Food Engineering Reviews (2023) 15:24–40
https://doi.org/10.1007/s12393-022-09324-0

QbD/PAT—State of the Art of Multivariate Methodologies in Food and Food-Related Biotech Industries

Christian H. Pérez-Beltrán[1] · **Ana M. Jiménez-Carvelo**[1] · **Anabel Torrente-López**[1,2] · **Natalia A. Navas**[1,2] · **Luis Cuadros-Rodríguez**[1,2]

Received: 26 May 2022 / Accepted: 7 October 2022 / Published online: 24 October 2022

Abstract
Several investigations have been made at lab scale considering the quality by design (QbD) and process analytical technology (PAT) approaches. Nonetheless, such applications have been focused on the analyzers or multivariate tools used at small scales. This comprehensive review presents the state of the art of both QbD and PAT. In addition, the key historical events since 1940 which have influenced the development of the QbD/PAT system are also highlighted. Moreover, the application of the recommended PAT tool of *multivariate tools for design*, *data acquisition*, *and analysis* (design of experiments, multivariate data analysis, and multivariate process control) is revised for the food and food-related biotechnology industries and describes the applications reported over the last 20 years. On this subject, only 34 studies were found in literature whose relation was close with both industries at industrial or pilot plant scales; a description of each of them focusing on multivariate tools is presented. Finally, some conclusions and future perspectives on this topic are given, with the aim of initiating a change in the field.

Keywords Quality by design · Process analytical technology · Design of experiments · Multivariate data analysis and process control · Food and food-related biotech industries

Introduction

During the initial steps in the design of the infrastructure for industrial quality, it was assessed once the product was finished, and, paradoxically, this goal is still common in many cases today. According to Koch [1], it was around the 1940s in Germany, after the World War II, when the concept of *process analytics* (PA) or *process analytical chemistry* (PAC) were proposed in the framework of the chemical and petrochemical industries. The distinctive features of this new approach were explained in detail in an outstanding and pioneering tutorial published by Callis et al. [2]. In such industries, PAC was implemented as the chemical or physical analyses of materials carried out during the elaboration process to understand the composition of molecules of interest, which was popularized for the following twenty years and adopted by refineries and nuclear plants.

Whilst PAC was being implemented, the quality trilogy was stated for the first time by Juran [3], work in which a new direction for managing quality was proposed through quality planning, control, and improvement. Later on, such processes gave rise to the *quality by design* (QbD) concept in 1992 by Juran [4], where the importance of planning the quality sought after by customers and consumers was outlined and explained through the use of the quality trilogy. After the development of the QbD Juran's approach, the Food and Drug Administration of the United States of America (US FDA) introduced in 2004 the concept of *process analytical technology* (PAT) [5] for its use in the pharmaceutical industry due to the loss of credibility of the industry before this time. The introduction of PAT was based on PAC, modifying the term *chemical* to *technology* and including more features, like microbiological, mathematical, and risk analysis, to enhance understanding and to control the production process. According to the above FDA guidance, PAT is considered *a system for designing*,

✉ Christian H. Pérez-Beltrán
christianpb@correo.ugr.es

1 Department of Analytical Chemistry, Faculty of Science, University of Granada, c/ Fuentenueva s/n, E-18071 Granada, Spain

2 Biohealth Research Institute (ibs.GRANADA), University of Granada, Granada, Spain

analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes, *with the goal of ensuring final product quality*. The same definition is also included in the ASTM E2363-14 [6] standard.

The early PAT implementation was based mainly on analytical chemistry and was focused on the development and application of at-line/on-line analytical instruments, particularly process analyzers, as evidenced in Koch et al. [7] and Bakeev [8]. Thus, the PAT concept was more linked to the use of analyzers than with the implementation of their paradigms. This fact caused confusion for some time, to the extreme of erroneously naming many analytical instruments as PAT analyzers or PAT tools. This is why a description of such instruments or their applications is not included in the paper as it is out of the scope of this review.

Few years after the acceptance of the PAT approach by the FDA, the International Conference on Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) produced the ICH Q8(R2) guideline [9], which retrieves the term QbD and adopted PAT definition and concept for the pharmaceutical industry. The ICH Q8(R2) defines QbD as *a systematic approach to development that begins with predefined objectives and emphasizes on product and process understanding and process control*, *based on sound science and quality risk management*. One important aspect to consider when using QbD is the design space, defined as *multidimensional combination and interaction of input variables and process parameters that have been demonstrated* to provide assurance of quality [9, 10]. The process should constantly work under the designed space, which allows for continuous quality improvement,but when deviations occur out of it, there is considered to be a change, and possible sources of no conformities should be valued [9]. This usually launches a post-approval requirement change process, which is aimed at establishing a new design space.

The aforementioned key historical events are summarized in Fig. 1 in which each action is represented by a black horizontal line in the rising arrow, representing a timeline and at the same time the way knowledge in quality has been increasing through the coming years represented by the shape of the arrow. A similar complementary historical review of PAT during its beginnings was performed by Chew and Sharratt [11].

In this regard, QbD and PAT share similar targets, being product and process understanding and process control. Nonetheless, QbD is considered the framework in which PAT is applied to achieve the goals previously mentioned. The QbD approach starts with the definition of the quality target product profile (QTPP). Thus, PAT is currently considered as an enabler whose analytical instruments and strategies make possible the creation of easier and smarter control plans, identification of critical process parameters (CPP) and detection of important interactions between process parameters, and relevant critical quality attributes (CQA) of the material to obtain the desired quality by designing it through the process with real-time or near real-time measurements. These measuring strategies are characterized by being performed during the process using analytical devices capable of non-destructive analysis, which is one of the PAT elements.

Besides the implementation of analytical devices, the PAT system encourages the inclusion of other elements [5], such as (i) multivariate tools for design, data acquisition, and analysis,(ii) process control tools; and (iii) continuous improvement and knowledge management tools. A conceptual diagram displaying the QbD framework and the different PAT tools used to achieve the ultimate goal can be observed in Fig. 2.

In this sense, the analytical instruments and process analyzers for analytical control used in the PAT system can provide different measurements, which are divided in on-line, in-line, at-line, and off-line analyses according to the way in which they are performed [5, 6]. On-line measurements consist on those determinations where the sample is deviated from the manufacturing process line and then returned to the process stream. For the in-line measurements, the sample is not removed from the manufacturing line, and it can be invasive or non-invasive. These types of measurements are performed by different analyzers or sensors, which can be implemented individually or in conjunction during the same process, producing different types of data that need to be treated and fused. This data fusion (DF) procedure is the current challenge and the next step in the evolution of PAT that could provide a more comprehensive understanding of the system and the opportunity to predict complex quality attributes, as described by Casian et al. [12]. In fact, the authors sate that most of DF applications can be found in the food sector, where it was used for food quality authentication, food and beverage characterization, and food quality assessment.

As it may be inferred, the aforementioned measurements produce enormous volumes of data, which need to be treated with different multivariate tools to mine the relevant and non-evident information. The main multivariate tools to be applied are design of multivariate experiments (DOE), multivariate data analysis (MVDA), and multivariate process control (MVPC), which are briefly described below.

The first described tool is DOE, which according to the ASTM [6], it makes reference to *the arrangement in which an experimental program is to be conducted, and the selection of the levels (versions) of one or more factors or factor combinations to be included in the experiment.* The quality of a product is linked to its production process, which must be all the time under control and the defined design

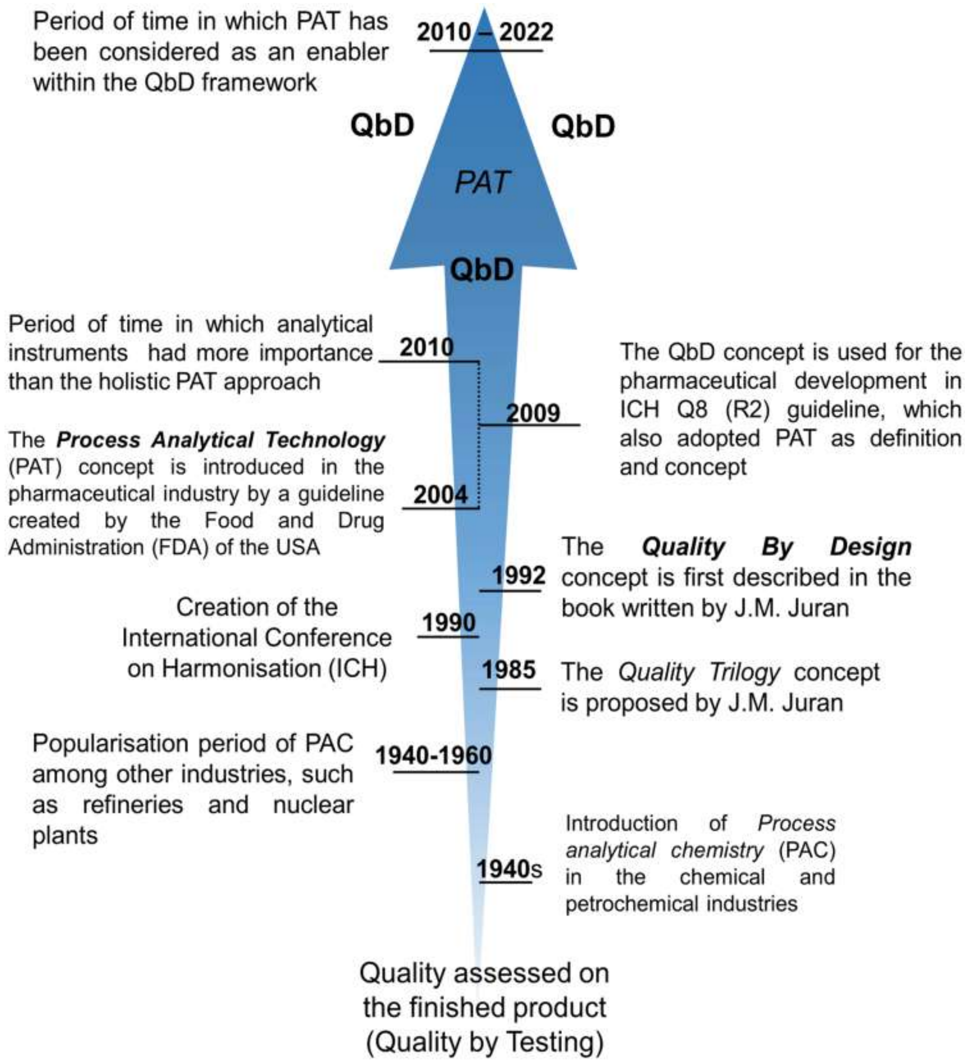

Fig. 1 Creation and evolution of the quality by design (QbD) and process analytical technology (PAT) concepts

