



This is a repository copy of *Serial artificial neural networks characterized by Gaussian mixture for the modelling of the Consigma25 continuous manufacturing line*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/233254/>

Version: Accepted Version

Article:

AlAlaween, W.H., Mahfouf, M. orcid.org/0000-0002-7349-5396, Omar, C. orcid.org/0000-0002-7839-608X et al. (3 more authors) (2024) Serial artificial neural networks characterized by Gaussian mixture for the modelling of the Consigma25 continuous manufacturing line. Powder Technology, 434. 119296. ISSN: 0032-5910

<https://doi.org/10.1016/j.powtec.2023.119296>

© 2023 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in Powder Technology is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Serial Artificial Neural Networks Characterized by Gaussian Mixture for The Modelling of The Consigma25 Continuous Manufacturing Line

Wafa' H. AlAlaween^{1*}, Mahdi Mahfouf², Chalak Omar³, Riyadh B Al-Asady³, Daniele Monaco³ and Agba D. Salman^{3*}

¹Department of Industrial Engineering, The University of Jordan, Jordan

²Department of Automatic Control and Systems Engineering, The University of Sheffield, UK

³Department of Chemical and Biological Engineering, The University of Sheffield, UK

(*Email addresses for corresponding authors: a.d.salman@sheffield.ac.uk (Agba D. Salman); w.alaween@ju.edu.jo (Wafa' H. AlAlaween))

Abstract

In this research, the Consigma25 Continuous Manufacturing (CM) Line is statistically analysed and modelled. First, the main effects plot is employed to examine the effects of different process parameters on the granules size and the tablet strength. Second, a modelling framework based on serial interconnected artificial neural networks is proposed to model the CM line by mapping these parameters to the granules size and the tablet strength. Then, Gaussian mixture models (GMMs) are adopted to characterize the error resulting from these networks in a way that helps in extracting more information and, as a result, improves the performance of the modelling framework. Validated on an experimental data set, the proposed interconnected framework can anticipate the characteristics of the granules and tablets produced using a specific blend of excipients with an absolute error percentage value of less than 12.3%. In addition, the GMMs have improved the predictive performance by 9.7%.

Keywords

Artificial neural network, Consigma25 continuous manufacturing line; Gaussian Mixture Model; Serial interconnected framework.

1. Introduction

The pharmaceutical industry has been adapting itself to the Fourth Industrial Revolution by more and more shifting to continuous manufacturing. Such a shift has several

advantages in terms of capital footprint, cost and time-to-market [1]. Shifting to continuous manufacturing means that the unit operations (e.g., mixing, granulation and tableting) need to work in a continuous mode. Recently, continuous manufacturing has become a leading technology for solid oral dosage forms in the pharmaceutical industry [2]. Therefore, the Continuous Manufacturing (CM) Line has attracted a lot of interest. Various research papers have, for instance, focused on the processes included in the CM line [3,4]. Among the various processes included in CM, wet granulation, as a process that aggregates fine particles to obtain granules with enhanced properties, has been extensively investigated [2, 4]. Twin screw wet granulation is commonly utilized in the CM line, this being due to its design and short time [5]. Many research papers have focused on such a process and studied various related topics including (i) understanding the influence of different granulation parameters on the granule properties and the different granulation mechanisms to provide a comprehensive understanding of this process; and (ii) developing data and physical based models (e.g., a fuzzy logic system and population balance models) to represent such a process at the micro and macro levels and to predict the granules properties in terms of size, moisture content and porosity [3,4,5]. For instance, a fuzzy logic system, as a transparent and cost-effective model, was established to linguistically represent the twin screw granulation process and to accurately predict the granules size [5]. In addition to the twin screw granulation process, the drying process in the CM line which is required to obtain the appropriate moisture content has also been examined [6]. Different process designs (e.g., a continuous fluidized bed dryer) leading to different monitoring and control strategies of the drying process were examined [6,7]. The effects of these different strategies and their related parameters on the granule size were also analysed [6]. Furthermore, the drying performance of the fluidized bed chambers, as the drying unit of the Consigma25 line, was investigated during continuous production over time [8]. In addition, the effects of the drying variables on the moisture level were also analysed and modelled [9].

Likewise, the breakage of granules during the drying process was studied [10]. The tableting process has also been investigated in the related literature [3]. For instance, it was modelled by developing a data-driven modelling paradigm that was utilized in a reverse-engineering framework to predict the tablet properties as well as control the process [3].

Many papers have focused on the development and control of the CM line [9,10,11]. Therefore, several algorithms have been presented for the acquisition of real data and online monitoring of tablet properties [11,12]. In addition, various process analytical technologies, which consist of interdisciplinary multivariate techniques for material and process analyses, were also utilized to verify whether the various CM line processes were in a control state or not [13]. Such technologies were implemented based on quality control concepts to test blend uniformity [14]. Likewise, multivariate statistical modelling tools based on chemometric approaches were employed to analyse the tableting process during the drug development stage [15]. Likewise, a parameterized model was developed based on a limited number of experimental data points to work as a soft sensor for anticipating the granules moisture content [9]. In addition and based on feeder data and residence time based models, soft sensors were utilized to detect out-of-specification materials and manage their discharge [16]. Such models were validated on wet granulation with a focus on the quality attributes of tablets in order to discharge the out-of-specification ones [16]. However, they could not maintain the process parameters close to the predefined values or manage the undesired events [1]. Therefore, advanced modelling and control strategies have been presented [17,18]. For instance, a robust manufacturing model was presented to predict, understand and control processes and products [17]. In addition, a model integrated with real-time tablet weight measurements was proposed to optimize tablet weight and hardness by controlling filling depth and compression height [18]. In general, the predictive modelling, control and understanding strategies presented in the related literature have their strengths and limitations. Thus, they have been employed according

to the challenges that need to be faced, and applied to the processes that need to be controlled. Therefore, in this research work, the Consigma25 CM line has been statistically analysed and modelled. The main effects plot has been employed to examine the effects of different process parameters on the granules size and the tablet strength. Then, a data based modelling approach based on serial artificial neural networks (ANNs) has been presented for modelling the various processes in the Consigma25 CM line. The ANN architecture was elicited in this paper because of its abilities to (i) capture highly dimensional non-linear behaviours; (ii) deal with limited number of data points; and (iii) take into account the unequal distribution of the data points in the space examined. In such an approach, various ANN models have been developed based on a limited experimental data set to represent the various processes in such a CM line. Such models are connected together in a way that the outputs of one ANN model in addition to the other process parameters of a process are considered as inputs for the next ANN model and so on. Such a way of connecting these models emanates from the fact that the output of one process in this line affects the downstream processes. Therefore, such a framework can better simulate the nature of the Consigma25 CM line. In order to improve the predictive performance of the serial ANNs, Gaussian mixture models (GMMs) have been adopted to characterize the error resulting from these networks in a way that extracts more information. In this paper, Section 2 presents the experimental work and the parameters investigated. Background about the ANN and the GMM is presented in Section 3, whereas Section 4 summarizes the implementation of the framework that is based on the ANNs and the employment of the GMMs, and the results obtained. Finally, conclusions are discussed in Section 5.

