European Journal of Pharmaceutics and Biopharmaceutics 153 (2020) 95–105

Contents lists available at ScienceDirect

European Journal of Pharmaceutics and Biopharmaceutics

journal homepage: www.elsevier.com/locate/ejpb

Deep learning for continuous manufacturing of pharmaceutical solid dosage form

Yves Roggo*,[1], Morgane Jelsch[1], Philipp Heger, Simon Ensslin, Markus Krumme

Novartis Pharma AG, Continuous Manufacturing (CM) Unit, CH-4002 Basel, Switzerland

ARTICLE INFO

Keywords:
Continuous manufacturing
Solid dosage form
Process monitoring
Process analytical technology
Deep learning
Process data science
Process data analytics

ABSTRACT

Continuous Manufacturing (CM) of pharmaceutical drug products is a new approach within the pharmaceutical industry. In the presented paper, a GMP continuous wet granulation line for production of solid dosage forms was investigated. The line was composed of the subsequent continuous unit: operations feeding – twin-screw wet-granulation – fluid-bed drying – sieving and tableting. The formulation of a commercial entity was selected for this study. Several critical process parameters were evaluated in order to probe the process and to characterize the impact on quality attributes. Seven critical process parameters have been selected after a risk analysis: API and excipient mass flows of the two feeders, liquid feed rate and rotation speed of the extruder and rotation speed, temperature and airflow of the dryer. Eight quality attributes were controlled in real time by Process Analytical Technologies (PAT): API content after blender, after dryer, in tablet press feed frame and of tablet, LOD after dryer and PSD after dryer (three PSD parameters: x10 x50 x90). The process parameter values were changed during production in order to detect the impact on the quality of the final product. The deep learning techniques have been used in order to predict the quality attribute (output) with the process parameters (input). The use of deep learning reduces the noise and simplify the data interpretation for a better process understanding. After optimization, three hidden layers neural network were selected with 6 hidden neurons. The activation function ReLU (Rectified Linear Unit) and the ADAM optimizer were used with 2500 epochs (number of learning cycle). API contents, PSD values and LOD values were estimated with an error of calibration lower than 10%. The level of error allow an adequate process monitoring by DNN and we have proven that the main critical process parameters can be identified at a higher levelof process understanding. The synergy between PAT and process data science creates a superior monitoring framework of the continuous manufacturing line and increase the knowledge of this innovative production line and the products that it makes.

1. Introduction [1,2]

Continuous Manufacturing (CM) of pharmaceutical drug products is a new approach within the pharmaceutical industry, opposing traditional batch manufacturing processes based on its potential to increase manufacturing flexibility and efficiency [1]. In CM, all process units are directly connected to each other. Starting material is continuously charged into the first process unit at the beginning of the line, while final product is simultaneously discharged at the end [1,2]. Process control strategy of the CM lines is designed comprising suitable PAT-tools (PAT = process analytical technologies) that deliver real-time information about the process state and the product quality [3,4].

Near Infrared Spectroscopy (NIRS) has become a popular PAT-tool in the pharmaceutical industry, as it is a safe, fast, and non-destructive method, which does not require sample preparation. Nowadays, NIRS is frequently applied for the identification of raw materials, in-line monitoring by compound quantification of process steps like blending, granulation, and drying, as well as for process troubleshooting [5].

Deep learning, especially deep neural network (DNN) is a computer technique inspired from nature [6,7] and have been applied in a large number of fields like computer vision or image analysis [7]. Due to increase of computer power, the complex architecture of the artificial

Abbreviations: ANN, Artificial Neural Network; API, Active Pharmaceutical Ingredient; BU, Blend Uniformity; CM, Continuous Manufacturing; CPP, Critical Process Parameters; CU, Content Uniformity; DNN, Deep Neural Network; FBD, Fluid Bed Dryer; LC, Label Claim; LOD, Loss-On-Drying; NIRS, Near Infrared Spectroscopy; PAT, process analytical technology; PSD, Particle Size Distribution; ReLU, Rectified Linear Unit; RMSEC, Root Mean Square Error of Calibration; RMSEV, Root Mean Square Error of Validation; TSG, Twin-Screw Granulation

* Corresponding author at: Novartis Pharma AG, WSJ-27.4.021.01, Continuous Manufacturing (CM) Unit, CH-4002 Basel, Switzerland.
E-mail address: yves.roggo@novartis.com (Y. Roggo).

[1] The first two authors listed contributed equally to this work.

https://doi.org/10.1016/j.ejpb.2020.06.002
Received 13 March 2020; Received in revised form 25 May 2020; Accepted 2 June 2020
Available online 11 June 2020
0939-6411/

neural network can be trained rapidly. In our study, the possibility to predict quality attributes by DNN is evaluated in order to monitor a pharmaceutical process and to improve the process understanding. Process parameters are used as inputs of the neural networks and the quality parameter are predicted as output of the DNN.

The main objective of this paper is to use DNN for offline process monitoring and to increase the process understanding of the CM production line with online PAT tools and deep learning models. One commercial product was selected and continuous process variations were performed in order to evaluate the impact on the quality attributes (API content, LOD content and PSD) of the continuous process.

2. Materials and methods

2.1. Formulation

Diclofenac Sodium (Acros Organics, Geel, Belgium) was used as Active Pharmaceutical Ingredient (API). The standard formulation contained 25% Diclofenac Sodium, 5% Sodium Starch-Glycolate, 5% Sodium Stearyl Fumarate, 4% Hypromellose (Cellulose-HP-M 603), 12.2% Calcium Hydrogen-Phosphate Anhydrous and 48.8% Microcrystalline Cellulose PH102 (all excipients supplied by Novartis Pharma AG, Stein, Switzerland). To generate granules at varying API content, formulations containing 70–130% of the original drug content were prepared, by adapting the amounts of Hypromellose and Calcium Hydrogen-Phosphate Anhydrous accordingly, while keeping their ratio constant at 80:20. Purified water was used as granulation liquid at a Liquid/Solid-ratio of 0.3. The targeted tablet weight was 240 mg with a targeted label claim (LC) of 60 mg API/tablet (LC = 100%).

2.2. Process

The main process units of the continuous wet granulation line are shown in Fig. 1. The five main process steps are blending, extrusion, drying, tableting and coating.

2.2.1. Preparation of powder blends

Placebo has been pre-blend before the continuous process. One feeder is used for excipients/placebo and a second one for the active ingredient (API). 60 kg of placebo blend was prepared in batch mode by weighing all ingredients into a 100 L container and blending for 10 min at 11 rpm with a Pharma Telescope Blender PTM 300 (LB Bohle GmbH, Ennigerloh, Germany). API and placebo are mixed during the continuous process in a Modulomix continuous modular paddle blender (Hosokawa Micron BV, Doetinchem, The Netherlands).

2.2.2. Twin-screw wet granulation

Continuous wet-granulation was performed on a Thermo Fisher Pharma 16 Twin Screw Granulator (TSG) (Thermo Fisher Scientific, Karlsruhe, Germany) with screw diameter (D) of 16 mm and screw length of 53 ¼ × D. The powder blend was fed into the barrel through the first feeding port by a loss-in-weight powder feeder (K-Tron T20, Coperion K-Tron GmbH, Niederlenz, Switzerland). Granulation liquid at room temperature was fed through a 2.5 mm wide nozzle in the third port by a custom made dispensing pump system (based on Watson Marlow, Zollikon, Switzerland). Five different Liquid to Solid (LS) ratio were tested by varying the liquid feed rate while the solid feed rate was kept constant.