space. When the pre-defined design space is modified, new requirement approvals using characterization studies must be performed over the new process to confirm that it can operate and perform properly yielding the desired product quality. Thus, DOE promotes innovation, problem solving, and discovery, and it is used to screen a process or response, create a space design, optimize a response(s), and study the interrelationships among multiple factors of interest [13, 14] A similar approach for the process control, optimization, and creation of future process trajectory is known as model predictive control (MPC). However, this model is not based on multivariate experimental designs, but in the reduction of the squared error of the process trajectory over the time horizon, minimizing the short-term effects of unknown and erratic signals, manipulating multiple input variables [15]. Nonetheless, only DOE is reviewed in this work, since it is the considered tool within the PAT system. The reader is kindly referred to literature [15] to find out more about MPC.

The second methodology, MVDA, is described by ASTM as *an appropriate tool for exploring and handling large sets of heterogenous data, mapping data of high dimensionality onto lower dimensional representations, exposing significant correlations among multivariate variables within a single data set or significant correlations among multivariate variables across data sets* [16]. MVDA is a key enabler for process understanding and decision making, and for the release of intermediate and final products after being appropriately validated using a science and risk-based approach. It involves the use of multivariate data regression and dimension reduction that promote the development of different unsupervised and supervised multivariate models to explore/screen the data, to analyze the similarity of data set, or to predict (classify/quantify) qualitative features or physicochemical characteristics of products from the data.

The third multivariate tool is MVPC, which is described by the ISO 7870–7 [17] as *a process monitoring of problems in which many related variables are of interest using*

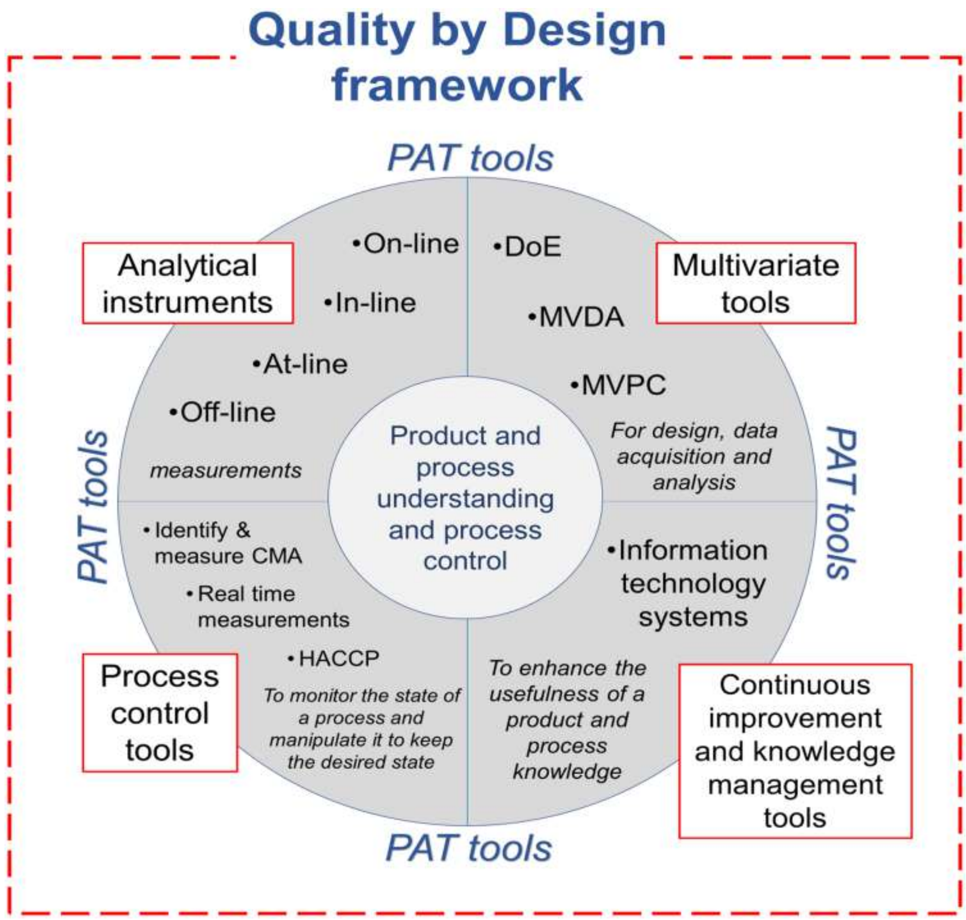

Fig. 2 Conceptual diagram of the quality by design (QbD) framework in which different process analytical technology (PAT) elements and tools are considered to achieve the main central objective of understanding the process and product and the control of the process. DOE, design of multivariate experiments; MVDA, multivariate data analysis; MVPC, multivariate process control; CMA, critical material attributes; HCCP, hazard analysis critical control points

multivariate control charts as the main mechanism to statistically evaluate and control the process. Based on this statistical evaluation and control, deviations can be identified through statistical signals, which make feasible the elimination of the possible cause of variation, bringing back the process to the state of statistical control to predict the performance and assess the capability to meet the future specifications.

From the different tools employed in the PAT system, special attention is paid to multivariate tools for design, data acquisition, and analysis. On this subject, this paper focuses on the multivariate tools for QbD/PAT applied both at industrial plant and pilot plant scales. Specifically, the application of DOE, MVDA, and MVPC in the food industry is described and reviewed in the "QbD/PAT Tools in the Food Industry" section, as well as in the food-related biotech industry in the "QbD/PAT Tools in the Food-Related Biotech Industry" section, considering the studies developed within a time span of 20 years. The pharmaceutical industry is not considered here since the QbD and PAT principles were concisely developed for it, and thus, a huge number of reports, applications, and discussions on the subject can be found in literature. Finally, the "Conclusions and Final Remarks" section intends to make evident the lack of QbD/PAT principles in the food and food-related biotech industries at industrial scale, encouraging a closer interaction among academic researchers and industry. This section also contains some ending remarks about the current review and future perspectives for both industrial sectors.

Note that the studies described in this review were selected considering, mainly (1) the scale at which they were performed—industrial or pilot plant scale—but some lab-scale studies were also considered because they involved the directly cooperation with the industry or company, but at once, (2) they had to explicitly mention that take into account QbD as a framework whose principles and goals must be achieved with PAT tools. Such studies were sought in the Web of Science and Google Scholar research tools, using "QbD," "PAT," "food," "biotechnology," "industry," "industrial-scale," and/or "pilot plant" as principal search words.

QbD/PAT Tools in the Food Industry

The QbD/PAT bases and tools are still rarely implemented in the food industry. This may be because each company has its preferred data acquisition methods and processes, from which may assess the quality of products off-line using laboratory-based analytical methods, many of them destructive in nature, time-consuming, costly in terms of complex sample preparation, and require highly skilled operators, making the implementation of a PAT system difficult, as pointed out by Teixeira et al. [18] and Hitzmann et al. [19]. Although QbD/PAT tools have been demonstrated to have a huge impact for improving the process understanding and control, and the end-quality of a product as evidenced in different research reports and review papers [20–23], they have been applied to a limited extent at industrial scale in the food sector.

The next subsections expose those cases in which PAT multivariate tools are applied in the food and food-related biotechnology industries, considering both industrial plant and pilot plant scales.

Design of Multivariate Experiments (DOE)

DOE is characterized by determining the relationships between input factors that promotes variation over the output responses. Only two studies with similar industrial-scale operating conditions were found in literature, which are presented in this subsection.

Fissore et al. [24] investigated the design space of a freeze-drying process of coffee extract. The study was focused on the primary drying of the process, since it accounts for most of the energy consumption. The design space was constituted by a set of operating conditions named heating fluid temperature (T_{fluid}) and chamber pressure (P_c). This research group found that, if only drying time is considered for a cycle with constant values of T_{fluid} and P_c, the values that should be used, according to the design space, are high values of shelf temperature and low values of chamber pressure (−5 °C and 5 Pa).

The second most similar study, in which DOE was applied at industrial scale, was recently performed by Tessarini et al. [25]. The authors developed and optimized the beer containing malted and non-malted substitutes under a QbD framework. For this study, a simplex-centroid mixture design was employed in which different proportions of maize, oat, malt, rice, rye malt, and sorghum cereals were tested. The starting QTPP of the Ale beer was 45% of barley malt substitutes as maximum, with $5.5 \pm 0.55\%$ v/v of alcohol content. CQA were color and pH, since they are directly related to the sensory part of the QTPP. Afterwards, a screening design took place where maize, oat, and rice were selected as barley malt substitutes for the optimization design. The optimization design consisted on performing a new mixture design with the three barely malt substitutes used to manufacture new formulations. The results reported that the following parameters should be considered to create an optimized formulation similar to the Ale beer: (i) employ high proportions of barley and oat malt (55 and 42%, respectively) and 3% of maize in the manufacturing process, (ii) use malt substitute for a greater level of color on the "European Brewery Convention" (EBC) scale or non-malted to lower it, and (iii) add higher concentrations of non-malted substitutes to reach optimal sensory scores.