2. EXPERIMENTAL WORK

In this research work, the Consigma25 line (GEA Pharma systems, Collette, Wommelgem, Belgium), as a CM line, was utilized to produce granules and tablets. Such a CM line, which was installed at the Diamond Pilot Plant (DiPP) at The University of Sheffield,

mainly consists of several unit operations (i.e., equipment), namely, twin screw granulator, fluidized bed dryer, milling, blending and tableting, as shown in Figure 1. Lactose (Pharmatose® 200M, DFE Pharma) was fed to a co-rotating twin screw granulator at different feed rates (i.e., 9, 10 and 11 kg/h). Deionised water, as a liquid binder, was added at a constant L/S ratio of 0.1. The twin screw granulator has 25mm diameter screws with a length-to-diameter ratio of 20:1. The screw configuration that consists of kneading and conveying elements was kept the same for all experiments. In addition to the powder feed rate, three levels of screw speed (i.e., 250, 500 and 750rpm) were examined in this research work.



Figure 1. The Consigma25 continuous manufacturing line at The University of Sheffield.

Once the granulation process is completed, the wet granules were transferred to the fluidized bed dryer that consists of six chambers, each one was filled in 5 minutes. The temperature, airflow rate and drying time were all kept constant at values of 40°C, 360m³/h and 12 minutes, respectively. The dried granules were then transferred to a conical mill

(Quadro comil) equipped with a 1575 μ m smooth screen. Three levels of milling time (i.e., 1, 2 and 3 minutes) were also studied, whereas the mill speed was kept constant at 1200rpm. As it is a common lubricant that has several advantages such as reducing friction and preventing sticking, the milled granules were then blended with Magnesium Stearate (Acros Organics, Netherlands) at 1% for 60 seconds. The granules blended with Magnesium Stearate were, then, compressed to produce tablets using 12mm diameter die. All the parameters of the tableting machine were kept constant. For instance, the compression force was kept constant at 5KN.

Table 1 summarizes all the parameters that kept constant.

Table 1. The fixed process parameters.

Processes	Parameters	Values
Twin screw granulation	L/S ratio	0.1
	Screw configuration	Conveying and kneading
	Granulator temperature	25°C
Fluidized bed	Filling time	5 min
	Temperature	40°C
	Air flow	360m ³ /h
	Drying time	12min
Milling	Mill speed	1200rpm
	Screen size	1575 μ m
Blending	Blending time	60 sec
	Lubricant percentage	1%
Tabletting	Die diameter	12mm
	Weight target	0.7gm
	Force	5KN

The granules and tablets in this experimental work were characterized. After the granulation process was completed, three samples of wet unmilled granules from the twin screw granulator were collected in trays and left to dry at a room-temperature overnight. Each sample was of approximately 20gm, which is considered as a representative sample collected directly from the twin screw granulator. The three dried samples were then characterized using CAMSIZER (Retsch Technology GmbH, Haan, Germany) to measure the granules size based on a real-time dynamic image analysis. For this purpose, such samples were placed in the

funnel of the CAMSIZER and were automatically fed to it through a feeder at a constant feed rate. The automatic feeding process ensures a monolayer of well-dispersed granules in front of two digital cameras. These granules pass in front of LED light sources. The resulting shadow projections are usually captured with two digital cameras: a basic camera that captures the images of relatively large granules; and a zoom camera that captures the images of small granules. It is worth emphasising that such a dual-camera setup detects the shadow projections of the granules at different positions and angles to take into account the different (perhaps non-spherical) shapes of the granules. Likewise, it allows for capturing and analysing a wide range of granules' sizes.

The tablets strength was also measured using automated tablet hardness tester (P5 Series, Kraemer-Elektronik GmbH, Darmstadt, Germany). It is worth emphasizing that three parameters, namely, feed rate, screw speed and milling time are investigated in this research work, this being due to their effects on the granules size and tablet strength [9,10,11]. Table 2 shows the nine experiments conducted in this research work and their processes' parameters. It can be seen that these data points are not equally distributed in the space examined.

Table 2. The process parameters and the output of the nine experiments.

Experiments	Powder Feed rate (Kg/hr)	Screw Speed (rpm)	Milling time (min)
1	10	250	1
2	10	250	2
3	10	250	3
4	10	250	1
5	10	500	1
6	10	750	1
7	9	250	1
8	10	250	1
9	11	250	1

3. Models Development

3.1 Artificial Neural Networks (ANNs)

The recent computational advancements have been the main reason behind the extensive use of efficient data based models in many areas such as logistics, pharmaceuticals, diseases and manufacturing [19,20,21]. For instance, several data based models (e.g., radial based functions and fuzzy logic systems) have been utilized to represent and simulate various dry/wet batch/continuous granulation processes [3,4,5]. For instance, an integrated network based on radial functions was proposed to model the high shear process, as batch wet granulation [21]. Likewise, a data-driven modelling paradigm based on an interval type-2 fuzzy logic system was also suggested to model the twin screw granulation process [5]. Such modelling paradigms in addition to other models presented in the related literature were able to predict the granule properties for these processes [22]. However, these paradigms were applied to a single process. Therefore, these paradigms cannot be transferable to the CM line processes without careful consideration of the continuous nature of such a line. Therefore, in this research paper, an interconnected framework based on serial data-driven paradigms is proposed to model the Consigma25 CM line. Various data based models can be used. Among them all, the ANN has been extensively used to replace or complement physical based models, this being due to its ability to simulate the human cognitive process [23].