2.2.3. Continuous fluid-bed drying

Continuous fluid-bed drying of wet granules was performed on a Glatt GPCG 2 CM fluid-bed dryer (Glatt GmbH, Binzen, Germany), directly connected to the twin-screw granulator. Drying settings were as follows: 85 °C drying temperature, 90 m^3/h air flow rate, dryer rotation speed 17 rph. Before starting the first trial of the day, or in case the process was stopped in-between, the empty dryer was pre-heated for 120 min at the intended drying temperature and airflow settings.

2.2.4. Sieving and tableting

Dried granules were sieved through a 1.25 mm sieve (Oscillowitt-Lab type MF-lab, Frewitt, Granges-Paccot, Switzerland) directly before compaction in a rotary tablet press (1200i, FETTE Compacting, Schwarzenbek, Germany), with round punches of 12 mm.

2.3. Process target values and variations

Seven critical process parameters (CPPs) were selected for process monitoring, as listed in Table 1. Selection was done based on a previously conducted risk evaluation [8]. Each CPP was monitored by considering the parameters actual process values over time (every 1 s). During the 85 conditions, 148,000 timestamps were recorded at frequency of 1 s. The total time of production was approximately 41 h.

The process factors were investigated by means of systematic univariate variation (step tests) to evaluate their impact on product quality. The trials were performed on 3 consecutive days, each parameter setting was tested for approximately 30 min, which is in line with the system dynamics of the line. Table 1 summarizes the design matrix.

The experiment aimed to investigate the influence of the process factors on drug product's critical quality attributes (CQAs) and other quality characteristics. In detail, intermediate CQAs of the dried granules are API contents (after dryer, in tablet feed frame and of tablets), particle size distribution (PSD) and moisture content (loss-on-drying,

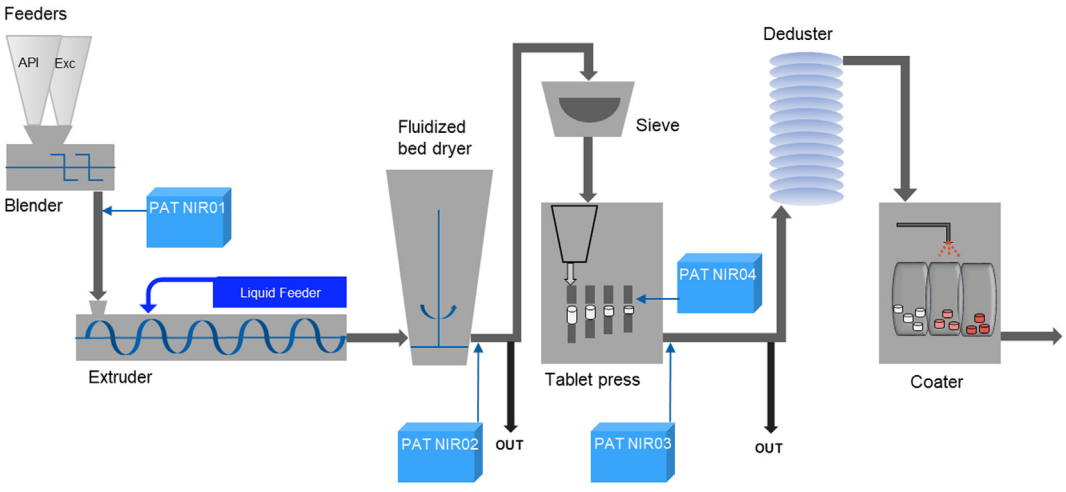

Fig. 1. Continuous wet granulation process and position of the PAT probes.

Table 1
Process parameters and ranges.

Process unit	CPP	Unit	Target value	Steps in % of the target
Feeder	API powder feed rate	kg/h	1.02	+/−10%, +/−5%
	Excipient powder feed rate	kg/h	2.98	+/−10%, +/−5%
Twin screw granulator	liquid feed rate	kg/h	1.05	+/−10%, +/−5%
	granulator screw speed	rpm	450	+/−10%, +/−5%
Dryer	inlet air temperature	°C	85	+/−8.5 °C, +/−4.25 °C
	inlet airflow tower	m^3/h	90	+/−10%, +/−5%
	dryer rotation speed	rph (rotation per hours)	17	+/−20%, +/−10%

Fig. 2. Critical process parameters and the process variations, Axis: x: process time (s) - y: Process values (normalized values – arbitrary unit– moving average 120 s), Legend: in blue calibration set and in green validation set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

LOD).

2.4. Process monitoring

2.4.1. PAT

Four NIRS probes of two different types were installed in-line to analyze dried granules and tablets. A SentroPAT FO instrument (Sentronic®, Dresden, Germany) configured with three immersion probes (SentroProbe DR LS NIR) was used for the monitoring of the blend after the continuous blender, for the monitoring of granules after the fluid-bed dryer (FBD) and of sieved granules in the tablet press feed frame. Spectra were measured in reflection mode using 60 scans of 0.011 s, a 2 nm resolution, and a spectral range of 1150–2200 nm. A VisioNIR LS instrument (VisioTec GmbH, Laupheim, Germany) was used for 100% in-line control of tablet content uniformity. Spectra of tablets were measured in reflection mode in the tablet press using one scan of 0.004 s, an 8 cm^{-1} resolution, and a spectral range of 1100–2100 nm. With both instruments, a background spectrum was acquired before the run for in-line acquisition.