The usefulness of DOE at the industrial scale in the food industry is demonstrated with the two previously discussed studies. In the same regard, it is also evident the lack of studies at industrial scale applying DOE, which shows that this PAT tool is still novel in the development of food products or that industries do not require it, since they already have well-established quality characteristics of their products. Nonetheless, DOE could be used to enhance the quality features of existing products and to reduce costs.

Multivariate Data Analysis (MVDA)

The potential of MVDA lies in its ability to extract meaningful information from complex data in a fast and easy manner. In fact, more studies applying MVDA than DOE were found in literature. Different techniques and foodstuffs are discussed in the following paragraphs.

The first study described in this subsection was performed by Xing et al. [26], who investigated the potential of visible spectroscopy for the classification and the color attribute prediction using as CQA the coordinates based on the CIE $L^*A^*B^*$ color space (L^*, lightness of the color of the sample; A^*, red and green characteristics; and B^*, yellow and blue characteristics). The experiments were conducted with a total of 189 samples at room temperature between 20 and 22 °C.

A stepwise discriminant analysis (STEPDISC) is performed to select a subset of the quantitative variables for use in discriminating among the classes. Canonical discriminant analysis (CDA) method was used to classify the samples into different groups. The partial least squares (PLS) regression was used to predict color attributes based on the reflectance spectra. Preprocessing methods such as Norris 1st derivative, Savitzky–Golay (SG) 2nd derivative, mean normalization (MN), and range normalization were tried. The authors found a significant difference among pale and red meat classes according to their mean reflectance spectra and to the Duncan test. Thus, they decided to classify the meat samples into pale and red classes. For the development of the first CDA model, four wavelengths (420, 440, 580, and 620 nm) were used to differentiate among the two classes of meat, obtaining a classification accuracy of 85%. Additionally, a discrimination model to differentiate between pale, firm, and nonexudative (PFN) meat and pale, soft, and exudative (PSE) meat classes was created with five wavelengths (420, 580, 440, 550, and 600 nm). The accuracy of the model for the PFN class was 82% whilst for the PSE class was 84%. Finally, a PLS model was developed using the full wavelength of the reflectance spectra to predict the color of the meat classes. Indeed, the model was capable to predict better the L^* coordinate than A^* and B^* ones, using as preprocessing technique the SG 2nd derivative. Nonetheless, the authors commented that further testing was needed to assess the effectiveness of the prediction model.

In the same subject of meat, Sørensen et al. [27] reported a method using spatially resolved near-infrared (NIR) spectroscopy to determine the fat quality of porcine carcasses by estimating the iodine value (IV). The study included 35 carcasses from a slaughterhouse, which were from a daily production stock and belonged to three different categories based on feeding regimes to ensure variability in the composition of the fatty acids. The NIR spectra were obtained 5 cm from the belly split line and 5 cm from the neck separation cut. Reference analyses were performed using gas chromatography of 287 disks of fat and skin collected close to the spot of spectroscopic measurements. Principal component analysis (PCA) was used to assess the GC results and screen whether the feeding groups could be distinguished based on their fatty acid composition. PLS was applied and used to predict the measured IV from the recorded spectra after mean centering and extended multiplicative signal correction (EMSC) and interval partial least squares (iPLS). The authors found that, combining EMSC and iPLS, the model yielded good results for IV predictions. Finally, the model was applied to predict the IV as a function of fat layer depth. The authors were able to measure the fat quality and fat layer quality differences of porcine meat at full abattoir processing speed with an on-line NIR transmission spectroscopy method refined with PCA, preprocessing techniques, and PLS.

Moreover, Achata et al. [28] also studied pork meat, but hyperspectral imaging (HSI) was used in this occasion. The authors used HSI with chemometrics to develop classification models of brined and non-brined pork loins and to predict the salt concentration used to brine the samples. For this study, 144 fresh pork loins (FPL), which were obtained from 16 animals, were brined using salt solutions concentrated at 5, 10, and 15% (w/v). The hyperspectral images were obtained in the Vis–NIR and NIR ranges at 20 °C from both sides of the samples. The authors used PCA, PLS-DA, and PLS in combination with several pre-treatments on both reflectance and logarithm-transformed data. In this sense, researchers could observe that samples grouped according to the experimental treatment they suffered, moreover, the authors obtained a classification model capable to correctly discriminate all brined and non-brined samples using the 957–1664 nm spectral range and quite good prediction models to estimate the salt concentration of raw and cooked FPL using the 957–1664 nm spectral range.

Furthermore, Pullanagari et al. [29] performed on-line quantifications of fatty acids (FA) of lamb meat under commercial abattoir conditions by means of visible-near-infrared spectroscopy (Vis–NIR), since they influence in the determination of the meat quality. The experiment consisted of the Vis–NIR analyses of 500 lambs from two commercial lamb processing plants. A PLS-based genetic algorithm (PLS-GA) was applied and is capable to compute and find optimization solutions useful for informative wavelength selection. Half of the 500 samples were randomly included in the calibration of the GA-PLS model,the rest were employed to validate it. The authors found that the majority of the studied FA and monounsaturated FA (MUFA) could be predicted with moderate accuracy. They concluded that, although the exact quantification of FA was not absolutely reliable with such procedure, it is evident that it could be adequate and implemented for screening purposes as part of a quality control process in the food industry.

Moreover, Lintvedt et al. [30] performed in-line measurements using Raman spectroscopy and PLS regression to estimate the concentration of eicosapentaenoic + docosahexaenoic acids in 63 salmon samples from three farming locations and the residual bone concentration (% ash) in 66 samples of mechanically recovered ground chicken, provided by a processing plant. The samples were measured with similar operating conditions to those used in the industry. Each sample was placed on plate covered with aluminum foil and passed through a dark cabinet using a conveyor belt at speeds 0.30, 0.15, 0.075, and 0.03 m/s, corresponding to exposure times of 1, 2, 4, and 10 s, respectively. The authors were able to obtain a model with good and acceptable results of prediction,they found that appropriate data preprocessing is fundamental to reduce the noise of the spectra, allowing them to acquire good results with models built using exposure times of 1 or 10 s.

Ørnholt-Johansson et al. [31] interpreted the data from companies to evaluate if MVDA could increase their production yield. For this research, 60 Atlantic salmons (Salmo solar) from three slaughterhouses in Norway were used. The authors found a meaningful weight loss through the processing of the fishes. They detected that mechanical filleting provided the initial weight difference. However, through the study of the data, the authors concluded that the main cause for the weight loss was the cut that divided the fillets from the skeletal frame. From these observations, three different PLS prediction models were built to estimate the yield after mechanical filleting. After comparing the three models, the authors decided to keep model two, which was built with variables shape ratio: W/LT, weight divided by length and thickness, K factor (W/L^3), weight divided by the cubed length, length, thickness, and weight. This model allowed the authors to differentiate salmons according to the slaughterhouse, using score and correlation loading plots, and allowed them to find that salmons from companies 2 and 3 were more similar between them than salmon from company 1. Finally, PCA was applied to study 13 deviating samples from the PLS model, discovering that two groups of salmon could be found according to the cut on their belly (straight or angled belly cut). The findings of this research ended up with valuable information that can help production companies in the decision-making process.

MVDA has also been particularly applied in the dairy sector. Lyngaard et al. [32] performed a real-time modeling of milk coagulation during the coagulation of twelve cheese batches to attain meaningful insights on this process through the application of near-infrared (NIR) reflection measurements, formulation, and testing of models. Twelve milk coagulation experiments were performed using 5 L of reconstituted milk, which were transferred to a 6 L of cheese vat imitating an industrial setting with normal operation conditions (NOC). Two PCA models were created and assessed according to the NIR measurements. The first model monitored the entire coagulation process with an s-shaped profile, involving enzymatic proteolysis of k-casein, paracasein aggregation, and gelation. The second model focused on each coagulation phase to obtain a more robust model for real-time use. The first model presented an excellent performance, but it was not appropriate for industrial screen of on-line parameters due to large variations in some estimation, caused by small alterations in the NIR measurements. The second one also performed well, and it led to acceptable parameter fittings, which make it a good option for a real-time application.

Following up with the dairy industry, Rimpiläinen et al. [33] tested a data-driven approach in an industrial-scale powder plant to predict and evaluate the end-point properties of milk powder through one function property known as sediment. A standard offline laboratory sediment test was performed to four consecutive production processes with 339, 300, 273, and 284 samples. Thirteen plant variables were examined to make possible the evaluation and implementation of a real-time quality control process. A prediction model was built based on conditional probability distributions (CPDs) using the first 75 sediment samples as training data set and then used to predict only the next sediment value, updating the model after every new sediment result. The prediction results showed that production processes 1, 3, and 4 were consistent with the established Gaussian assumption, but for process 2, the sediment results did not fit very well. It was observed that process pairs 1–2 and 3–4 were similar, indicating a possible change in the operating conditions of the procedure. According to the obtained results, the authors claim that sediment values could be controlled and decreased in each process. For production process 1, dryer temperatures (T_{D1} and T_{D2} should be increased; in process 2, direct contact heater temperature (T_{DC}) should be also increased; in process 3, steam injector temperature (T_{SI}) should be decreased and T_{DC} increased; and for process 4, T_{SI} should be decreased. However, when all processes were considered together, T_{SI} and milk concentrate temperature (T_C) had the highest influence on sediments. The authors found that the average prediction error levels of the CPD model were comparable with the PLS model. Nonetheless, they decided to keep the CPD model since it offered them a good manner to evaluate the influence of each predictor variable over the sediments. In this sense, it was verified that MVDA can be used to develop prediction models to find suggestions on how to adjust plant variables to improve the sediment values.