In general, an ANN multi-input single-output structure consists of three types of layers, as shown in Figure 2. These layers are: (i) an input layer that represents the n inputs considered ($\mathbf{x}=x_1, x_2, \dots, x_n$), where each input is schematically represented as an input neuron; (ii) a hidden layer that consists of m hidden neurons representing transfer functions that are usually employed to mathematically extract the input/output relationships; and (iii) an output layer that represents a single output neuron (y) for such a structure. It is worth mentioning that different

numbers of hidden layers can be used depending on the complexity of the relationships and behaviours that need to be captured. In this research work, a one hidden layer was used in all the ANNs developed. Various types of transfer functions including, for instance, linear and Gaussian functions can be used. The selection of the transfer function depends on the nature of the input/output relationships and its complexity. Because of its continuity and easiness to derive, the sigmoid function was utilized in this research paper for all the hidden neurons. The sigmoid function ($\Phi_j(\mathbf{x})$) for the j^{th} hidden neuron can be mathematically expressed as follows [24]:

$$\Phi_j(\mathbf{x}) = \frac{1}{1 + e^{-\sum_{i=1}^n \lambda_{ij}x_i + \lambda_{oj}}} \quad \text{Equation (1)}$$

where λ_{ij} and λ_{oj} stand for the weights connecting the i^{th} input neuron with the j^{th} hidden neuron and the bias, respectively. The predicted output (y) can usually be written as a linear function of the outputs of the hidden neurons. Therefore, it can be expressed as follows [24]:

$$y = \sum_{j=1}^m \eta_j \Phi_j(\mathbf{x}) + \eta_o \quad \text{Equation (2)}$$

where η_j and η_o stand for the weights connecting the j^{th} hidden neuron with the output and the bias, respectively. All the weights used in the network are initially defined randomly and, then, optimized using one of the common optimization algorithms such as Levenberg-Marquardt and Scaled Conjugate Gradient. Because of its proven efficiency and effectiveness and the limited number of experimental data points, the Bayesian algorithm was utilized to optimize the network parameters using two performance measures, namely, the root mean square error (RMSE) and the coefficient of determination (R^2) [24]. Such an algorithm was embedded with the backpropagation structure to optimize the weights connecting the i^{th} input with the j^{th} hidden neuron and the weights connecting the j^{th} hidden neuron with the output [20].

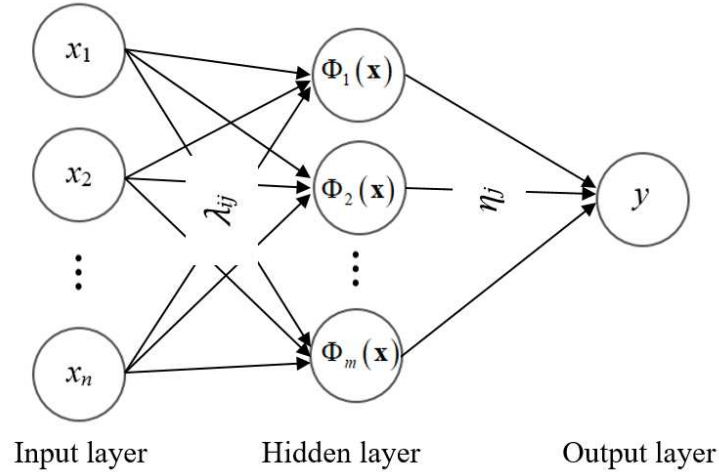


Figure 2. The artificial neural network.

The numbers of neurons in the input and output layers are commonly determined from the case under investigation. Although several outputs are investigated in this research work, a multi-input single-output structure was selected in order to improve the predictive performance of each output. It is also worth emphasising that the number of neurons in the hidden layer plays a crucial role in determining the performance of the ANN. Several algorithms have been proposed in the literature to determine the optimal number of hidden neurons that should be used [25]. However, there is no universally accepted approach that can be used to identify the best number of hidden neurons. Therefore, in this research paper, such a number was identified according to the maximum performance obtained. In other words, it is the one that can lead to the optimal performance and error [19].

3.2 Gaussian Mixture Model

In models development, information may sometimes not be fully extracted, this being due to the error normality assumption. Therefore, extracting such information can significantly enhance the predictive performance of a model [26-28]. Various paradigms have hitherto been proposed in the literature to extract the possible hidden information by analysing errors [21]. Such paradigms include the GMM, time series error residuals and fuzzy clustering [27,28].

Among the various paradigms, the GMM has been utilized in many research articles due to its proven efficiency when it comes to error characterization [21]. This can also be attributed to its capability to analyse the probability density with a satisfactory performance when a good number of Gaussian functions is employed [21]. Due to complexity of the processes examined in this paper and measurement uncertainties, it was, therefore, utilized to provide a representative insight into the probability density function, which can be used refine the predictive performance of the models developed.

In general, the GMM, as a probabilistic paradigm, can be expressed as a combination of Gaussian components that have their own parameters in terms of mean (μ_l), mixing coefficient (π_l) and covariance (Σ_l). Therefore, such a combination with L components can simply be expressed as follows [24]:

$$p(x^e) = \sum_{l=1}^L \pi_l N(x^e \mid \mu_l, \Sigma_l) \quad \text{Equation (3)}$$

where x^e stands for the inputs data that will be used in this model which includes process inputs and errors obtained from the predictive model, and the letter e is employed to distinguish the GMM inputs from the ANN ones. The best values of the l^{th} Gaussian's parameters are commonly identified by optimizing the log likelihood function leading to the following equations [21]:

$$\begin{aligned}
\eta(z_{kl}) &= \frac{\pi_l N(x_k^e | \mu_l, \Sigma_l)}{\sum_{l=1}^L \pi_l N(x_k^e | \mu_l, \Sigma_l)}, \quad \forall l \\
\left\{ \begin{aligned}
\mu_l &= \frac{\sum_{k=1}^K \eta(z_{kl}) x_k^e}{\sum_{k=1}^K \eta(z_{kl})} \\
\Sigma_l &= \frac{\sum_{k=1}^K \eta(z_{kl}) (x_k^e - \mu_l)(x_k^e - \mu_l)^T}{\sum_{k=1}^K \eta(z_{kl})} \\
\pi_l &= \frac{\sum_{k=1}^K \eta(z_{kl})}{K}
\end{aligned} \right\}, \quad \forall l
\end{aligned} \tag{4}$$

where $\eta(z_{kl})$ represents the probability associated with the belonging of the k^{th} data point to a Gaussian component (l), where z_{kl} stands for a L -dimensional vector variable that has a value of 1 when the k^{th} point belongs to the l^{th} Gaussian component and zero otherwise. Identifying an analytical solution for Equations (4) is not as simple as it may seem. However, one of the common algorithms that can be used to estimate a solution is the Expectation Maximization (EM) one [29]. Initially, K-means clustering can be utilised to identify the Gaussian components' parameters. This is followed by estimating the value of $\eta(z_{kl})$. Such a step is referred to as E-step. The $\eta(z_{kl})$ is then used to re-estimate the Gaussian components' parameters, the so-called M-Step. Then, the new values of the parameters are employed to re-evaluate the $\eta(z_{kl})$ value. Such iterations are continued until one of the stopping criteria is met [21].