2.4.2. NIR method development and validation

Offline spectra were taken with the probe directly submerged into a

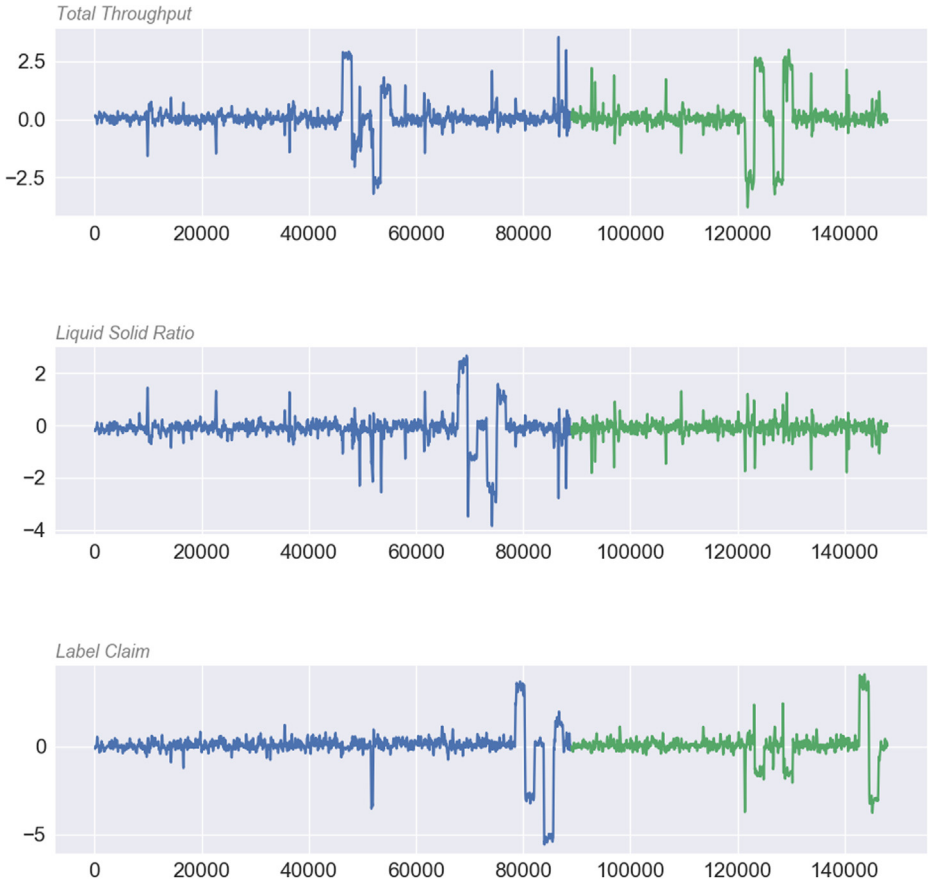

Fig. 3. Computed parameters (based on process values), Axis: x: process time (s) - y: Values (normalized values – arbitrary unit - moving average 120 s), Legend: in blue calibration set and in green validation set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

granule sample at three different positions and five spectra per position, in-line spectra were continuously collected directly in the tablet feed frame during processing. The whole dataset containing 441 offline- and 1809 in-line spectra, was split into two independent calibration and validation sets (221 offline- and 905 in-line spectra for calibration and 220 offline and 904 in-line spectra for validation). The granulation calibration was used after the fluid bed dryer and for the blend uniformity after the blender.

Offline spectra were taken in static mode from 10 tablets per label claim, in-line spectra were taken directly in the tablet press at tableting speeds between 17.000 and 70.000 tablets/hour. The full dataset containing 4500 offline and 6750 in-line spectra, was split into two independent calibration and validation sets, each containing 50% of the overall data. During calibration development it was ensured that calibration and validation spectra were acquired on different days with different batches of powder blends. Partial least squares (PLS) regression was carried out on the preprocessed NIRS spectra with a full cross-validation. The results of the calibration is out of scope of this publication and can be seen in previous publications [9–11].

2.4.3. De-noising and deep learning

- De-noising

Classical methods like moving average [12], Savitzky Golay [13] filter and Discrete Wavelet transform (DTW) [14] were applied on the process data. The moving average was computed on a 120 s time window. The Savitzky Golay filter was applied with a window length of 121, a polynomial order of 3 without derivative. The wavelet filtration has been done with Daubechies function (db) order 1. The reconstruction of the signal is performed with a threshold of 0.01.

- Deep Neural Networks

The DNN were used to predict the quality attributes as output (PSD, LOD, API content) with the seven process parameters as inputs. The following parameters have been applied:

- Activation function: rectified Linear unit (reLU)
- Iteration (Epochs): 2500
- Inputs: 7 process parameters after normalization
- Output: Critical quality attribute – one at a time

All inputs are normalized. One output was used for each neural network. Therefore, the output variables were not normalized. Number of hidden layer, number of hidden neurons by layer and choice of the optimizer is part of the optimization process. In this study, four optimizers were tested : Adaptive Moment Estimation Algorithm ADAM [15,16], Stochastic Gradient Descent SGD [17,18], Adaptive - Gradient Algorithm Adagrad [19,20,21] and Nesterov accelerated Adaptive Moment Estimation Algorithm Nadam [22,23]. Additionally, an internal test set is used (split of 20% of the calibration set) during the training of the model.

2.4.4. Software for computation

SentroSuite package version 2 (Sentronic®, Dresden, Germany) was used for spectra acquisition in the tablet feed frame and after the fluid-bed dryer. The calibration for the Sentronic spectrometer was developed with SIMCA software (version 13.3, Umetrics/Sartorius, Umea, Sweden).

Spectra acquisition of tablets in the tablet press was done with the NovaPAC/NovaMath package (Prozess Technologie Inc. ®, St. Louis, Missouri, USA). Unscrambler® version 10.5 (CAMOs Software AS, Oslo, Norway) was used for the preprocessing and for the PLS computation.