MVDA has also been applied in the olive oil industry, where Tamborrino et al. [34] studied the effect of calcium carbonate during the extraction process over the olive oil quality, energy consumption, and rheological properties to improve the extraction process adjusting malaxation parameters. The olive pastes and olive oil were obtained with a continuous olive oil mill plant. The energy consumption was assessed with measurements of the active, reactive, and apparent power. The electricity consumption was calculated regarding the rates of operation. Additionally, the viscosity of the pastes was measured to obtain their rheological properties. Once the chemical results were obtained, they applied PCA to make evident the main variables that affected oil samples. Researchers decided to use the effect plot on the volatile compounds and trans-2-hexanal to show the experimental trials trend.

Furthermore, Picouet et al. [35] used the QbD approach to predict the final acrylamide content of deep-fried potatoes "chips," which is a parameter related with safety. The final chips were defined according to twelve quality target parameters (QTPs), including 3 color coordinates (CIE $L^*A^*B^*$), 5 sensory attributes (odor roast, flavor rancid, flavor roast, crunchy, and oil mouth feel), 3 concentrations of volatile compounds (hexanal, pentylfuran, and 2.4diecadenal), and total acrylamide content. In this case, the utilized MVDA tool was a classical least-squares multilinear regression (MLR) coupled to a step-wise model, which was created with 65 frying experiments for the calibration set and 33 for the validation set, using a mid-level fusion approach of four different QTPs parameters (color coordinate A^*, "flavor roast" sensory descriptor, acrylamide, and pentylfuran contents). The results suggested that the predictive models for the acrylamide content were unsatisfactory, since it is still not clear some complex mechanisms and factors that influence the quality parameters of the potato chips.

Two recent applications of MVDA in the food industry were performed using hyperspectral imaging (HSI). Liu et al. [36] evaluated the potential use of HSI to detect sucrose adulteration in tomato paste, to compare the detection and prediction performance of different chemometrics methods, and to identify the lowest proportion of sucrose in tomato paste that can be safely detected. The study consisted of the multispectral analysis at 19 wavelengths of two batches of pure concentrated tomato paste provided by the industry. The tomato pastes and sucrose mixtures were made at 1–9% proportion levels (w/w). The authors were able to observe two clear groupings of batch 1 and batch 2 of tomato pastes using PCA. Furthermore, the authors quantified the level of adulteration in tomato paste using calibration models generated by PLS, least squares-support vector machine (LS-SVM), and back propagation neural network (BPNN). Researchers found that LS-SVM provided the best predictive results for both batches of tomato pastes. In this study, 100% accuracy was obtained in the prediction set with a detection limit of sucrose of 1% using HSI with multivariate methods.

Additionally, HSI was applied over milk powders by Munir et al. [37] to determine if it could be used as a process analyzer for the real-time quality control, coupled to a predictive regression model. The whole milk powder samples were obtained from three different factories with the same specifications, but with some equipment differences. The authors studied the effects of preprocessing the signal either by smoothing and/or differentiating it, and they applied PCA and PLS as multivariate analysis methods to find possible trends among the powders and to construct a model capable to predict the origin of an unknown powder, respectively. In this sense, the authors demonstrated through PCA that a high degree of smoothing is a suitable preprocessing step capable to maximize the differentiation performance among the three factories and between "poor" and "good" quality milk powders. Moreover, the constructed PLS model yielded 79–87% of accuracy regarding the dispersibility predictions related with "good" and "poor" powders. With the current study, the authors developed an analytical method using hyperspectral imaging and MVDA capable to distinguish among milk powders from different factories with diverse qualities and properties, which could be implemented online by the industry.

Continuing with the application of MVDA in the food industry, Moscetti et al. [38] developed, optimized, and predicted the desalting process by electrodialysis of soy sauce through the application of on-line NIR spectroscopy, level and conductivity probes, and a control strategy of the electric current generator. The raw soy sauce was analyzed in a laboratory-scale electrodialyzer plant in which monitoring and controlling activities were performed. The experiments were performed at four different electric current profiles, where only the first one was employed as the calibration set and the other three as prediction sets. The concentration of salt, non-salt solids, and amino nitrogen were determined with prediction models based on NIR spectroscopy to completely constitute the desalting process. The PLS model was performed with the 1100–1925-nm spectral range, over which standard normal variate (SNV), SG, and mean centering (MC) preprocessing steeps were applied, obtaining good predictions for the salt and non-salt concentration of the desalting process of soy sauce, as well as for the conductivity, osmotic pressure, and density.

Another study was performed by Lan et al. [39], who compared the ability of NIR, mid-infrared (MIR), and Raman spectroscopies and HSI to assess the composition and texture characteristics of apple purees produced in-house that mimic an industrial process. In this study, 62 samples from two different processes were analyzed and further used to build predictive and discriminative models, which were PLS, SVM, and random forest (RF). Several classification models were developed to discriminate among five characteristics of apple puree,prediction models were built to foresee eight rheological and structural properties and nine biochemical properties of the apple purees. Researchers found that the MIR technique coupled with RF and SVM had a higher discrimination accuracy of purees than the PLS-discriminant analysis (PLS-DA) and that NIR coupled with PLS resulted in better predictions of the quality puree parameters than the SVM and RF quantitative models.

MVDA has also been applied for the production of beer. Tessarini et al. [40] proposed a real-time monitoring process of beer parameters applying infrared spectroscopy with MVDA to predict the final quality of beer. The beer formulations were manufactured in a pilot plant where samples were collected from mashing, fermentation, and maturation processes for further physicochemical analysis using attenuated total reflection-FTIR (ATR-FTIR) spectroscopy. In this case, the PLS regression model was used to predict the final alcohol content, density, pH, and color, obtaining acceptable results. Furthermore, the predicted values for each manufacturing stage were contrasted with their corresponding experimental value using analysis of variance (ANOVA), which indicated that no statistical difference was observed among them. As a matter of fact, the authors could predict the desired quality of the finished beer through the manufacturing process, making possible the identification of deviations in the system, taking preventive or corrective actions if necessary.

Moreover, Schorn-García et al. [41] developed a PAT-based methodology to monitor and control a wine alcoholic fermentation process using spectroscopy and MVDA. Five alcoholic fermentations of *Saccharomyces cerevisiae* were performed and analyzed on-line and at-line to obtain the reference values. The authors built a PCA model to observe the trend of the density evolution during the fermentation process, which was easily followed by the plot elaborated with the first PC against time. Moreover, a PLS model was performed to predict the density (g/mL) along the alcoholic fermentation and to predict the biological time of the process, obtaining high and good correlation values between the predicted and the measured values. Furthermore, the authors developed a PLS-DA model to determine if the samples were under or out of control, which was built using five normal alcoholic fermentations and five contaminated fermentations. Results of this model allowed the authors to properly classify the samples in the corresponding classes with satisfactory values of sensitivity and specificity.

Another study performed by Wei et al. [42], consisted on achieving on-line measurements in continuous acquisition mode with an optical fiber probe system of 2300 tobacco leaves, was collected in three different years. The spectral data obtained from the NIR analyses were used to develop PLS, SVM, and convolutional neural network (CNN) quantitative models to predict changes in moisture, starch, protein, and soluble sugars of the samples during a flue-curing process, which lasted 7–8 days. In this regard, the authors were able to demonstrate that, for this particular case, CNN model performed better in the monitoring process than PLS and SVM. Moreover, the authors also created a strategy to include seasonal and temperature variability into the model to predict samples from a new harvest season in a curing barn, providing a potential and practical method to overcome performance degradation by seasonal differences and temperature oscillations.

Finally, Upadhyay et al. [43] studied ready-to-cook (RTC) food products, specifically, instant noodles. The authors performed in-line and at-line measurements of NIR and visible spectra at a pilot plant noodle manufacturing line of Nestle R&D Centre India Private Limited (Haryana, India). The spectra obtained from the in-line measurements were used to monitor some quality parameters during the process, such as moisture, crude protein, total fat, and total ash, whilst the at-line measurements perform the prediction of their content in the final product. The authors took advantage of some MVDA tools, such as PCA to study sample distribution patterns and detect probable outliers and PLS and SVM to perform the prediction/calibration models, together with different preprocessing techniques and competitive adaptive reweighted sampling (CARS) selection algorithm to improve the results of the models. According to the results presented in this work, the authors were able to obtain excellent prediction models with SVM under full wavelength for all the quality parameters, except for total ash, demonstrating that the quality monitoring of instant noodles produced under pilot plant facility is effectively achievable using NIR on the manufacturing lines.

As it could be observed in this subsection, MVDA analysis has been more applied than DOE in the food industry, which indicates its importance to monitor processes and gain knowledge about them. Nonetheless, the number of studies is still low, showing a good opportunity of improvement for industries.

Multivariate Process Control (MVPC)

Multivariate process control (MVPC) increased its popularity within the statistical process control, since the applied techniques reduce the amount of information contained in the variables of the process down to two or three metrics through the application of statistical modeling, according to Bersimis et al. [44]. Despite of its well-known benefits, only three applications were found in literature in which MVPC has been applied under very similar operating conditions to the food industry.

The first study was reported by Tokatli et al. [45] in which the critical control points (CCP) of a continuous food pasteurization process were monitored with MVPC, and fault detection and diagnosis methods were developed. The study was performed in a high-temperature short-time pasteurization pilot plant. According to the authors, they found that the studied monitoring and diagnosis charts were able to show deviations in the holding tube-outlet temperature measurements caused by variations in the holding tube-inlet temperature sensor, in the preheater temperature sensor, and in the steam valve of the plant. From this information, corrective actions can be performed in advance and avoid undesired effects on the pasteurized product temperature.