Since the best number of components (L) is not known in advance, the EM algorithm is also adopted in this research paper to identify such a best number. For this purpose, the Bayesian information criterion (BIC) is used [29]. Upon the convergence of the EM algorithm,

the conditional mean and standard deviation are used to compensate for the bias (i.e., enhancing the performance) and estimate the confidence level [21]. For this purpose, the estimated conditional mean is then added to the predicted value in order to compensate for the bias. This can enhance the model performance. The estimated conditional standard deviation is utilized to identify the confidence level [21].

4. Implementation and Results

4.1 Experimental Results: Main Effects Analysis

To investigate the influence of the three parameters that were examined in this research work, the main effects plots were employed to analyse the effects of these three parameters at their different levels considered. In general, the main effects plot is a visual representation of the mean output values at the different levels investigated for each parameter [30]. Such a plot can be used to determine and, thus, compare the relative effects of various parameters in terms of the nature of the relationships (i.e., direct or inverse relationships) and their relative strength. Figure 3 presents the main effects plots of the three parameters (i.e., powder feed rate, screw speed and milling time) on two outputs, namely, D_{50} (i.e., the size that 50% of the granules have size that is below this value and 50% of the granules have size that is above this value) which was used as an indicated example of the granules size, and the tablet strength. The results shown in such a figure reveal that these three variables have significant however different effects on D_{50} and the tablet strength. To illustrate, increasing the feed rate from 9 kg/h to 10kg/h led to a decrease in D_{50} . However, increasing it from 10kg/h to 11kg/h led to a significant increase in D_{50} . This is an indication that the relationship between the feed rate and D_{50} is a nonlinear one. Increasing the feed rate, however, led to an increase in the tablet strength. It can be noticeable that the relationship between the feed rate and D_{50} is more considerable than the relationship between the feed rate and the tablet strength. It is also apparent in Figure 3 that the screw speed can significantly affect D_{50} and the tablet strength in

different ways. To elucidate, increasing the screw speed from 250rpm to 500rpm led to a significant increase in D_{50} and a decrease in the tablet strength. However, increasing the screw speed from 500rpm to 750rpm led to a decrease in D_{50} and an increase the tablet strength. Likewise, the change in the milling time effects can be clear, where increasing the milling time from 1 minute to 2 minutes and from 2 minutes to 3 minutes increased the tablet strength. Likewise, the increase in the tablet strength when the milling time increased from 1 minute to 2 minutes was much more than it when the milling time increased from 2 minutes to 3 minutes. The effect of the milling time on D_{50} was not considered in such statistical analyses as the milling process was performed after characterising the granules size. In other words, the milling time does not affect the granules size in the Consigma25 CM line. However, it affects the downstream processes and their outputs. Based on the main effects analyses conducted in this section, it was found that the three parameters investigated in this research paper have significant effects on the granules size represented by D_{50} and the tablet strength. In addition, the relationships between them may not be linear. Therefore, these parameters need to be considered in the framework developed in this research work.

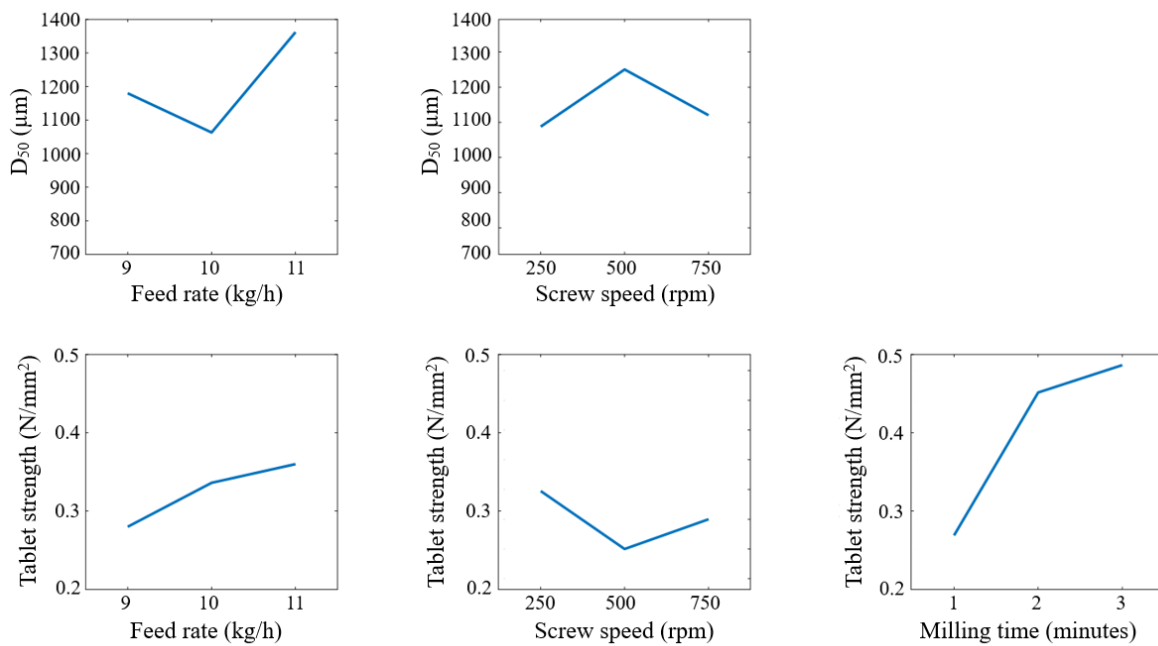


Figure 3. The influences of the three parameters on D_{50} and the tablet strength.

4.2 The Results of the Serial Artificial Neural Networks Framework

In order to develop a data-driven framework based on serial artificial neural networks, the parameters that are investigated and their effects on the downstream processes need to be carefully considered. Figure 4 shows the flowchart of the Consigma25 CM line with its processes and parameters. Since the effects of feed rate, screw speed and milling time on dry granules size and the tablet strength were investigated in this research, two ANN models were established to represent the Consigma25 CM line with these parameters. Such models are also presented in Figure 4. The first model was established to simulate the twin screw granulation process and to predict the granules size which was mapped to the screw speed and the powder feed rate, whereas the second model was developed to map the granules' properties (i.e., granules size) obtained from the first model and the milling time to the tablets strength.