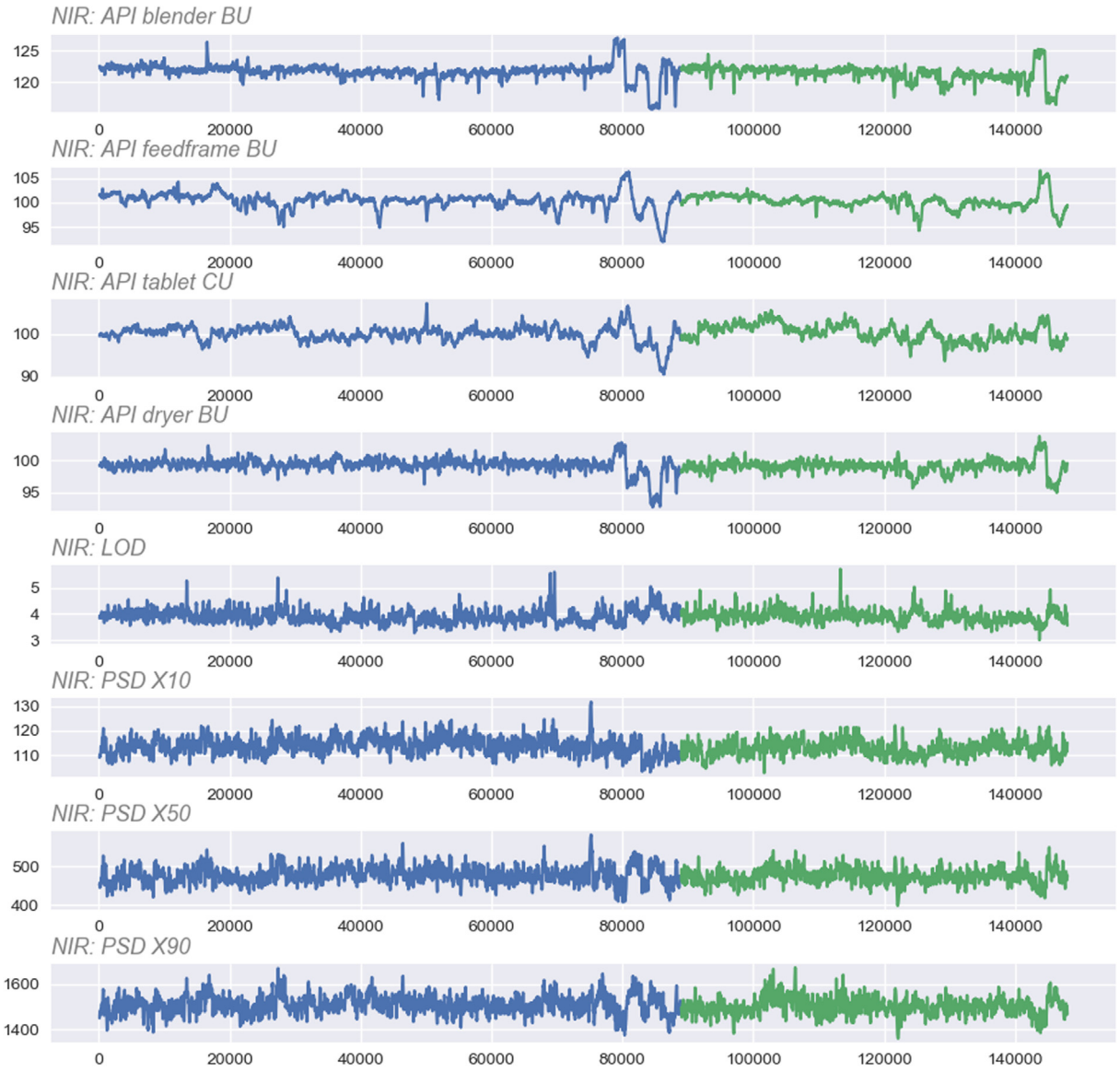

Fig. 4. PAT measured values, Axis: x: process time (s) - y: PAT values (API content in %, LOD in %, PSD in µm – moving average 120 s), Legend: in blue calibration set and in green validation set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Statistics of the PAT values (8 critical quality attributes – std = standard deviation).

Parameters	mean	std	min	25%	50%	75%	max
NIR: API blender BU (%)	121.5	1.5	109.0	120.9	121.7	122.4	129.7
NIR: API feed frame BU (%)	100.3	1.7	90.4	99.6	100.5	101.3	108.4
NIR: API tablet CU (%)	100.1	2.1	88.3	98.8	100.2	101.5	108.4
NIR: API dryer BU (%)	99.1	1.4	86.4	98.5	99.2	99.9	108.7
NIR: LOD (%)	3.9	0.4	1.8	3.6	3.8	4.0	7.8
NIR: PSD X10 (nm)	113.4	5.6	61.5	109.5	112.9	116.7	153.7
NIR: PSD X50 (nm)	477.4	34.0	313.1	454.4	474.5	497.5	743.4
NIR: PSD X90 (nm)	1508.2	64.5	1132.0	1465.5	1506.8	1547.7	1942.4

Hierarchical Calibration development module (version 1.5 - CAMOs Software AS, Oslo, Norway) was used to create the in-line calibration.

Process Data Analytics were computed with Python 3.6.5 (Anaconda Package version 4.6.14, Continuum Analytics) using the Spyder (The Scientific Python Development EnviRonment) 3.2.8 interface. The data structures and analysis tools were provided by Pandas 0.23.0 while the fundamental package for scientific computing with Python was included in Numpy 1.14.3. The Scikit-learn 0.19.1 toolbox enabled the computation of Savitzky Golay filter. The graphics were displayed with Matplotlib 2.2.2 and Seaborn 0.8.1 and the derivative pretreatments with Scipy 0.19.0. Wavelet were available in the pyWT toolbox 0.5.2 and Deep learning was performed with Keras 2.2.4 using Tensorflow backend.

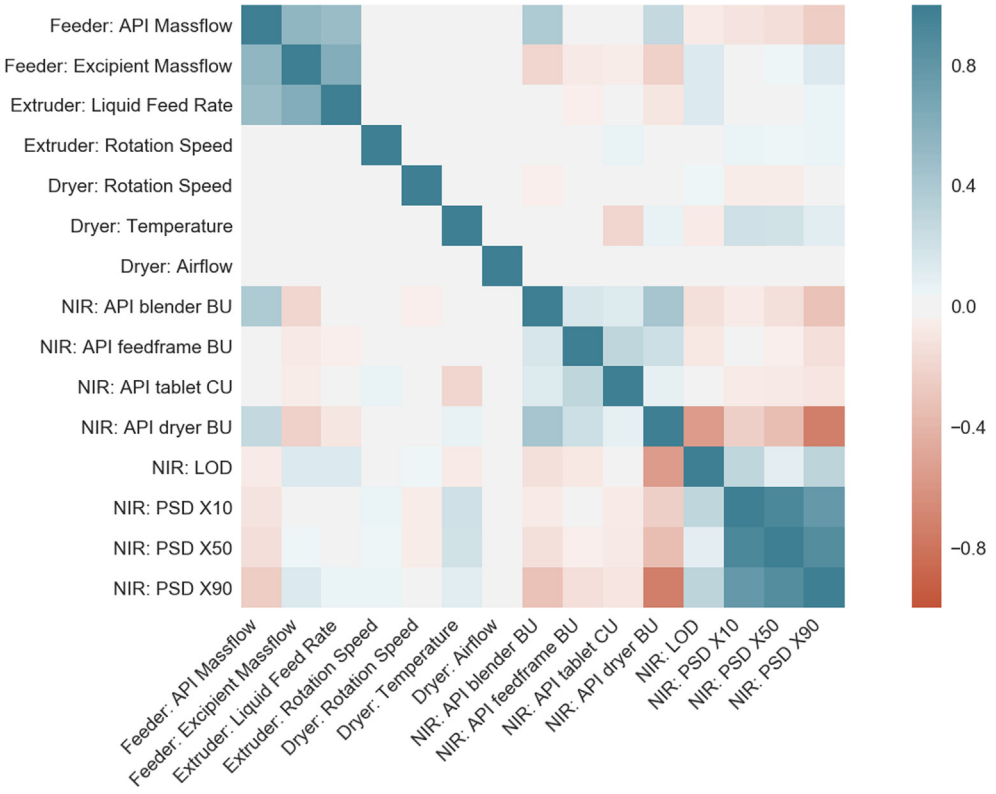

Fig. 5. Correlation map (process variables and PAT values).

3. Results and discussion

3.1. Interpretation of the process values, computed values and PAT monitoring

The Fig. 2 presents the variation of the process parameters and the Fig. 3 shows three computed parameters (Total throughput (kg/h) of the line, liquid/solid ratio (kg/h) and the label claim (% of the target). Concerning the feeder, we can observe small periodic variations (for example 3 spikes between 0 and 40,000 s), this variations are due to the re-fill of the feeder. Nevertheless, these variations are not critical for the product quality. The three computed parameters help the process understanding by translating process values into interpretable variables: throughput, liquid/solid ratio and drug potency (label claim). As remark, we can detect a change in the label claim at the time 120 000 to 130 000 s. Normally no variations were planned. This experiment should be only a change of the total throughput (e.g. a repetition for the validation of the experiment at 40 000 s to 60 0000 – Fig. 3). A computing error (during the setting of the process parameter) leads to these variations. However, data can be used for the DNN calibration.

The variations of the API and excipient mass flow and of the liquid feed rate have been applied in order to observe different total throughput of the line, to modify the label claim (API %) and to change Liquid/Solid ratio of the line. These parameters are quality relevant indicators: API content has a direct impact on patient safety and drug efficacy. LOD and PSD have an influence on the quality of product (e.g. stability and dissolution).

