International Journal of Pharmaceutics 605 (2021) 120808

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

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

Powder composition monitoring in continuous pharmaceutical solid-dosage form manufacturing using state estimation – Proof of concept

Francesco Destro [a], Salvador García Muñoz [b], Fabrizio Bezzo [a], Massimiliano Barolo [a,*]

[a] CAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering, University of Padova, via Marzolo 9, 35131 Padova PD, Italy
[b] Synthetic Molecule Design & Development, Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46074, USA

ARTICLE INFO

Keywords:
Continuous pharmaceutical manufacturing
Quality-by-design
Loss-in-weight feeder
Powder feeder
State estimation
Soft sensing

ABSTRACT

In continuous solid-dosage form manufacturing, the powder feeding system is responsible for supplying downstream the correct formulation of the drug product ingredients. The composition of the powder delivered by the feeding system is inferred from the measurements of powder mass flow from the system feeders. The mass flows are, in turn, inferred from the loss in weight measured in the feeder hoppers. Most loss-in-weight feeders post-process the mass flow signal to deliver a smoothed value to the user. However, such estimated mass flows can exhibit a low signal-to-noise ratio. As the feeders are critical elements of the control strategy of the manufacturing line, better instantaneous estimates of mass flow are desirable for improving the quality assurance. In this study, we propose a model-based approach for monitoring the composition of the powder fed to a continuous solid-dosage line. The monitoring system is based on a moving-horizon state estimator, which carries out model-based reconciliation of the feeder mass measurements, thus enabling accurate composition estimation of the powder mixture. Experimental datasets from a direct compression line are used to validate the methodology. Results demonstrate improvement with respect to current industrial solutions.

1. Introduction

Continuous manufacturing is gaining interest in the pharmaceutical community, due to the many advantages with respect to the so-far-dominant batch production mode (Lee et al., 2015; Plumb, 2005). Although many pharmaceutical operations are already carried out continuously, their integration in an end-to-end continuous processing framework still remains challenging. Focusing on solid-dosage manufacturing, an important bottleneck for the transition to continuous processing lays on powder feeding and blending. In the direct compression and direct encapsulation of powders, the powder composition cannot be changed after the feeders. Appropriate estimation and monitoring of the fed composition are therefore fundamental to develop robust control strategies in a continuous line.

Besides the active pharmaceutical ingredient (API), pharmaceutical tablets and capsules include excipients and lubricants. The materials must be fed according to formulations specifically designed for meeting either target physicochemical properties or proper manufacturability (e.g., API dissolution profile and powder flowability). In industrial operations, each material is supplied to the process through a separate loss-in-weight feeder. The powder, stored in a hopper, is gently pushed by a horizontal agitator into a rotating screws system, located below the hopper, from where the material is pushed out of the feeder.

Monitoring feeding operations is a challenging task. Despite recent progress in image analysis for powder composition monitoring (Facco et al., 2017; Galata et al., 2021) and the increasing availability of PAT tools for powder composition assessment along the manufacturing line, these advanced technologies are not established or widespread yet. As a result, the dispensed powder mixture composition is not directly measured in real time in pharmaceutical manufacturing. Instead, the mass composition of a given ingredient (e.g., the API) in the mixture is obtained by calculation, namely by dividing the ingredient mass flow by the sum of the mass flows of all ingredients. The mass flow of each ingredient is itself not measured, but is estimated numerically by the feeder as the ratio between a powder mass weight difference (loss in weight) in the hopper and the time interval across which this difference is measured. This estimated mass flow can exhibit a low signal-to-noise ratio, due to the propagation of noise from the powder net weight measurements (Gyürkés et al., 2020), and is further filtered and smoothed by proprietary algorithms embedded in the equipment. The effects of variations in the API fed concentration can potentially be seen (after axial mixing) through downstream concentration assessment

* Corresponding author.
E-mail address: max.barolo@unipd.it (M. Barolo).

https://doi.org/10.1016/j.ijpharm.2021.120808
Received 8 April 2021; Received in revised form 25 May 2021; Accepted 13 June 2021
Available online 16 June 2021