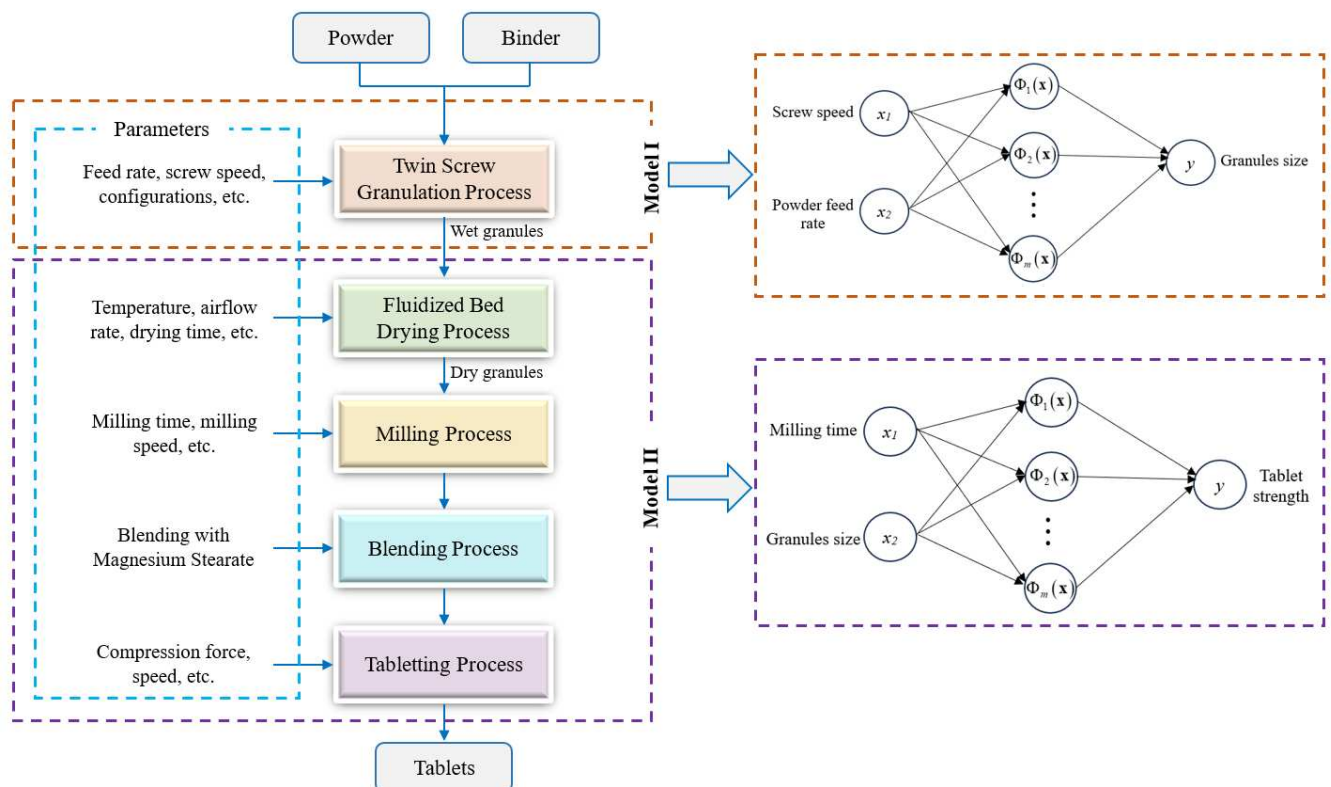


Figure 4. A flowchart of the Consigma25 CM line with its processes and parameters and the models developed to represent them.

In order to develop the first model, the experimental data were split into a training set (70%), which is employed to determine the relationships and tune the network parameters (i.e., the weights connecting neurons from different layers and bias); and a testing set (30%), whose performance is utilized as a stopping criterion to terminate the training when the network generalization capabilities stop improving [4]. Different numbers of the hidden neurons in the range of 1 to 7 were employed, where the number that led to the best network performance in terms of the performance measures and minimum error was chosen. It is worth emphasizing that the values of the network parameters were randomly assigned and they, then, were optimized using the Bayesian algorithm which was embedded with the backpropagation network. Such an optimization algorithm was selected because of its proven ability to deal with small and, perhaps, noisy data sets [23,24]. By using 5 hidden neurons, Figure 5 shows a sample of the predictive performance of the first ANN model developed for the twin screw granulation to predict D_{50} , where the target values represent the experimental values and the predicted values represent the values obtained by the model. The R^2 values for the training and testing sets are 0.86 and 0.82, respectively. The RMSE values for these sets are 98.59 and 111.49 μm , respectively. It is apparent that the R^2 for the testing set was around 4% lower compared to the one for the training set. This can be attributed to the relatively low number of data points in the testing set. In addition, it is also apparent that the testing RMSE is approximately 12% higher than the training RMSE. This is not an indication of overtraining, as such a value can be attributed to the actual D_{50} values that are in the testing set, where all of them are greater than 1100 μm . This can be proved by determining the error percentage values of the data points in the testing set, where it was found that the absolute error percentage values for these data points are in the range of 7.2% to 9.6%, and, thus, they are within the 90% confidence interval. Similarly, two ANN models were developed to predict the D_{10} and D_{90} values, as granule

properties representing the size. All the predicted points in these two models fit nicely around the best fit line with absolute error percentage values in the range of 5.4% to 10.9%.

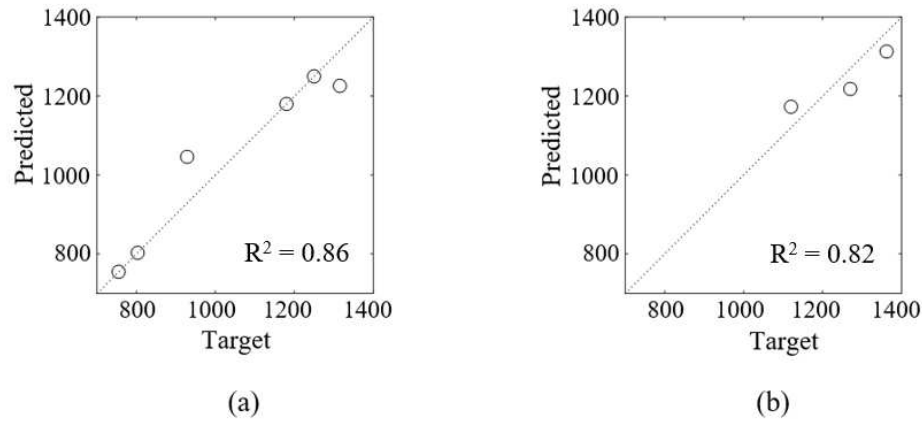


Figure 5. The ANN performance for D_{50} (μm): (a) Training set and (b) Testing set.