An additional parameter was selected for the extruder (screw speed) to see if this parameter can influence the particle size. Concerning the drying process, the three main parameters (Airflow, temperature and rotation speed) have been evaluated in order to estimate the drying effectiveness. At a given time, only one parameter changed from the target condition. The objective was to identify the most relevant process parameters and their impacts on the API content, water content and PSD.

During the continuous process, the main critical attributes were measured by the PAT sensors. Fig. 4 presents the evaluation of the API content, LOD and PSD values during the production runs and Table 2 gives a statistical overview of the process values. At a first glance, an impact of the feeder API mass flow on the API content measured by NIRS (time stamps 80,000–90,000 s) can be detected. This change has also an impact on the particle size (less API will produce higher particle size) and on the LOD (less API, more water and higher LOD). A simple correlation between variables do not allow a better process understanding (Fig. 5). Feeder mass flows of excipient and of API have a positive correlation, LOD and PSD are correlated as well. Dryer temperature has negative correlation with the LOD and is correlated with PSD (due to a change of LOD value).

Concerning the BU after the blender, The observed average is 121.58%. A systematic error (bias) was detected. The BU prediction of the blend was performed with a calibration developed with granules and not with the blends, which can explain the bias. Nevertheless, the calibration is adequate to detect trending as shown in Fig. 4. Therefore, the values of NIRS after the blender are used in this study without any bias correction.

3.2. Problem statement and use of deep learning

A visual inspection of the values provided by the PAT sensors exhibits a high noise level. The noise is influencing the data interpretation and a simple correlation map cannot be clearly interpreted (Fig. 5). The hypothesis is that the data interpretation can be simplified when the noise is reduced. Classical de-noising methods have been applied: rolling mean, Savitzky Golay filter and Discrete Wavelet transform. However, those methods did not improve the readability and the interpretation of the data in order to detect the most import process parameters (Fig. 6). Therefore, the strategy was to use deep learning in order to improve the understanding of the process. The process variables have been used as X variables (input) and the NIR prediction as Y variable (output). Moreover, after training, the DNN can be used for

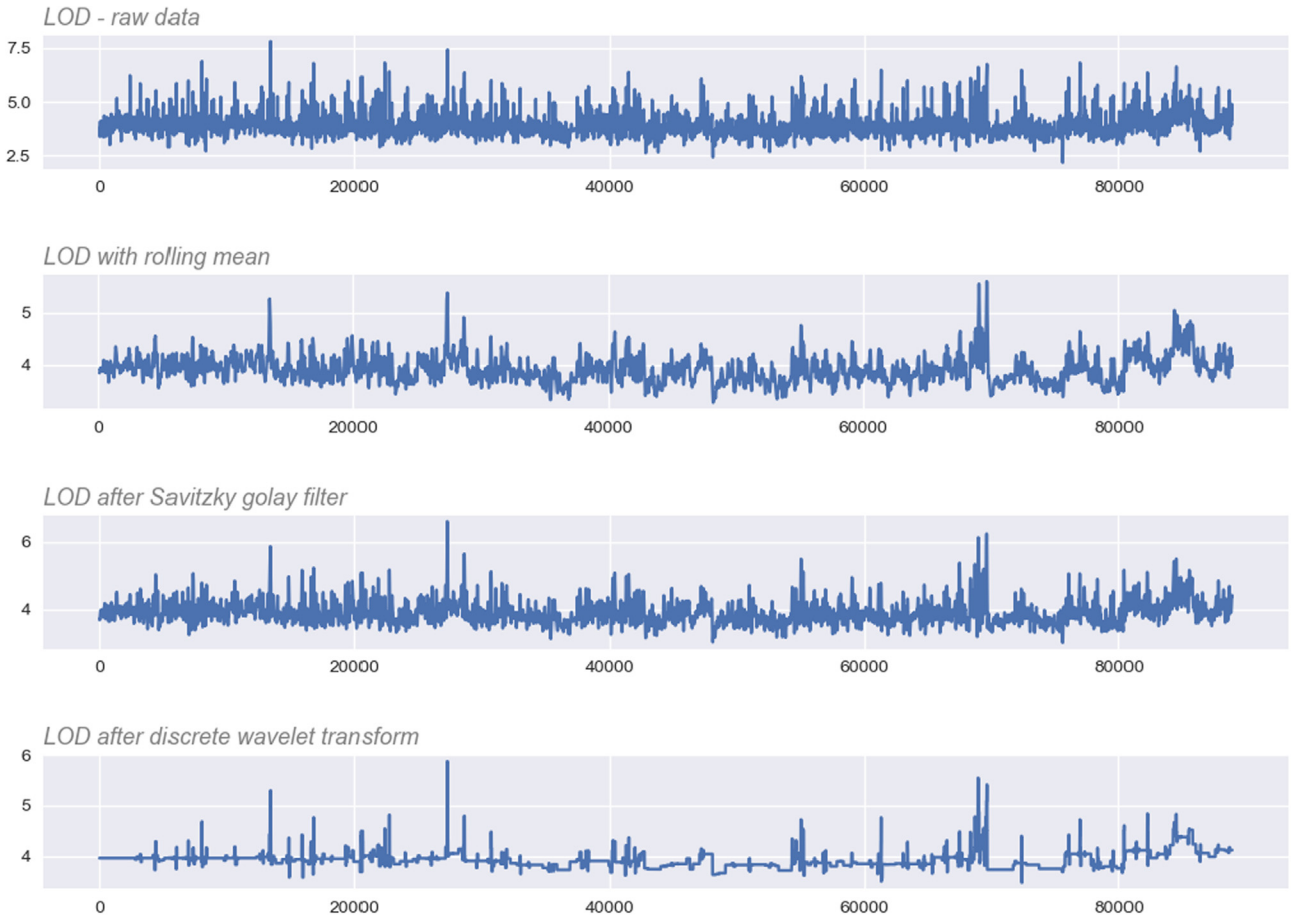

Fig. 6. De-noising of PAT values in the calibration set: moving average (window 120 s), Savitzky Golay and discrete wavelet transform (x: process time in s, y: LOD value in %).

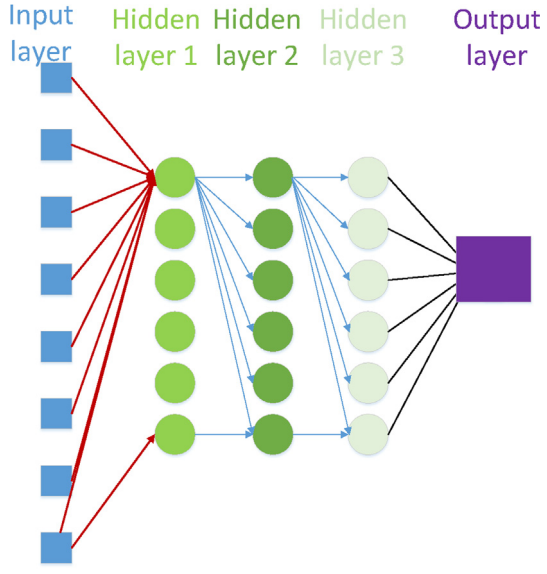

Fig. 7. General schematics of the architecture of the Deep Neural Network (7 inputs, 3 hidden layers with 6 neurons, 1 output) - (remark: only the links of the first neuron are displayed to simplify this figure – All the neurons are fully connected to next layer).