The second study was recently reported by França et al. [46], who monitored the whole production process of craft beer, using NIR spectroscopy and MVPC. In this study, seven batches of Belgian Pale Ale (BPA) craft beer were produced using the same standard machinery (32 L capacity) that most of the home brewers employ. Four of the seven batches with NOC were used to establish the control chart and to study the variability within and among batches,the validation of the model was done using the three remaining batches, two were out of order and only one was under NOC. The control chart was created using PCA of the NOC batches of beer with Hotelling's T^2 and sum of square residuals Q statistics, established at 95% of confidence interval. The authors created a PCA model that successfully associated NIR information with the different steps of the beer production, since the PCs provided essential information concerning biochemical changes in the saccharification process, appearance of fermentable sugar, fermentation, and ethanol transformation by the yeast. Moreover, the authors used the PCA information to build the calibration control chart in which most of the observations were within the established T^2 and Q established limit. When external and validation batches were analyzed with the developed calibration control chart, it was observed that two batches were extremely out of the NOC, specifically in the fermentation step. In this regard, researchers could monitor and control the overall process of beer production in each step of its production through the combination of NIR and MVPC methodologies.

The third reported study in which MVPC was used is the work presented by Schorn-García et al. [41]. In this case, the authors monitored and controlled the possible contaminations with lactic acid bacteria in a wine alcoholic fermentation process using ATR-MIR spectroscopy. Mainly, the authors utilized 10 normal alcoholic fermentations and 4 contaminated fermentations, whose evolutions were properly monitored and detected through a Q-residuals plot, obtained from a previously elaborated PCA model. Thus, the authors were able to follow this process and point out the fermentations out of control, and to detect process deviations using Q-residuals plots and contribution plots, which allowed to assign the cause of such deviations to specific regions of the spectra that are used to differentiate normal and abnormal process samples.

The current section demonstrates that PAT tools, such as DOE, MVDA, and MVPC, are being used in the food industry and have demonstrated their usefulness in the sector, as summarized in Table 1. However, the lack of these studies at industrial and pilot plant scales is notorious.

QbD/PAT Tools in the Food-Related Biotech Industry

The previous section dealt with the multivariate tools recommended by the QbD and PAT system in the food industry. The same three tools are identified in diverse studies at industrial or pilot plant scales in the food-related biotechnology industry and further described in this section. One of the most important features applied within this industry are the bioreactors, which are the key unit operation to perform different processes. As well noted by Boudreau and McMillan [15], the process control of bioreactors tries to influence the reactions inside the cell by regulating the environment that surrounds it, in order to obtain a specific product. Please note that bioreactors are mainly related to biopharmaceutical and biochemical industries,however, these industries are out of the scope of this work, since the use of the PAT multivariate tools are well established and used within them. Instead, emphasis is made on food-related biotech processes and products, such as in the fermentation process, which was the most common topic of research, as it is shown in the following subsections.

Design of Multivariate Experiments (DOE)

Despite of the well-known benefits of DOE to promote innovation and solve problems, only these three studies were found, as shown in this section.

Harms et al. [47] developed a stepwise approach for defining the design space for the production of a protein, which involved the fermentation of a methylotrophic yeast *Pichia pastoris* in a pilot plant facility executing two 300 L runs. The authors designed three different studies taking into account the (i) absorbance and feed rate screening, (ii) culture parameters, and (iii) protein stability. During the first study, the authors used a DOE named fractional factorial screening design with a resolution of IV. Eight factors were tested at two levels in four blocks with one center point per experimental block. Despite of finding statistically significant effects over the final absorbance, the authors considered them as non-key operating parameters, since they were of small magnitude. Regarding the second study, pH and temperature were characterized using a two-level full factorial design (FFD). Results showed that only temperature had a statistically significant effect on titer. The third study was performed with a two-level FFD to estimate all main and interaction effects for the growth and productivity in the induction phase. According to the reported results, the interaction among temperature and dissolved oxygen had a statistically significant effect on the percentage of solids and titer. Additionally, temperature also had a statistically significant effect on titer, considering temperature, pH, and absorbance as key parameters for the growing process. The second part of this study involved the characterization of temperature and pH for the post-induction process using again a two-level FFD. With this study, the authors found that neither pH, temperature, nor their interaction were considered to have a significant effect on post-induction product protein concentration, demonstrating that the product is stable and there was no proteolytic degradation. In this sense, the authors established the design space for the fermentation process and identified temperature, pH, and absorbance as key operating parameters for process characterization through the use of risk analysis and DOE.

Moreover, Bayer et al. [48] proposed the use of DOE with hybrid modeling for process characterization, using 20 L cultivations of *Escherichia coli* fed-batch. The study was split in two phases: (i) finding the model with the best performance in describing the biomass concentration and soluble product titer and (ii) determining which model was most accurate to predict the entire process. The studied models in the first phase were response surface model (RSM) with a FFD, artificial neural network (ANN), and hybrid model (RSM + ANN) and only AAN and hybrid models in the second phase. The authors demonstrated that the hybrid model was superior to the ANN model in predicting the biomass concentration and the soluble product titer. In most of the cases, the hybrid model correctly matched the predicted values with the analytical measurements with small prediction intervals. Regarding these results, an approach was developed to characterize and

Table 1 QbD/PAT implementation in the food industry on a pilot plant and industrial scale

QbD/PAT tool	Food or related item	Goal	Multivariate methodology	Ref
DOE	Coffee	Definition of the design space to optimize the process in terms of energy losses and efficiency	NM	Fissore et al. [24]
	Beer	Manufacturing and optimization of the formulation process	S-CMD	Tessarini et al. [25]
MVDA	Pork meat	Potential of visible spectroscopy to classify and predict meat quality	CDA, PLS	Xing et al. [26]
	Porcine carcasses	Determination of fat quality using spatially resolved NIR spectroscopy	PCA, PLS	Sørensen et al. [27]
	Brined pork	Classification of brined and non-brined pork loins and prediction of salt concentration with Vis–NIR hyperspectral imaging	PCA, PLS-DA, PLS	Achata et al. [28]
	Lamb meat	On-line quantification of fatty acids by Vis–NIR spectroscopy	PLS	Pullanagari et al. [29]
	Deboned chicken and salmon	In-line determination of fatty acids in Salmon and residual bone concentration in chicken using Raman spectroscopy	PLS	Lintvedt et al. [30]
	Salmon	Optimization of the production process to increase the yield	PLS, PCA	Ørnholt-Johansson et al. [31]
	Cheese	Real time modeling of milk coagulation	PCA	Lyngaard et al. [32]
	Milk powder	Prediction of functional properties based on manufacturing data	CPDs, PLS	Rimpiläinen et al. [33]
	Olive oil	Improvement of the olive oil extraction process	PCA	Tamborrino et al. [34]
	Potato 'chips'	Identification of main quality and process parameters	MLR-SW	Picouet et al. [35]
	Tomato paste	Qualitative and quantitative detection of sucrose adulteration	PCA, PLS, LS-SVM, BPNN	Liu et al. [36]
	Milk powder	Development of a real time quality control process using hyperspectral imaging spectroscopy (HIS)	PCA, PLS	Munir et al. [37]
	Soy sauce	Development, optimization and prediction of the desalting process by electrodialysis using on-line NIR spectroscopy	PLS	Moscetti et al. [38]
	Apple puree	Prediction of the composition and texture characteristics using NIR, MIR, Raman spectroscopies and HIS	PLS, RF, SVM	Lan et al. [39]
	Beer	Prediction of the final quality of beer through a real-time monitoring process using ATR-FTIR spectroscopy	PLS	Tessarini et al. [40]
	Wine	Monitoring and control process of the alcoholic fermentation using ATR-MIR spectroscopy	PCA, PLS, PLS-DA	Schorn-García et al. [41]
	Tobacco leaves	Monitoring and prediction of different parameters involved in the flue-curing process using NIR spectroscopy	PLS, SVM, CNN	Wei et al. [42]

Table 1 (continued)

QbD/PAT tool	Food or related item	Goal	Multivariate methodology	Ref
	Instant noodles	Monitoring and prediction of several quality parameters during the production process using NIR-vis spectroscopy	PCA, PLS, SVMR	Upadhyay et al. [43]
MVPC	Milk	Monitoring and control of critical points to detect faults in sensors during early stages of the process	NM	Tokatli et al. [45]
	Craft beer	Monitoring of the whole production process with NIR spectroscopy and control possible deviations of the process	PCA	França et al. [46]
	Wine	Monitoring and identification of contaminated alcoholic fermentations using ATR-MIR spectroscopy	PCA	Schorn-García et al. [41]

BPNN back propagation neural network, *CDA* canonical discriminant analysis, *CPDs* conditional probability distributions, *CNN* convolutional neural networks, *DOE* design of experiments, *LS-SVM* least squares-support vector machines, *MLR-SW* multilinear regression step-wise model, *MVDA* multivariate data analysis, *MVPC* multivariate process control, *NM* not mentioned, *PLS* partial least squares, *PLS-DA* partial least squares-discriminant analysis, *PCA* principal component analysis, *PAT* process analytical technology, *QbD* quality by design, *S-CMD* simplex-centroid mixture design

optimize the entire process using a dynamic hybrid model, making possible to obtain the desired product at the end of the process controlling the CPP. Such results were achievable due to the structure of the model, which differentiate if the variations of the process are caused by the metabolism of the bacteria or due to the process operations.

Lastly, the control and optimization of lactose production through its crystallization process at industrial scale was studied by Galvis et al. [49], who developed a novel strategy based on retrospective QbD approach and new experiments coming from DOE. After the use of MVDA, the authors identified 4 out of 32 variables as critical process parameters (CPPs). These four CPPs were included in a face-centered DOE considering low, medium, and high levels each and with three replications of the design center. The results of these experiments allowed researchers to analyze the effects of the 4 CPPs and their different interactions over the mass percentage of total fines. However, the authors mentioned that the experimental design was not fully performed as it was planned and the experimental factors were not completely independent, thus, the statistical significance of the factors was not reliably quantified. Nonetheless, the authors were able to compare the historical data with the new obtained data using contour plots, finding important insights that allowed them to improve the quality of the final product by up to 7%.