Since D_{50} does not represent the volume fractions of the granules obtained, various models should be developed to predict the volume fractions of the granules in the various size classes. Therefore, by using these models, the granules size distribution can be predicted. For this purpose, different ANNs were developed for the different size classes in a way that is similar to the abovementioned procedure. Based on the predictive performance, the optimal numbers of hidden neurons for these ANNs were in the range of 3 to 7. Figure 6 presents an example of the predicted and the actual size distribution of one experiment which was conducted at a powder feed rate of 10kg/h and screw speed of 750rpm. For this experiment, the average error percentage value was 7.93%, where three size classes out of 33 had error percentage values of more than 10%. For all the ANNs developed at this stage, it was noticeable that the predicted value for each class was close to the actual one, where the overall R^2 values for these models were in the range of 0.83 to 0.91. Likewise, the error percentage values for all the models and size classes were in the range of -12.3% to 11.11%.

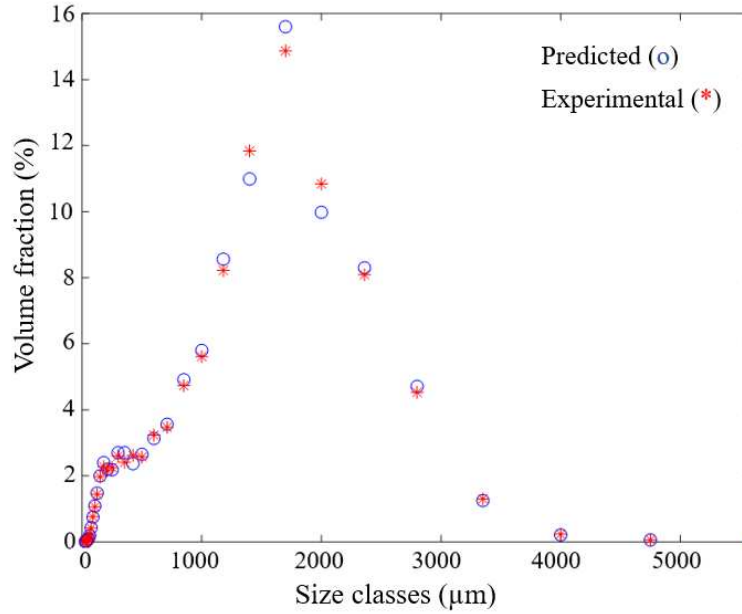


Figure 6. The predicted (○) and the experimental (*) volume fractions (feed rate= 10kg/h and screw speed=750rpm).

Another ANN model was also established to map the granules size, as a granule property, and other process parameters (i.e., milling time) to the tablet strength in order to predict it. In order to avoid the curse of dimensionality that can happen when including the whole size distribution (i.e., 33 classes), the three predicted granules' diameters (i.e., D_{10} , D_{50} and D_{90}) were considered as inputs to this model in addition to the milling time. By using 6 neurons in the hidden layer, Figure 7 shows the predictive performance of the ANN used to predict the tablets strength, where all the predictive points fit nicely around the best fit line. The training and testing R^2 values are 0.92 and 0.90, respectively. It is apparent from Figures 5 and 7 and the predictive performance measures that the predictive performance of the ANN developed to predict the tablet strength is better when compared to the predictive performance values of the ANNs developed to predict the three granules' diameters and the volume fractions of the granules produced by the twin screw granulation process. Such a behaviour can be attributed to (i) the limited number of data points (i.e., experiments) used in this research work; and (ii) the complicated granulation process and its nonlinear input/output relationships [21].

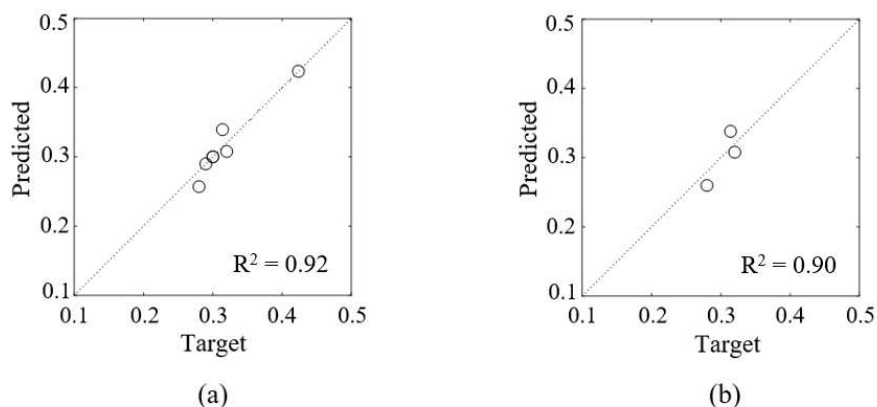


Figure 7. The ANN performance for tablet strength (N/mm^2): (a) Training set and (b) Testing set.

4.3 The Results of the Gaussian Mixture Model

To enhance the performance of the serial ANNs developed in Section 4.2, the error values resulting from these networks were characterized by two GMMs, one for each model. Since the number of inputs is limited in the first model, the feed rate and screw speed in addition to the error resulting from the first ANN were considered in the development of the first GMM. For such a model, the training set was used to train the GMM and identify its best parameters, whereas the testing set was kept hidden. By using 5 components, Figure 8 shows the predictive performance after analysing the error and extracting the hidden information. The R^2 values for the training and testing sets are 0.94 and 0.93, respectively. The RMSE values for these sets are 67.54 and $78.03\mu\text{m}$, respectively. It is apparent that the difference in the R^2 value between the training and the testing sets was reduced after the error characterization by the GMM when compared to the same difference between these sets for the ANN. It is also noticeable that a significant improvement of 9.3% in the R^2 values was obtained. In addition, a significant overall improvement in the RMSE values of 31.6% was also noticeable. Likewise, Figure 8 shows that the data points are, in general, closer to the best fit line when compared to those shown in Figure 5. This can be attributed to the unrepresented information, where such information were successfully extracted by using the GMM. Similarly, the GMM was also employed to characterize the ANNs used to predict the fractions of the granules. By using different optimal numbers of Gaussian components in the range of 2 to 6, the overall improvement in the R^2 was in the range of 5.3% to 21.4%.