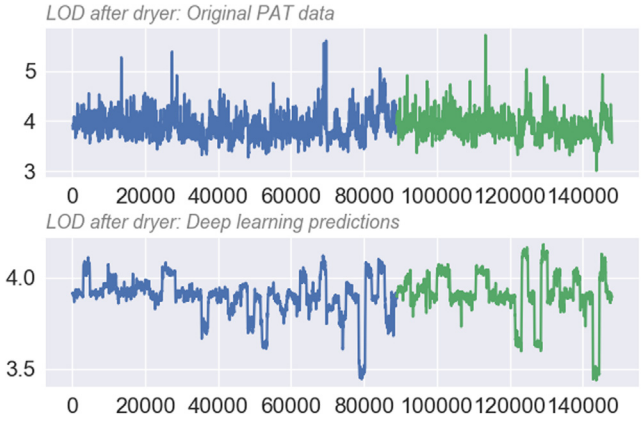

Fig. 8. Qualitative visual comparison - LOD DNN predictions and rolling mean of PAT data (x axis: process time in s, y: LOD value in %).

process monitoring as a redundant information.

3.3. Deep learning optimization

The LOD has been selected for the optimization of the architecture. The input data were centered and reduced and, therefore, the range after normalization can be between [-square root(N), square root (N]. This in fact does not really undermine the results, since the ranges are similar. However, for future applications, other normalization methods should be used in order to have exactly the same range for all the inputs (e.g. Min-Max normalization or normalization on the range [0, 1]).

Table 3
Summary of root mean square errors (RMSE) – 8 parameters – calibration and internal test set).

Parameter	RMSE calibration	RMSE relative calibration	RMSE Test set	RMSE relative Test set
Content after blender	2.8	2.3%	5.0	4.1%
Content after dryer	2.3	2.4%	3.5	3.5%
Content in feedframe	3.3	3.2%	6.1	6.0%
Content of tablets	3.3	3.3%	5.0	5.0%
PSD x10	7.0	6.1%	10.2	8.9%
PSD x50	30.9	6.4%	37.1	7.7%
PSD x90	61.1	4.0%	84.9	5.6%
LOD	0.2	7.0%	0.3	8.9%

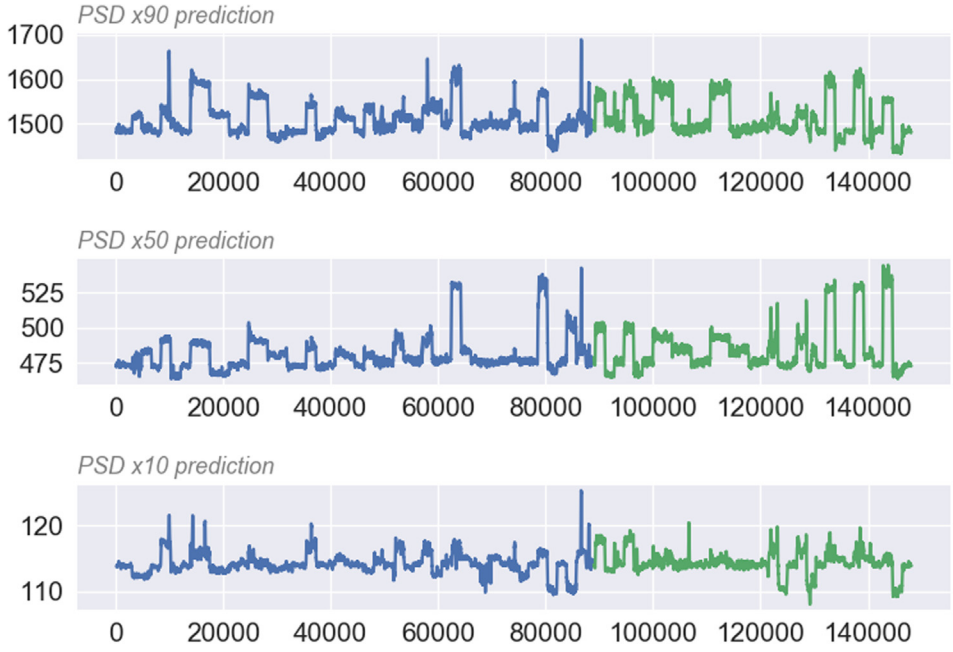

Fig. 9. PSD DNN prediction vs process time (x: process time (s) – y: predicted values PSD value in µm).

Only one output has been used for one DNN while the seven process variables are the inputs of the DNN. The DNN architecture was optimized in two steps: the first one was the selection of the number of hidden layers and the numbers of hidden neurons. The second one was the selection of the optimizer.

The process data have been split into two sets: 0–98 888 s is the calibration set and 98 889–140,000 s is the validation set. The validation is a repetition of the calibration experiments one day later. The process variations are mostly identical in the two data sets. Due to technical issues, the full validation set was not completely recorded: for example, the dryer airflow steps and the last variations of the API mass flow were not repeated in the validation set. Nevertheless, this technical issue has no impact on the quality of the data and on the training of the DNN.

The increase of the number of hidden layers and of neurons increase the speed of the learning (Supplementary material - Fig. 1). The DNN with only one hidden layer have higher RMSE (Root Mean Square Error). Therefore, a deep learning architecture was selected. The complexity of the model was kept low in order to avoid overfitting. In our case, two layers of nine neurons or three layers of six neurons provided the prediction with a reasonably low RMSE. The architecture with three layers of six neurons were selected (Supplementary material - Fig. 2). By further increasing the complexity of the DNN, the RMSE in the internal test set is not improved. The final ANN architecture is presented in Fig. 7 in order to better understand the approach. The optimization of the architecture was done with the LOD. Several architectures for API and LOD have been also tested. However, no significant improvement has been detected. Therefore, the same architecture was applied for all output parameters in order to simplify the strategy in this publication.

Several optimizers have been compared (Supplementary material - Fig. 3). A clear difference in the learning speed can be detected. The convergence for example of the SGD was quick compare to others. If the number of epochs is large, the optimizers achieved similar RMSE. In our case, the computation speed was not critical due to a small data set. The SGD algorithms fail sometime to converge to an optimal solution. Therefore, The ADAM optimizer was chosen as a more robust alternative with still manageable training time and same precision.