As already outlined in the earlier section focused to the food industry, there are few reported instances using DOE. Thus, more awareness of this tool and effort are needed within the food-related biotech industry to apply DOE.

Multivariate Data Analysis (MVDA)

Due to the importance of MVDA, it has also been applied in food-related biotechnology sector. In fact, this section deals with five studies at industrial plant or pilot plant scale in which MVDA is applied. In this sector, it is common to produce or utilize cell cultures, which need to be monitored during a complex process.

According to this, Abu-Absi et al. [50] decided to assess and monitor different parameters in a cell cultivation process in 500 L bioreactors, using off-line and in-line Raman spectroscopy coupled to MVDA. Samples and measurements were taken and performed from four bioreactor runs,the data from the first three were used for the calibration data set and the last one for the validation data set. The authors intended to predict parameters with PLS and some preprocessing techniques, such as 1st and 2nd derivatives, variance scaling, and SNV path length correction. As reported by the authors, the calibration of the models was good for glutamine, glutamate, glucose, lactate, ammonium, viable cell density (VCD), and total cell density (TCD). Afterwards, these models were validated, and predictions were contrasted with the measured values. Researchers found a model using the three lots for glutamine which differed 30% between the measured and predicted values. For glutamate, the average difference was 12%, for glucose was 15%, for lactate was 13%, and for VCD and TCD the difference was 15%. The only model that did not match the predicted values with the measured values was the ammonium model. Nonetheless, the authors considered that these performances were good enough for the purpose of their study. In this sense, researchers were able to provide immediate feedback and control the process performance using real-time measurements, ensuring consistent manufacture of the mammalian cell cultures using MVDA.

Mercier et al. [51] also used MVDA to monitor the early development of a cell cultivation process. This study consisted on 17 and 10 cell cultivation runs of 2 L and 10 L, respectively. During the initial steps of the process, the evolution of the behavior of the batches was checked, and PLS was employed to relate the data of the process to a response variable which represented the run maturity. Then, batch level modeling was created, considering each batch as single unit. At this point, PCA was used to explore the data and then PLS was employed to understand how the initial conditions of the process influenced over it. When the authors analyzed the score plots of the off-line and on-line variables model, they realized that clusters were clearly observed according to the scale of the cultivations, causing operational differences in both on and off-line process variables. The authors attributed this behavior to the consumption of O_2 and CO_2, which was higher for the 2 L cultures than for 10 L cultures that was associated to the aeration strategy, which was not linearly scaled between the two bioreactor volumes. Additionally, analysts found that the cell diameter for the 10 L cultivations was on average 1.1 µm smaller than the 2 L cultivations that was attributed to the cross flows inside the fibers of the equipment, which were distinct in the two bioreactor scales. PCA was also employed for batch diagnosis in which 7 batches were further analyzed, which showed deviations due to the concentrations of additives in feed medium during the perfusion, inoculation at twice the target cell density for the off-line variables, a change in the procedure for medium preparation, and due to a deviation in absorbance probe calibration for the on-line variables. Moreover, PLS was used to establish correlations between process parameters and process responses; however, the authors reported that the generated PLS models showed a poor fit. In this sense, the authors proved that PCA could be used as valuable tool to identify deviations in early development of cell cultivation processes, being the scale effect a relevant factor to take in to consideration when developing a process as the presented here.

The third study was performed by Ferreira et al. [52], who studied if multiway PCA (MPCA) and multiway PLS (MPLS) could be used to (i) model 16 industrial fermentation processes of *Streptomyces clavuligerus* strain for the production of clavulanic acid using a pilot plant and (ii) to predict the fermentation yield. The acquired data were preprocessed using SG filter and then explored using MPCA. The authors were able to differentiate among batches based on the trajectories of variables measured on-line, being the most different batches 3, 5, 6, and 7 from the rest. This difference was caused in batch 3 for its high conductance profile and for keeping low values of temperature for a long period of time. Batches 5–7 presented different conditions for the substrate addition, producing changes in the quality variables (biomass, absorbance, and conductance). The MPLS was performed to predict the final concentration of the clavulanic acid for each batch and also to evaluate what variables influenced the most over the productivity. With this model, batches 3, 5, 6, and 7 were no different from the others, attributing this behavior to the basis of each method, since MPCA focused on the covariance of the variance, whilst MPLS focused on the covariance of the X-block (process variables) that is more correlated with the Y-block (response variables). Moreover, the authors found that capacitance was the principal variable for the prediction of the final product concentration using the weight contribution plot. In this regard, researchers could improve the knowledge through MVDA on a fermentation process carried out in a pilot plant in which dissimilarities were detected according to abnormal changes in quality variables, predicting the final product with moderate accuracy and detecting the most important variables that influenced the most over the productivity prediction (capacitance and absorbance).

Furthermore, Alves-Rausch et al. [53] performed a real-time multiparameter monitoring during a fermentation process in a 50 L bioreactor, intending to produce *Bacillus* spores and introducing into the study floor vibrations and high humidity as in the industrial environment. The fermentation process was divided in 5 batches, where temperature was controlled at 39 °C and pH at a specific set point by addition of NaOH (50% v/v) during the growth phase or H_2SO_4 (38%) during the sporulation phase. In this case, PCA was performed with spectra collected directly from the reactor at 0.25 and 0.5 h before inoculation. The media formulation was considered "good" or "bad" according to the final production yield of the fermentations. Two models were tried; without any preprocessing and applying SNV, finding that SNV removed most physical effects from the spectra. The SNV-PCA model was used together with the distance to the model in the X-space (DModX), allowing to identify that one batch was different from the other four, which was mainly attributed to a reduction in the content of yeast extract. Regarding the use of XLS (extension of PLS implemented in a specific software), five calibration models with no preprocessing were performed for acetoin, absorbance at $\lambda = 600$ nm (A_{600}), dry mass, and two sum parameters for sugar and analytes, which were considered as indicators of sugar consumption and overall metabolism. With these models, the authors were able to monitor a large-scale industrial fermentation process getting important insights of the process, such as the in-line prediction of acetoin concentrations that gives information of the metabolic state of bacillus, and that A_{600} and biomass in-line values provide important information about the growth and sporulation of the culture growing for further process and medium optimization.

Lastly, Galvis et al. [49] developed a novel strategy based on retrospective QbD approach to control and optimize the crystallization process for lactose production. Such strategy was developed using long-running historical data of 2 years obtained from an industrial production facility, using expert knowledge and including new experiments. The authors intended to improve production quality by reducing the mass percentage of small crystal fines produced as critical quality attributes, and they used different MVDA along the process to achieve this goal. In fact, PCA was used in the first place to detect and remove outlying samples, and PLS was applied to identify the variables of the process that were more critical for the production quality. In this sense, the authors were able to properly identify 4 critical process parameters out of 32 studied variables, using PLS and variable importance in projection (VIP) that once they were optimized, the product quality improved up to 7%.

From this subsection, is evident how MVDA is of great importance to monitor, optimize, and get important information of the fermentation and crystallization processes. Additionally, it is evident that such studies need to be performed and adopted by the industry to take all the advantages of these methodologies.

Multivariate Process Control (MVPC)

From the previous subsection, Alves-Rausch et al. [53] performed a real-time multiparameter monitoring during a fermentation process, which intended to produce *Bacillus* spores. The authors monitored the process by performing a batch evolution model (BEM) based on the NIR data, which captured the variations in the spectra over time, and a reference batch trajectory was built including process control limits based on $\pm$ 3 SD. The BEM was based on the PLS models in which the PLS scores were averaged for each time point. Four well-behaved batches were used to build the model, leaving one batch out for validation. The BEM on the SNV preprocessing data gave them better insights of the three different metabolic stages identified in the BEM plot than the BEM with no preprocessing. In the first stage, microorganisms started to grow and consume the sugar sources, then, microorganisms started to consume the metabolite produced in the first stage, and finally, the metabolites in the media were completely consumed, and the spectral changes were smaller, which may be an indicator that cell growth stopped and microorganisms started to sporulate. In this sense, researchers could monitor a fermentation process using BEM, making possible to create a reference batch trajectory and to detect future deviations for the coming batches.

Another study, performed by Krause et al. [54], monitored seven aerobic fermentation batches of *Saccharomyces pastorianus*, variety *calsbergensis*. This process was carried out at pilot scale in an industrial fermentation tank with 70-L capacity in which information of seven sensors was studied through MVPC and "particle swarm optimization" (PSO). MVPC was based on "unfold-PLS" and was used to create statistically supported process trajectories for process control. Two levels of MVPC were used: level 1 (maturity prediction) and level 3 (residual standard deviation (RSD)). The seven initial input data coming from the sensors were extended by fully polynomial extension of second order including mixed terms, obtaining a total of 35 input variables, but variable importance in the projection (VIP) was applied since not all variables were of the same importance to model the target of interest. Once the VIP was applied, twelve variables were considered for the elaboration of the multivariate process trajectory control charts. In this sense, the authors showed the result for three-time sector through the use of MVPC charts, where all sensors demonstrated to work properly within the established boundaries, showing good similarity among each individual input trend and the historical data. Researchers reported that 12 inputs were used in 90.8% of the modeled cases and 11 in only 8.7% of the other ones. Scientists reported that all trajectories always kept the direction between the established 3 σ limits, developing a successful approach to monitor the trajectory progress of the fermentation process and capable to predict false input information.