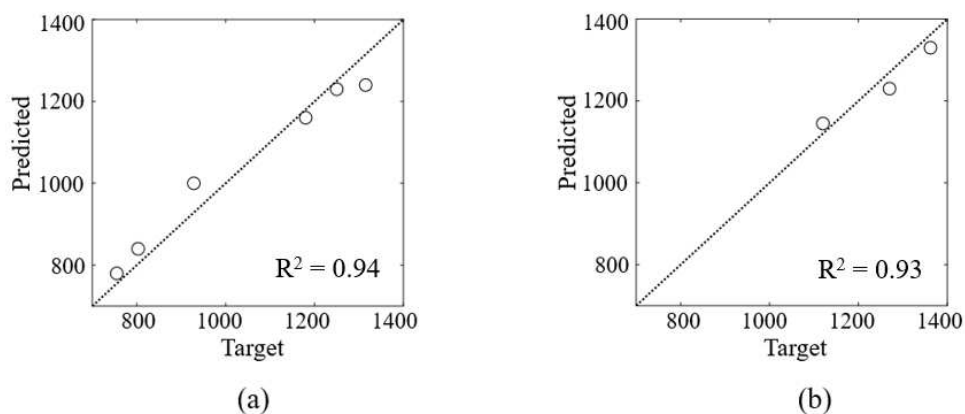


Figure 8. The ANN performance for D_{50} (μm) after the error analyses using the GMM: (a) Training set and (b) Testing set.

Similarly, another GMM was developed to analyse the error resulting from the second ANN established to predict the tablet strength. Since the main parameters' effects were considered in the second ANN, the milling time, the three predicted granules' diameters (i.e., D_{10} , D_{50} and D_{90}) and the error resulting from the second ANN were used as inputs for this model. By using 5 Gaussian components, the predictive performance measures in terms of the R^2 and the RMSE values were improved by 3.3% and 2.4%, respectively. As it was expected the improvement obtained for the first model is more than three times the improvement obtained for the second model. This can be attributed to that the predictive performance of the second ANN model was already better when compared to the first. Likewise, the complexity of the granulation process in addition to the limited number of data points need more complicated (perhaps stochastic) paradigms to be modelled successfully.

In summary, the proposed framework, which is based on serial ANNs characterized by GMMs successfully predicted the granules size and tablet strength produced using the Consigma25 CM line. Such a framework can be considered as a promising development in the pharmaceutical industry. In the future, such a framework can be a roadmap for the development of dynamic as well as continuous modelling paradigms that are based on the wealth of data to

represent all the processes in the CM line, where they can also be utilized to control these processes by exploiting them in a reverse-engineering structure that is embedded with deterministic and stochastic optimization algorithms.

5. Conclusions

In this research project, the Consigma25 Continuous Manufacturing (CM) line was analysed and modelled. The main effects plot was utilized first to examine the nature and the strength of the relationship between various process parameters and the granules size and the tablet strength. A modelling framework based on serial interconnected artificial neural networks (ANNs) was, then, presented to model the unit operations of the Consigma25 CM line. Such interconnected networks were employed to predict the granules size and the tablet strength with an absolute error percentage value of less than 12.3%. In order to improve their performance, Gaussian mixture models (GMMs) were developed to characterize the error resulting from these models by extracting the hidden information. The integration of the GMMs led to an average improvement of 9.7%. Therefore, the modelling framework proposed in this paper can be advantageous to the pharmaceutical industry as it can be used not only to predict the granule and tablet properties but also to systematically control the CM line, in particular when the interpretability of the model is improved. In the future, such a framework can be adapted to be a continuous and dynamic modelling paradigm based on the data available. It can also be incorporated with physical based models and fuzzy logic in order to model the related processes at the micro and macro levels, and to enhance the interpretability and handle process as well as measurement uncertainties.

References

[1] S. Sacher, S. Celikovic, J. Rehr, J. Poms, M. Kirchengast, J. Kruisz, M. Sipek, S. Salar-Behzadi, H. Berger, G. Stark, M. Horn, Towards a novel continuous HME-Tableting line: Process development and control concept, *European Journal of Pharmaceutical Sciences*, 142 (2020) 105097. <https://doi.org/10.1016/j.ejps.2019.105097>.

- [2] M.J. Kreiser, C. Wabel, K.G. Wagner, Direct Tableting on a Continuous Manufacturing Line—Impact of Mixing Parameters, Material Densities, and Drug Load on Subsequent Process Parameters and Tablet Quality. *AAPS PharmSciTech*, 24 (2023) 70. <https://doi.org/10.1208/s12249-023-02525-8>.
- [3] W.H. AlAlaween, M. Mahfouf, A.D. Salman, When swarm meets fuzzy logic: Batch optimisation for the production of pharmaceuticals, *Powder Technology*, 379 (2021) 174-183. <https://doi.org/10.1016/j.powtec.2020.10.066>.
- [4] W.H. AlAlaween, B. Khorsheed, M. Mahfouf, G.K. Reynolds, A.D. Salman, An interpretable fuzzy logic based data-driven model for the twin screw granulation process, *Powder Technology*, 364 (2020) 135-144. <https://doi.org/10.1016/j.powtec.2020.01.052>.
- [5] W.H. AlAlaween, B. Khorsheed, M. Mahfouf, I. Gabbott, G.K. Reynolds, A.D. Salman, Transparent predictive modelling of the twin screw granulation process using a compensated interval type-2 fuzzy system, *European Journal of Pharmaceutics and Biopharmaceutics*, 124 (2018) 138-146. <https://doi.org/10.1016/j.ejpb.2017.12.015>.
- [6] D. Monaco, C. Omar, G.K. Reynolds, P. Tajarobi, J.D. Litster, A.D. Salman, Drying in a continuous wet granulation line: Investigation of different end of drying control methods, *Powder Technology*, 392 (2021) 157-166. <https://doi.org/10.1016/j.powtec.2021.07.004>.
- [7] V. Pauli, F. Elbaz, P. Kleinebudde, M. Krumme, Methodology for a variable rate control strategy development in continuous manufacturing applied to twin-screw wet-granulation and continuous fluid-bed drying, *Journal of Pharmaceutical Innovation*, 13 (2018) 247-260. <https://doi.org/10.1007/s12247-018-9320-6>.
- [8] M. Fonteyne, J. Arruabarrena, J. de Beer, M. Hellings, T. Van Den Kerkhof, A. Burggraeve, C. Vervaet, J.P. Remon, T. De Beer, NIR spectroscopic method for the in-line moisture assessment during drying in a six-segmented fluid bed dryer of a continuous tablet production line: validation of quantifying abilities and uncertainty assessment, *Journal of pharmaceutical and biomedical analysis*, 100 (2014) 21-27. <https://doi.org/10.1016/j.jpba.2014.07.012>.
- [9] J. Rehrl, S. Sacher, M. Horn, J. Khinast, End-point prediction of granule moisture in a ConsiGmaTM-25 segmented fluid bed dryer, *Pharmaceutics*, 12 (2020) 452. <https://doi.org/10.3390/pharmaceutics12050452>.
- [10] F. De Leersnyder, V. Vanhoorne, H. Bekaert, J. Vercruysse, M. Ghijs, N. Bostijn, M. Verstraeten, P. Cappuyns, I. Van Assche, Y. Vander Heyden, E. Ziemons, Breakage and drying behaviour of granules in a continuous fluid bed dryer: Influence of process parameters and wet granule transfer, *European Journal of Pharmaceutical Sciences*, 115 (2018) 223-232. <https://doi.org/10.1016/j.ejps.2018.01.037>.
- [11] M. Fonteyne, J. Vercruysse, F. De Leersnyder, B. Van Snick, C. Vervaet, J.P. Remon, T. De Beer, Process analytical technology for continuous manufacturing of solid-dosage forms, *TrAC Trends in Analytical Chemistry*, 67 (2015) 159-166. <https://doi.org/10.1016/j.trac.2015.01.011>.
- [12] S. Sacher, S. Celikovic, J. Rehrl, J. Poms, M. Kirchengast, J. Kruisz, M. Sipek, S. Salar-Behzadi, H. Berger, G. Stark, M. Horn, Towards a novel continuous HME-Tableting line: Process development and control concept. *European Journal of Pharmaceutical Sciences*, 142 (2020) 105097. <https://doi.org/10.1016/j.ejps.2019.105097>.
- [13] S. Laske, A. Paudel, O. Scheibelhofer, S. Sacher, T. Hoermann, J. Khinast, A. Kelly, J. Rantannen, O. Korhonen, F. Stauffer, F. De Leersnyder, A review of PAT strategies in secondary solid oral dosage manufacturing of small molecules, *Journal of Pharmaceutical Sciences*, 106 (2017) 667-712. <https://doi.org/10.1016/j.xphs.2016.11.011>.
- [14] J.M. Vargas, S. Nielsen, V. Cárdenas, A. Gonzalez, E.Y. Aymat, E. Almodovar, G. Classe, Y. Colón, E. Sanchez, R.J. Romañach, Process analytical technology in continuous manufacturing of a commercial pharmaceutical product, *International journal of pharmaceutics*, 538 (2018) 167-178. <https://doi.org/10.1016/j.ijpharm.2018.01.003>.