3.4. Deep learning results

The results of the selected DNN for the water content prediction is presented in Fig. 8. The similarity between DNN and NIR data can be observed for the LOD predictions. Moreover, the DNN predictions have less noise and therefore the interpretation of trends can easily and visually performed. The response to the process variations can be seen in the results of DNN. The two main objectives have been achieved: on one hand LOD prediction by DNN have a lower noise intensity and on the other hand, DNN prediction seems adequate for process monitoring. The calibration errors (RMSE – Table 3) are relatively low. Moreover, the trends observed during the validation have a significant noise level and the process variation can be clearly detected. The results of the DNN respond typically much faster to the changes in process setpoints than the PAT system. At the same time the noise level is reduced, so there is opportunity to identify the system characteristics with a

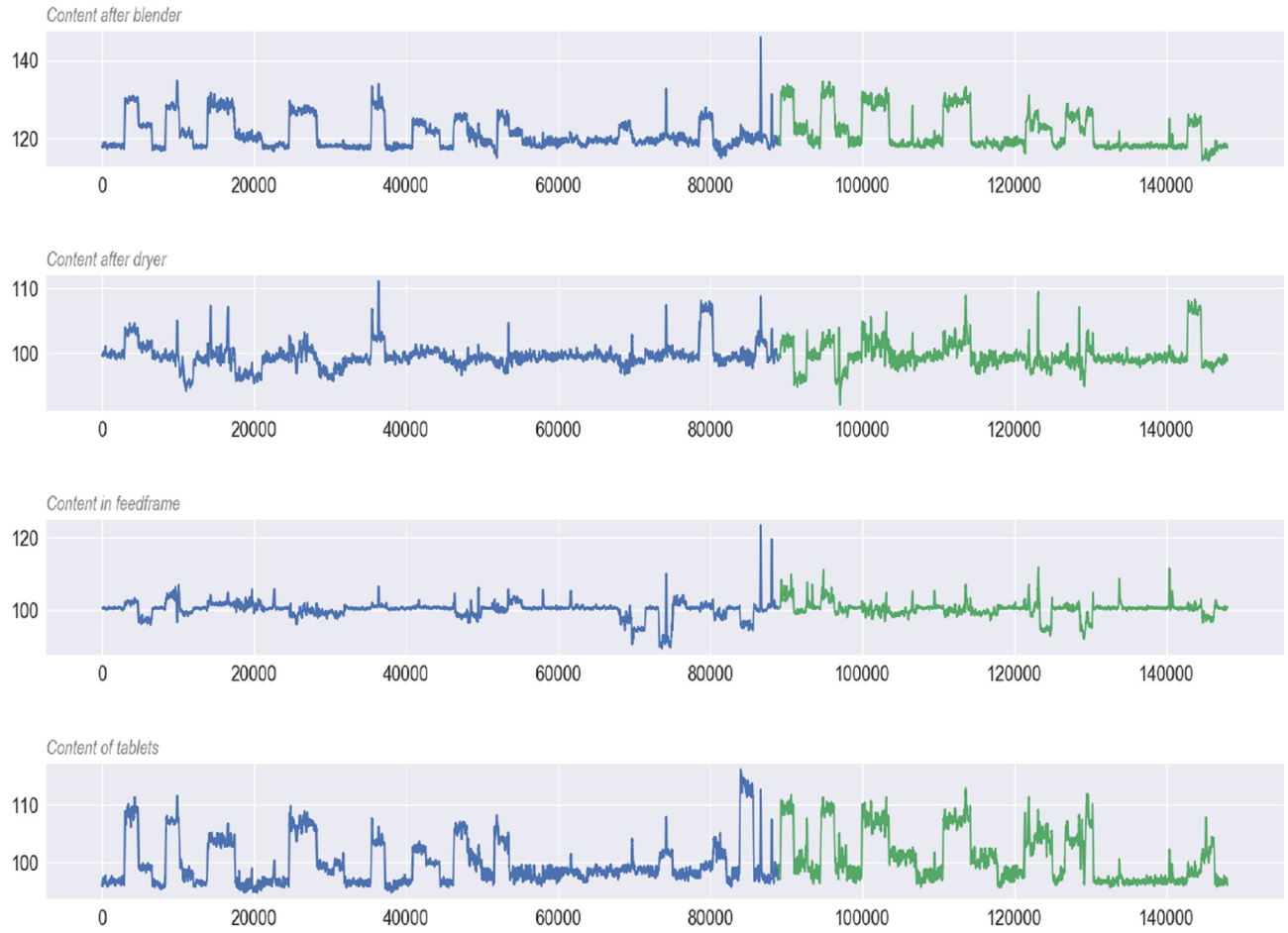

Fig. 10. Content DNN predictions vs process time (x: process time (s) – y: predicted values API content in %).

significantly sharper view than the individual real measured data can make accessible. By its faster response it holds promise to help to keep the process under better control as is shortens cycle times in feedback cycles, which will be investigated further.

Therefore, the DNN calibration can be used with significant promise for online process monitoring. The current comparison of DNN and PAT predictions is only visual. For a fully quantitative comparison of both outputs (DNN and PAT), a validation of the DNN output should be performed. Nevertheless, this DNN validation is not in scope of this publication.

Similarly, DNNs have been created for the prediction contents after blender, after dryer, in tablet freed frame and of tablets (Supplementary material - Fig. 4) and for the 3 PSD values (x10, x50 and x90 – Supplementary material - Fig. 5). For all the networks, the same architecture (3 Hidden layers and 6 hidden neurons by layers) was selected with the ReLU activation functions and ADAM optimizers. An internal test set is used (split of 20% of the calibration set). Relatively low calibration errors have been calculated for the API content and the PSD parameters. The trends obtained with the DNN have a higher signal to noise ratio compared to the PAT values and the process variations can be visually detected earlier. The DNN can be used online for process monitoring in order to predict the API content, the LOD values or the PSD parameters from the process values.

3.5. Critical process parameters

The DNN predictions during the process runs are displayed in Figs. 9 and 10 for the PSD and the API content respectively. Like the LOD results in Fig. 8, the DNN predictions present a significant noise reduction and adequate monitoring is demonstrated. The impact of the process variation can be detected in the DNN predictions and allow the identification of the critical process parameters. A new correlation plot has been computed with the DNN predictions improving the correlation map from PAT data (Fig. 11).

The critical parameters have been divided into two groups: main critical parameters when all the variations have an impact on the quality attribute (Y) and secondary critical parameters when only some variations have impacts.

Solid mass flow and liquid feed rate/mass flow are the main critical parameters for the LOD content. Dryer temperature, airflow and FBD rotation speed are the secondary critical parameters for the LOD content. Concerning the API, the API mass flow and the parameters influencing the LOD (Dryer speed and liquid feed rate) are the primary CPP. The main critical parameters for the PSD are API mass flow and liquid feed rate. A detailed summary is provided in Table 4.

4. Conclusion

A deep learning strategy has been used in order to improve the process understanding and to propose an innovative monitoring strategy of a continuous manufacturing line. During this study, the impact of the process parameters on the three main quality attributes (amount of water, API content and PSD) have been evaluated. The process monitoring by DNN has been demonstrated in-silico. The DNN calibration will be installed later on the production line.

