ISA Transactions 35 (1996) 259-266.

[16] T. Kourti, Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions, Journal of Chemometrics

17 (2003) 93-109.

[17] P. Nomikos, J.M. MacGregor, Multivariate SPC charts for monitoring batch processes, Technometrics 37 (1995) 41-59.

[18] T. Kourti, Application of latent variable methods to process control and multivariate statistical process control in industry, International Journal of Adaptive Control and Signal Processing 19 (2005) 213-246.

[19] H.J. Ramaker, E.N.M. Van Sprang, S.P. Gurden, J.A. Westerhuis, A.K. Smilde, Improved monitoring of batch processes by incorporating external information, Journal of Process Control 12 (2002) 569-576.

[20] E.N.M Van Sprang, H.J. Ramaker, J.A. Westerhuis, S.P. Gurden A.K. Smilde, Critical evaluation of approaches for on-line batch process monitoring, Chemical Engineering Science 57 (2002) 3979-3991.

[21] T.J. Rato, R. Rendall, V. Gomes, S.-T. Chin, L.H. Chiang, P. Saraiva, M.S. Reis, A Systematic Methodology for Comparing Batch Process Monitoring Methods: Part I – Assessing detection strength, Industrial & Engineering Chemistry Research, 55 (2016) 5342-5358.

[22] T.J. Rato, R. Rendall, V. Gomes, P. Saraiva, M.S. Reis, A Systematic Methodology for Comparing Batch Process Monitoring Methods: Part II—Assessing Detection Speed, Industrial & Engineering Chemistry Research, 57 (2018) 5338-5350.

[23] K.A. Kosanovich, M.J. Piovoso, K.S. Dahl, J.F. MacGregor, P. Nomikos, Multi-way PCA applied to an industrial batch process, Proceedings of American control conference, Baltimore, Maryland, June 1994, pp. 1294-1298.

[24] C.E. Schlags, M.P. Populе, Industrial application of SPC to batch polymerization processes, Proceedings of American Control Conference, Arlington, VA. June 25-

27, 2001, pp. 337-343.

[25] T. Kourti, J. Lee, J.F. Macgregor, Experiences with industrial applications of projection methods for multivariate statistical process control, Computers and Chemical Engineering 20 (1996) S745-S750.

[26] A.A. Tates, D.J. Louwerse, A.K. Smilde, G.L. Koot, H. Berndt, Monitoring a PVC batch process with multivariate statistical process control charts, Industrial Engineering and Chemistry Research 38 (1999) 4769-4776.

[27] B. Lennox, G.A. Montague, H.G. Hiden, G. Kornfeld, P.R. Goulding, Process monitoring of an industrial fed-batch fermentation, Biotechnology and Bioengineering 74 (2001) 125-135.

[28] A. Ferrer, Multivariate statistical process control based on principal component analysis (MSPC-PCA): some reflections and a case study in an autobody assembly process, Quality Engineering 19 (2007) 311-325.

[29] T. Darkow, R. Dittmar, H. Timm, Real-time application of multivariate statistical methods for early event detection in an industrial slurry stripper, Proceedings of the 19th World Congress of The International Federation of Automatic Control, Cape Town, South Africa, August 24-29, 2014, pp. 8879-8884.

[30] C. Duchesne, J.J. Liu, J.F. MacGregor, Multivariate image analysis in the process industries: A review, Chemometrics and Intelligent Laboratory Systems 117 (2012) 116-128.

[31] M.H. Bharati, J.F. MacGregor, Multivariate image analysis for real-time process monitoring and control, Industrial Engineering and Chemistry Research 37 (1998) 4715-4724.

[32] E. Lietta, D. Colucci, G. Distefano, D. Fissore, On the use of infrared thermography

for monitoring a vial freeze-drying process, Journal of Pharmaceutical Sciences 108 (2018) 391-398.

[33] S.M. Patel, T. Doen, M.J. Pikal, Determination of the end point of primary drying in freeze-drying process control, AAPS PharmSciTech 11(1) (2010) 73-84.

[34] J.P. De Villiers, F.W. Leuschner, R. Geldenhuys, Centi-pixel accurate real-time inverse distortion correction, Proceedings of SPIE (2008) 7266:11-1-11-8.

[35] R.C. Gonzalez, R.E. Woods, S.L. Eddins, *Digital image processing using Matlab*. Pearson education Inc. Upper Saddle River, New Jersey, 2004.

[36] J.M. Prats-Montalbán, A. Ferrer, Statistical process control based on Multivariate Image Analysis: a new proposal for monitoring and defect detection, Computers and Chemical Engineering 71 (2014) 501-511.

[37] S. Wold, Cross-validatory estimation of the number of components in factors and principal components models, Technometrics 20 (1978) 397-405.

[38] J. Camacho, J. Picó, A. Ferrer, Multi-phase analysis framework for handling batch process data, Journal of Chemometrics 22 (2008) 632-643.

[39] F. Arteaga, A. Ferrer, Framework for regression-based missing data imputation methods in on-line MSPC, Journal of Chemometrics 19 (2005) 439-447.

[40] F. Arteaga, A. Ferrer, Dealing with missing data in MSPC: several methods, different interpretations, some examples, Journal of Chemometrics 16 (2002) 408-418.

[41] S. Garcia-Muñoz, T. Kourti, J.F. MacGregor, Model predictive monitoring for batch processes, Industrial and Engineering Chemistry Research 43 (2004) 5929-5941.

651 **List of Tables**

652

655

List of Figures

Figure 1. Image preprocessing and segmentation example: (A) Original image as obtained from the thermal camera, (B) Image after optical aberration correction; the white line at the vials bottom is the line detected by the Hough transform. (C) Result of the pre-treatment and segmentation process, the features are extracted from the pixels corresponding to each one of the black rectangles.

Figure 2. Notation used in the missing data prediction algorithms.

Figure 3. Pirani Baratron ratio profile (graphs A) and temperature features trajectories (graphs B to E) extracted for batch 2.

Figure 4. Comparison of the score of the first principal component obtained by TSR (graphs A and C) and KDR (graphs B and D) method on both vial-wise (graphs A and B) and batch-wise (graphs C and D) unfolding approach.

Figure 5. Comparison of FPRESS (graphs A and B) and FPMSE (graphs C and D) obtained with KDR (graphs B and D) and TSR (graphs A and C) methods for the vial-wise approach. Gray lines: value for each one of the single vials, Black line with symbols: average trajectory.

Figure 6. Comparison of FPRESS (graphs A and B) and FPMSE (graphs C and D)

obtained with KDR (graphs B and D) and TSR (graphs A and C) methods for the batch-wise approach. Gray lines: value for each one of the single batches, Black line with symbols : average trajectory.

Figure 7. Control charts for vial 105, batch 11 (dark gray line), 51, batch 6 (signaling only in SPEI, light gray line), and 75, batch 8 (signaling in both SPEI and T^2, light gray). Thick black line: tuned empirical UCLs.

Figure 8. Control charts for 3 batches of the test set when using the *BW* approach. Gray lines: batch 12 (NOC); dark gray lines: batch 8 (test set, clearly highlighted as an outlier in all cases); light gray lines: batch 10 (test set, signaled only by the SPEI control chart). Thick black line: tuned empirical UCLs.

Figure 9. Contribution plot obtained for batch 6 (A), 8 (B), and 11 (C) simulated in the batch-wise configuration using the TSR algorithm.