The last study found in literature in which MVPC was applied at industrial scale was performed by Gunther et al. [55], who applied PCA and MVPC to industrial fermentation data obtained from the industry. These methodologies were used to detect and diagnose possible abnormal conditions from both on-line and off-line analyses. A total of ten batches with 1084 samples each monitored through 11 process variables were analyzed from 300 L reactors. Batches 1–8 with NOC were used to develop the PCA model, batch 9 to validate it, and batch 10 to detect problems within the process. In fact, the score plot of the PCA model showed a similar trend of the first 9 batches, but batch 10 was clearly different. Furthermore, the authors performed the monitoring process of these batches using T^2 and Q statistics, as part of the MVPC, from the off-line analyses, resulting in the same results as in PCA. These results were further confirmed studying the on-line data of the fermentation batches 9 and 10 on which T^2 and squared prediction error (SPE) were applied as part of the MVPC. Results led to the same conclusions as PCA and T^2 and Q plots; however, in this comparison, SPE evidenced more clearly the fault detections than the T^2 plot. Hence, it was demonstrated that MVPC could help in the monitoring process of a fermentation process identifying NOC and abnormal batches. All the discussed studies in which PAT tools are applied in the food-related biotech industry are summarized in Table 2.

Table 2 QbD/PAT implementation in the biotechnology industry on a pilot plant and industrial scale

QbD/ PAT tool	Process	Goal	Multivariate methodology	Ref
DOE	Fermentation	Definition of the design space for a product	FFD, 2L-FFD,	Harms et al. [47]
	Cultivation of *Escherichia coli*	Fast characterization and optimization of the process	RSM, FFD	Bayer et al. [48]
	Crystallization	Control and optimization of the process using historical and new data	FC	Galvis et al. [49]
MVDA	Cultivation	Assessment and monitoring of the process using different parameters	PLS	Abu-Absi et al. [50]
	Cultivation	Monitoring of the early stage of the process	PCA, PLS	Mercier et al. [51]
	Fermentation	Model the industrial fermentation process and predict its yield	MPCA, MPLS	Ferreira et al. [52]
	Fermentation	Real-time multiparameter monitoring of a fermentation process	PCA, PLS	Alves-Rausch et al. [53]
	Crystallization	Selection of critical process parameters	PCA, PLS	Galvis et al. [49]
MVPC	Fermentation	Real-time multiparameter monitoring of a fermentation process	PLS	Alves-Rausch et al. [53]
	Fermentation	Monitoring of fermentation process	U-PLS	Krause et al. [54]
	Fermentation	Detection of possible abnormal batches	PCA	Gunther et al. [55]

2L-FFD 2-level full factorial design, *BBD* Box-Behnken design, *DOE* design of experiments, *FC* face centered, *FFD* fractional factorial design, *JY-PLS* Join-Y partial least squares, *MVDA* multivariate data analysis, *MVPC* multivariate process control, *MPLS* multiway partial least squares, *MPCA* multiway principal component analysis, *NM* not mentioned, *PLS* partial least squares, *PBD* Plackett–Burman design, *PCA* principal component analysis, *PAT* process analytical technology, *QbD* quality by design, *RSM* response surface model, *U-PLS* unfold partial least squares

Conclusions and Final Remarks

As observed from the "QbD/PAT Tools in the Food Industry" section to the "QbD/PAT Tools in the Food-Related Biotech Industry" section, the multivariate tools for design, data acquisition, and analysis (DOE, MVDA, and MVPC) recommended by the PAT system under a QbD framework were described and discussed within the food and food-related biotech industries. In the case of the food industry, 23 studies were addressed and 11 in the food-related biotech industry. This makes evident the narrow circumstances where DOE, MVDA, and MVP are used at industrial and pilot plant scales. Moreover, several different studies applying similar analytical techniques that the ones exposed here at the industrial, pilot plant, or lab scale can be found in literature, but they are not reported in this review because they were not performed under the QbD/PAT framework nor follow the QbD/PAT principles, such as the investigations gathered by Grassi and Alamprese [56]. The same observation was noted by Djekic et al. [57] when performing a survey to 203 companies from the European Union and abroad. The authors found that the application of models in the food industry consists of simplified models that do not evaluate the processes, quality, or safety conditions and environmental impact, thus, revealing that the application of mathematical models in food companies has not been a matter of interest yet, identifying it as a research gap.

This absence of multivariate tools should be given by different factors, such as (i) poor knowledge on modeling, (ii) not user-friendly models/software, (iii) instability of processes when introducing experimental tests, (iv) additional cost of new experiments that the company is not willing to assume [58], or (v) the high confidentiality of the studies which hinders the free publication of the results in scientific journals. In this regard, it has been noticed that QbD and PAT tools have been applied in the academy mainly for research purposes at lab scale and in some companies to improve quality control and rapidly evaluate the final product to increase productivity, losing the holistic view and devaluating the real purpose of the QbD/PAT system, which is to ensure quality through continuous and real-time feedback (on-, in-line analysis). Researchers have made a great labor in proving the application of QbD/PAT at lab scale, in pilot plants, and, in some cases, at industrial scale, as exposed within this review paper.

However, it is time to stop basing, associating, and focusing QbD and PAT only to the use of analytical instruments and to start sharing, as much as the confidentiality of each company allows results on the use of DOE, MVDA, and MVPC. QbD and PAT have a more profound meaning than using novel analytical analyzers and multivariate data analysis to increase productivity; both approaches intend to help companies at getting better with their manufacturing and quality assurance processes, products, and final customer to whom is directed. For this to start changing in the food and food-related biotech industries, both of them with their corresponding academia and regulatory organizations should cooperate more to bring together all the diffused work produced by each of them to create a more solid QbD framework with PAT as an enabler. As noted by O'Donell et al. [59], adopting this strategy might create a society for the both industries in which QbD/PAT will be the core center of their activities, assembling chief executive officers and associated companies, government representatives, process engineers, scientists, and technicians, aiming to provide to these industries with a stronger, smarter, and more efficient working framework for the upcoming years.

Summarizing, the key historical aspects and fundamentals of the QbD framework and PAT system were reviewed, as well as their application within the food and food-related biotechnology industries. Special attention was given to the use of multivariate tools for design, data acquisition, and analysis in these industries. A total of 34 case studies were found in literature in which DOE, MVDA, and MVPC at lab scale, pilot plant, and industrial scale were applied for both industries. From this revision, it was observed that the implementation of these tools is still under research and that food and food-related biotech companies are not applying them within their processes, with the exception of some studies reported in this work. It was also noticed that QbD and PAT are being used indistinctly by these industries with emphasis on analytical instruments and multivariate tools to make their analyses and processes faster. In this sense, the authors make an appeal of encouragement to both industries, to researchers, and academia to work closer and improve the current practices, aiming to start a new direction for QbD and PAT in order to adopt them as the leading rules within their processes and industries.

Acknowledgements C.H. Pérez-Beltrán acknowledges the scholarship provided by the Autonomous University of Sinaloa (México). A. Torrente-López acknowledges the FPU predoctoral grant (ref.: FPU18/03131), which is currently receiving from the Ministry of Universities, Spain.

Author Contribution All the authors contributed to the review conception and design. Literature search was performed by Christian H. Pérez-Beltrán and Luis Cuadros-Rodríguez. The first draft of the manuscript was written by Christian H. Pérez-Beltrán, and all the authors revised and commented on subsequent versions. All the authors read and approved the final manuscript.

Declarations

Competing interests The authors declare no competing interests.

References

1. Koch KH (1999) Process analytical chemistry – control, optimization, quality, economy. Berlin, Germany
2. Callis JB, Illman DL, Kowalski BR (1987) Process analytical chemistry. Anal Chem. https://doi.org/10.1021/ac00136a001
3. Juran JM (1986) The quality trilogy: a universal approach to managing for quality. Qual Prog 19(8):19–24
4. Juran JM (1992) Juran on quality by design: the new steps for planning quality into goods and services. New York, USA
5. US FDA (2004) Guidance for industry: PAT – A framework for innovative pharmaceutical development, manufacturing, and quality assurance. U.S. Food and Drug Administration
6. ASTM E2363–14 (2014) Standard terminology relating to process analytical technology in the pharmaceutical industry. ASTM International, West Conshohocken, USA
7. Koch MV, VandenBussche KM, Chrisman RW (2007) Micro instrumentation for high throughput experimentation and process intensification – a tool for PAT. Weinheim, Germany
8. Bakeev KA (2010) Process analytical technology. England, Chichester
9. ICH Q8(R2) (2009) Pharmaceutical development. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Geneva, Switzerland
10. Orlandini S, Pinzauti S, Furlanetto S (2013) Application of quality by design to the development of analytical separation methods. Anal Bioanal Chem. https://doi.org/10.1007/s00216-012-6302-2
11. Chew W, Sharratt P (2010) Trends in process analytical technology. Anal Methods. https://doi.org/10.1039/C0AY00257G
12. Casian T, Nagy B, Kovács B, Galata DL, Hirsch E, Farkas A (2022) Challenges and opportunities of implementing data fusion in process analytical technology – a review. Molecules. https://doi.org/10.3390/molecules27154846
13. ASTM E456–13A (2017) Standard terminology relating to quality and statistics. ASTM International, West Conshohocken, USA
14. ISO 3534–3:2013 (2013) Statistics – Vocabulary and symbols – Part 3: Design of experiments. International Organization for Standardization, Geneve, Switzerland
15. Boudreau MA, McMillan GK (2007) New directions in bioprocess modeling and control: maximizing process analytical technology benefits. North Carolina, U.S.A.
16. ASTM ES891–20 (2020) Standard guide for multivariate data analysis in pharmaceutical development and manufacturing applications. ASTM International, West Conshohocken, USA
17. ISO 7870–7:2020 (2020) Control charts – Part 7: Multivariate control charts. International Organization for Standardization, Geneve, Switzerland
18. Teixeira JA, Vicente AA, Macieira da Silva FF, Azevedo Lima da Silva JS, da Costa Martins RM (2014) In: Teixeira JA, Vicente AA (eds) Engineering Aspects of Food Biotechnology, 1st edn. CRC Press, Boca Raton, USA
19. Hitzmann B, Hauselmann R, Niemoeller A, Daryoush Sangi D, Traenkle J, Glassey J (2015) Process analytical technologies in food industry – challenges and benefits: a status report and recommendations. Biotechnol J. https://doi.org/10.1002/biot.201400773
20. Rathore AS, Kapoor G (2017) Implementation of quality by design toward processing of food products. Prep Biochem Biotechnol. https://doi.org/10.1080/10826068.2017.1315601
21. Rifna EJ, Pandiselvam R, Kothakota A, Subba Rao KV, Dwived M, Kumar M, Thirumdas R, Ramesh SV (2022) Advanced process analytical tools for identification of adulterants in edible oils – a review. Food Chem. https://doi.org/10.1016/j.foodchem.2021.130898
22. Khan IA, Smillie T (2012) Implementing a "quality by design" approach to assure the safety and integrity of botanical dietary supplements. J Nat Prod. https://doi.org/10.1021/np300434j
23. Misra NN, Sullivan C, Cullen PJ (2015) Process analytical technology (PAT) and multivariate methods for downstream process. Curr Biochem Eng. https://doi.org/10.2174/2213385203666150219231836
24. Fissore D, Pisano R, Barresi AA (2014) Applying quality-by-design to develop a coffee freeze-drying process. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2013.09.018
25. Tessarini ES, De Almeida e Silva JB, Rebello Lourenço F (2021) Development and optimization of beer containing malted and non-malted substitutes using quality by design (QbD) approach. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2020.110182
26. Xing J, Ngadi M, Gunenc A, Prasher S, Gariepy C (2007) Use of visible spectroscopy for quality classification of intact pork meat. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2007.01.020
27. Sørensen H, Petersen KM, Engelsen SB (2012) An on-line NIT method for determining depth profiles of fatty acid composition and iodine value in porcine adipose fat tissue. Appl Spectrosc. https://doi.org/10.1366/11-06396
28. Achata EM, Inguglia ES, Esquerre CA, Tiwari BK, O'Donnell CP (2019) Evaluation of Vis-NIR hyperspectral imaging as a process analytical tool to classify brined pork samples and predict brining salt concentration. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2018.10.022