- [15] S. Zomer, J. Zhang, S. Talwar, S. Chatteraj, C. Hewitt, Multivariate monitoring for the industrialisation of a continuous wet granulation tableting process, *International Journal of Pharmaceutics*, 547 (2018) 506-519. <https://doi.org/10.1016/j.ijpharm.2018.06.034>.
- [16] J. Rehr, A.P. Karttunen, N. Nicolai, T. Hörmann, M. Horn, O. Korhonen, I. Nopens, T. De Beer, J.G. Khinast, Control of three different continuous pharmaceutical manufacturing processes: Use of soft sensors. *International journal of pharmaceutics*, 543 (2018) 60-72. <https://doi.org/10.1016/j.ijpharm.2018.03.027>.
- [17] Q. Su, S. Ganesh, M. Moreno, Y. Bommireddy, M. Gonzalez, G.V. Reklaitis, Z.K. Nagy, A perspective on Quality-by-Control (QbC) in pharmaceutical continuous manufacturing, *Computers & Chemical Engineering*, 125 (2019) 216-231. <https://doi.org/10.1016/j.compchemeng.2019.03.001>.
- [18] A. Bhaskar, F.N. Barros, R. Singh, Development and implementation of an advanced model predictive control system into continuous pharmaceutical tablet compaction process, *International journal of pharmaceutics*, 534 (2017) 159-178. <https://doi.org/10.1016/j.ijpharm.2017.10.003>.
- [19] W.H. AlAlaween, N.M Faouri, S.H. Al-Omar, E.M. Hendaileh, A.S.S. Almousa, A.H. AlAlawin, O.H. Abdallah, N.T. Albashabsheh, B. Khorsheed, O.A. Abueed, A dynamic nonlinear autoregressive exogenous model for the prediction of COVID-19 cases in Jordan, *Cogent Engineering*, 9 (2022) 2047317. <https://doi.org/10.1080/23311916.2022.2047317>.
- [20] W.H. AlAlaween, O.A. Abueed, A.H. AlAlawin, O.H. Abdallah, N.T. Albashabsheh, E.S. AbdelAll, Y.A. Al-Abdallat, Artificial neural networks for predicting the demand and price of the hybrid electric vehicle spare parts, *Cogent Engineering*, 9 (2022) 2075075. <https://doi.org/10.1080/23311916.2022.2075075>.
- [21] W.H. AlAlaween, M. Mahfouf, A. Salman, Predictive modelling of the granulation process using a systems engineering approach, *Powder Technology*, 302 (2016) 265-274. <https://doi.org/10.1016/j.powtec.2016.08.049>.
- [22] B. Nagy, D.L. Galata, A. Farkas, Z.K. Nagy, Application of artificial neural networks in the process analytical technology of pharmaceutical manufacturing—a review, *The AAPS Journal*, 24 (2022) 74. <https://doi.org/10.1208/s12248-022-00706-0>.
- [23] W.H. AlAlaween, A.H. AlAlawin, M. Mahfouf, O.H. Abdallah, M.A. Shbool, M.F. Mustafa, A new framework for warehouse assessment using a Genetic-Algorithm driven analytic network process, *Plos one* 16 (2021) 0256999. <https://doi.org/10.1371/journal.pone.0256999>.
- [24] C.M. Bishop, N.M. Nasrabadi, *Pattern recognition and machine learning*, 4th ed., Springer, New York, 2006.
- [25] M.I.C. Rachmatullah, J. Santoso, K. Surendro, Determining the number of hidden layer and hidden neuron of neural network for wind speed prediction, *PeerJ Computer Science*, 7 (2021) 724. <https://doi.org/10.7717/peerj-cs.724>
- [26] J.A. Mauricio, Computing and using residuals in time series models, *Computational Statistics & Data Analysis* 52 (2008) 1746-1763. <https://doi.org/10.1016/j.csda.2007.05.034>
- [27] J. Oliveira, W. Pedrycz, *Advances in fuzzy clustering and its applications*, Wiley-Blackwell (2007).
- [28] Y.Y. Yang, M. Mahfouf, G. Panoutsos, Probabilistic Characterisation of Model Error Using Gaussian Mixture Model– with Application to Charpy Impact Energy Prediction for Alloy Steel, *Control Engineering Practice* 20 (2012) 82-92. <https://doi.org/10.1016/j.conengprac.2011.10.001>
- [29] G. J. McLachlan, T. Krishnan, *The EM Algorithm and Extensions*, John Wiley and Sons Inc., Hoboken, New Jersey, 2008.
- [30] J. Antony, *Design of experiments for engineers and scientists*, Elsevier, 2023.