Deep neural networks can learn from noisy PAT values and the ANN predictions can be used to gain sharper views on process understanding and for the more discrimination in the identification of critical process

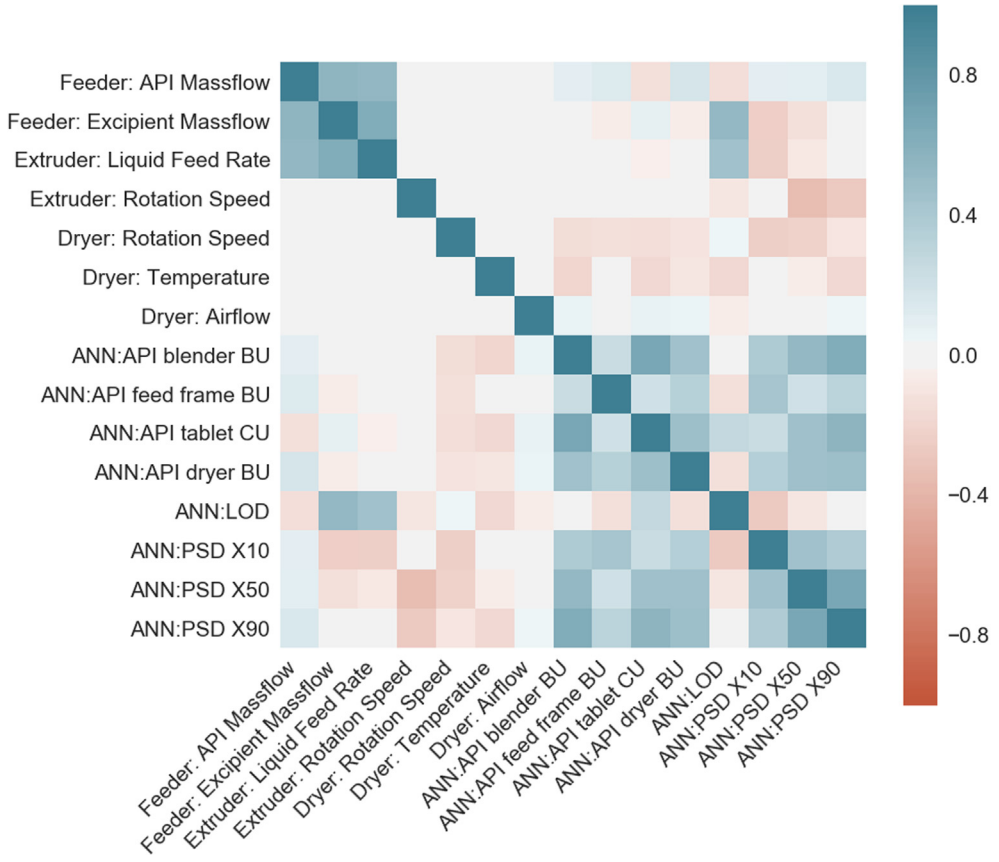

Fig. 11. Correlation map with process variables and DNN predictions.

Table 4
Identification of critical process parameters.

Critical Quality attibute	Primary critical process parameters	Secondary critical process parameters
LOD	API and excipient mass flow Liquid feed rate	Dryer temperature Dryer airflow Dryer rotation speed
API	API and excipient mass flow	Liquid feed rate Dryer rotation speed
PSD	API and excipient mass flow Liquid feed rate	Dryer temperature Dryer air flow Dryer rotation speed

parameters. In this sense they can be useful to effectively “de-noise” PAT data, even though they are not filtering noise out of the raw data but regenerate (simulate) PAT-equivalent data that reconstruct the system response in a sharper and faster way than the direct measurement that we have seen so far. Moreover, DNN can be used for online monitoring as a redundant information. If one of the PAT sensors is broken, the DNN provides information for process monitoring.

The synergy between PAT, deep learning and process data science creates an adequate monitoring framework of the continuous manufacturing line. An outlook of this study will be an extension to support the real time release: online prediction of the final quality of the drug product with deep learning techniques. The challenge with such a kind of purely pattern based learning techniques lies closely next to the opportunities: as there is no mechanistic understanding providing a plausibility check of the predicted values, they can help to distinguish information from random noise more powerful than many other techniques, but at the same time they can produce very uncertain predictions that need comprehensive validation. We found the DNN technique striking and will investigate its robustness in the future.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejpb.2020.06.002.

References

[1] S. Mascia, et al., End-to-end continuous manufacturing of pharmaceuticals: integrated synthesis, purification, and final dosage formation, Angew. Chem. Int. Ed. 52 (47) (2013) 12359–12363.
[2] P. Poechlauer, et al., Continuous processing in the manufacture of active pharmaceutical ingredients and finished dosage forms: an industry perspective, Org. Process Res. Dev. 16 (10) (2012) 1586–1590.
[3] N. Nicolaï, et al., Process control levels for continuous pharmaceutical tablet manufacturing, Chem. Eng. Pharmaceut. Ind.: Drug Product Design, Develop., Model. (2019) 561–584.
[4] M. Fonteyne, et al., Process Analytical Technology for continuous manufacturing of solid-dosage forms, TrAC, Trends Anal. Chem. 67 (2015) 159–166.
[5] Y. Roggo, et al., A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, J. Pharm. Biomed. Anal. 44 (3) (2007) 683–700.
[6] A. Krogh, What are artificial neural networks? Nat. Biotechnol. 26 (2) (2008) 195.
[7] J. Xie, L. Xu, E. Chen, Image denoising and inpainting with deep neural networks, Adv. Neural Informat. Process. Syst. (2012).
[8] V. Pauli, et al., Methodology for a variable rate control strategy development in continuous manufacturing applied to twin-screw wet-granulation and continuous fluid-bed drying, J. Pharmaceut. Innovat. 13 (3) (2018) 247–260.
[9] Y. Roggo, et al., Continuous manufacturing process monitoring of pharmaceutical solid dosage form: A case study, J. Pharm. Biomed. Anal. (2019) 112971.
[10] V. Pauli, et al., Real-time monitoring of particle size distribution in a continuous granulation and drying process by near infrared spectroscopy, Eur. J. Pharm. Biopharm. 141 (2019) 90–99.
[11] V. Pauli, et al., Process analytical technology for continuous manufacturing tableting processing: A case study, J. Pharm. Biomed. Anal. 162 (2019) 101–111.
[12] C.N. Babu, B.E. Reddy, A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data, Appl. Soft Comput. 23 (2014) 27–38.
[13] H. Azami, K. Mohammadi, B. Bozorgtabar, An improved signal segmentation using moving average and Savitzky-Golay filter, J. Signal Informat. Process. 3 (01) (2012) 39.
[14] M. Lang, et al., Noise reduction using an undecimated discrete wavelet transform, IEEE Signal Process Lett. 3 (1) (1996) 10–12.

[15] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
[16] N.S. Keskar, R. Socher, Improving generalization performance by switching from adam to sgd. arXiv preprint arXiv:1712.07628, 2017.
[17] H. Robbins, S. Monro, A stochastic approximation method, Ann. Math. Stat. (1951) 400–407.
[18] S. Ruder, An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747, 2016.
[19] F. Fazayeli, Adaptive subgradient methods for online learning and stochastic optimization, 2014.
[20] J. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, J. Machine Learn. Res. 12 (Jul) (2011) 2121–2159.
[21] F. Zou, L. Shen, On the convergence of adagrad with momentum for training deep neural networks. arXiv preprint arXiv:1808.03408, 2(3) (2018) 5.
[22] T. Dozat, Incorporating nesterov momentum into adam. Workshop track - ICLR 2016, 2016.
[23] J. Wang, Z. Cao, Chinese text sentiment analysis using LSTM network based on L2 and Nadam, in: 2017 IEEE 17th International Conference on Communication Technology (ICCT), IEEE, 2017.