29. Pullanagari RR, Yule IJ, Agnew M (2015) On-line prediction of lamb fatty acid composition by visible near infrared spectroscopy. Meat Sci. https://doi.org/10.1016/j.meatsci.2014.10.008
30. Lintvedt TA, Andersen PV, Afseth NK, Marquardt B, Gidskehaug L, Wold JP (2022) Feasibility of in-line Raman spectroscopy for quality assessment in food industry: how fast can we go? J Food Eng. https://doi.org/10.1177/00037028211056931
31. Ørnholt-Johansson G, Gudjónsdóttir M, Engelbrecht Nielsen M (2017) Analysis of the production of salmon fillet – prediction of production yield. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2017.02.022
32. Lyndgaard CB, Engelsen SB, Van den Berg FWJ (2012) Real-time modelling of milk coagulation using in-line near infrared spectroscopy. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2011.07.029
33. Rimpiläinen V, Kaipio JP, Depree N, Young BR, Wilson DI (2015) Predicting functional properties of milk powder based on manufacturing data in an industrial-scale powder plant. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2014.12.010
34. Tamborrino A, Squeo G, Leone A, Paradiso VM, Romaniello R, Summo C, Pasqualone A, Catalano P, Bianchi B, Caponio F (2017) Industrial trials on coadjuvants in olive oil extraction process – effect on rheological properties, energy consumption, oil yield and olive oil characteristics. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2017.02.019
35. Picouet PA, Gou P, Pruneri V, Diaz I, Castellari M (2019) Implementation of a quality by design approach in the potato chips frying process. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2019.04.013
36. Liu C, Hao G, Su M, Chen Y, Zheng L (2017) Potential of multispectral imaging combined with chemometric methods for rapid detection of sucrose adulteration in tomato paste. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2017.07.026
37. Munir MT, Wilson DI, Yu W, Young BR (2018) An evaluation of hyperspectral imaging for characterising milk powders. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2017.10.001
38. Moscetti R, Massantini R, Fidaleo M (2019) Application on-line NIR spectroscopy and other process analytical technology tools to the characterization of soy sauce desalting by electrodialysis. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2019.06.022
39. Lan W, Baeten V, Jaillais B, Renard CMGC, Arnould Q, Chen S, Leca A, Bureau S (2022) Comparison of near-infrared, mid-infrared, Raman spectroscopy and near-infrared hyperspectral imaging to determine chemical, structural and rheological properties of apples purees. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2022.111002
40. Tessarini ES, Rebello Lourenço F (2020) Real-time monitoring of beer parameters using infrared spectroscopy - a process analytical technology approach. J AOAC Int. https://doi.org/10.1093/jaoacint/qsaa057
41. Schorn-García D, Cavaglia J, Giussani B, Busto O, Aceña L, Mestres M, Ricard Boqué R (2021) ATR-MIR spectroscopy as a process analytical technology in wine alcoholic fermentation – a tutorial. Microchem J. https://doi.org/10.1016/j.microc.2021.106215
42. Wei K, Bin J, Wang F, Kang C (2022) On-line monitoring of the tobacco leaf composition during flue-curing by near-infrared spectroscopy and deep transfer learning. Anal Lett. https://doi.org/10.1080/00032719.2022.2046021
43. Upadhyay R, Gupta A, Niwas Mishra H, Bhat SN (2022) At-line quality assurance of deep-fried instant noodles using pilot scale visible-NIR spectroscopy combined with deep-learning algorithms. Food Control. https://doi.org/10.1016/j.foodcont.2021.108580
44. Bersimis S, Panaretos J, Psarakis S (2005) Multivariate statistical process control charts and the problem of interpretation – a short overview and some applications in industry. Proceedings of the 7th Hellenic European Conference on Computer Mathematics and Its Applications. Athens, Greece
45. Tokatli F, Cinar A, Schlesser JE (2005) HACCP with multivariate process monitoring and fault diagnosis techniques: application to a food pasteurization process. Food Control. https://doi.org/10.1016/j.foodcont.2004.04.008
46. França L, Grassi S, Pimentel MF, Amigo JM (2021) A single model to monitor multistep craft beer manufacturing using near infrared spectroscopy and chemometrics. Food Bioprod Process. https://doi.org/10.1016/j.fbp.2020.12.011
47. Harms J, Wang X, Kim T, Yang X, Rathore AS (2008) Defining process design space for biotech products: case study of *Pichia pastoris* fermentation. Biotechnol Prog. https://doi.org/10.1021/bp070338y
48. Bayer B, Von Stosch M, Striedner G, Duerkop M (2020) Comparison of modeling methods for DoE-based holistic upstream process characterization. Biotechnol J. https://doi.org/10.1002/biot.201900551
49. Galvis L, Offermans T, Bertinetto CG, Carnoli A, Karamujic E, Li W, Szymanska E, Buydens LMC, Jansen JJ (2022) Retrospective quality by design r(QbD) for lactose production using historical process data and design of experiments. Comput Ind. https://doi.org/10.1016/j.compind.2022.103696
50. Abu-Absi NR, Kenty BM, Ehly Cuellar M, Borys MC, Sakhamuri S, Strachan DJ, Hausladen MC, Jian Li Z (2010) Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. Biotechnol Bioeng. https://doi.org/10.1002/bit.23023
51. Mercier SM, Diepenbroek B, Dalm MCF, Wijffels RH, Streefland M (2013) Multivariate data analysis as a PAT tool for early bioprocess development data. J Biotechnol. https://doi.org/10.1016/j.jbiotec.2013.07.006
52. Ferreira AP, Almeida Lopes J, Cardoso de Menezes J (2007) Modelling industrial fermentation data with multiway multivariate techniques. IFAC Proc Vol. https://doi.org/10.3182/20070606-3-MX-2915.00044
53. Alves-Rausch J, Bienert R, Grimm C, Bergmaier D (2014) Real time in-line monitoring of large-scale Bacillus fermentations with near-infrared spectroscopy. J Biotechnol. https://doi.org/10.1016/j.jbiotec.2014.09.004
54. Krause D, Hussein MA, Becker T (2015) Online monitoring of bioprocesses via multivariate sensor prediction within swarm intelligence decision making. Chemom Intell Lab Syst. https://doi.org/10.1016/j.chemolab.2015.04.012
55. Gunther JC, Conner JS, Seborg DE (2007) Fault detection and diagnosis in industrial fed-batch cell culture. Biotechnol Prog. https://doi.org/10.1021/bp070063m
56. Grassi S, Alamprese C (2018) Advances in NIR spectroscopy applied to process analytical technology in food industries. Curr Opin Food Sci. https://doi.org/10.1016/j.cofs.2017.12.008
57. Djekic I, Mujčinović A, Nikolić A, Jambrak AR, Papademas P, Feyissa AH, Kansou K, Thomopoulos R, Briesen H, Kavallieratos NG, Athanassiou CG, Silva CLM, Sirbu A, Moisescu AM, Tomasevic I, Brodnjak UV, Charalambides M, Tonda A (2019) Cross-European initial survey on the use of mathematical models in food industry. J Food Eng. https://doi.org/10.1016/j.jfoodeng.2019.06.007
58. Pietraszek J, Radek N, Goroshko AV (2020) Challenges for the DOE methodology related to the introduction of Industry 4.0. Prod Eng Arch. https://doi.org/10.30657/pea.2020.26.33
59. O'Donnell CP, Fagan C, Cullen PJ (2014) Process analytical technology for the food industry. New York, USA