Nomenclature

Variables

c_{API}^{inst}	fed concentration of API calculated from raw instantaneous mass flow
c_{API}^{HPLC}	post-blending API concentration measured with HPLC on tablets
c_{API}^{tandem}	post-blending API concentration estimated with spectra collected from tablets with a Bruker Tandem IIIA FT-NIR instrument
$c_{API}^{theoretical}$	fed concentration of API calculated from moving-average mass flow
c_{API}^{NIR}	post-blending API concentration measured with NIR on powder mixture entering the feed frame
$\widehat{c}_{API}^{pre-blending}$	pre-blending API concentration obtained by state estimation
$\widehat{c}_{API}^{blended}$	post-blending API concentration obtained by state estimation
DC_i	drive command
F	number of feeders/ingredients
H	window size
k_i	effective density calibration parameter
m_i	powder net weight measurement
$\widehat{m}_i$	powder net weight obtained by state estimation
$\widetilde{m}_i$	model prediction of the powder net weight
$\dot{m}_i^{avg}$	moving-average mass flow
$\dot{m}_i^{inst}$	raw instantaneous mass flow
$\dot{m}_i^{refill}$	mass flow of powder refilling the hopper
$\dot{m}_i^{smooth}$	smoothed mass flow
$\widehat{\dot{m}}_i$	mass flow obtained by state estimation
$\widetilde{\dot{m}}_i$	model prediction of the mass flow
N_i	screw rotation speed
N_{max}	maximum rotation speed
n	number of starts of the screw thread
n_{exp}	number of experiments in a dataset
n_x	number of system states
n_u	number of system inputs
n_y	number of system measurements
P	screw pitch
T	receding horizon
t	time
t_k	discrete time step
t_f	final time step

Vectors and matrices

$\mathbf{F}$	continuous-time system model vector function
$\mathbf{f}$	discrete-time system model vector function
$\mathbf{g}$	measurement model vector function
$\mathbf{Q}$	model error covariance matrix
$\mathbf{R}$	measurement noise covariance matrix
$\mathbf{u}$	inputs vector
$\mathbf{v}$	measurement error vector
$\mathbf{w}$	model error vector
$\widehat{\mathbf{x}}$	states vector obtained by state estimation
$\widetilde{\mathbf{x}}$	model prediction of the states vector
$\mathbf{y}$	measurements vector
$\widetilde{\mathbf{y}}$	model prediction of the measurements vector

Greek symbols

η	feeder volumetric efficiency
$\theta_{blending}$	blending dead time
θ_i	vector of estimated parameters
ρ_i	powder effective density
ρ_i^0	stress-free powder density
$\sigma_{m_i}^2$	variance of the mass measurement
σ_V	vertical stress in the hopper
$\tau_{blending}$	blending time constant
τ_i	feeder time constant
Φ_{MLE}	maximum likelihood function

Subscript

i	feeder/ingredient number

Acronyms

API	active pharmaceutical ingredient
EIOT	extended iterative optimization technology
FOPDT	first-order-plus-dead-time
HPLC	high performance liquid chromatography
MHE	moving-horizon estimator
NIR	near-infrared spectroscopy
PAT	process analytical technology
SSE	sum of squared errors

(García-Muñoz et al., 2017) obtained by high-performance liquid chromatography (HPLC) or spectroscopic measurements (De Beer et al., 2011). However, as will be discussed in detail in Section 2, the dynamics of the variation of the API fed concentration, as calculated from the mass flow estimations, is not always in good agreement with the one assessed downstream. This issue is even more relevant with cohesive powders that are difficult to feed, and for which the net weight measurements are very noisy.

Better instantaneous estimation of feeder powder mass flows is desirable for improving the control strategy of a solid-dosage manufacturing line. Within the Quality-by-Design (Fisher et al., 2016; Food & Drug Administration, 2004) framework, practitioners are encouraged to modernize pharmaceutical manufacturing through the adoption of control strategies relying on risk analysis and solid process understanding. Fully understanding how the variability of process inputs (raw material properties and process operating conditions) impacts the process outputs (product critical quality attributes) is of paramount importance. In this perspective, improving the estimation of feeder powder mass flow is critical to implementing Quality-by-Design in solid-dosage manufacturing. However, this monitoring problem is largely unexplored in the literature, and industrial practice leaves this aspect up to the equipment vendors.

Recent studies on pharmaceutical powder feeding and blending design, monitoring and control for solid-dosage manufacturing are comprehensively reviewed elsewhere (Su et al., 2020). Several contributions address the selection of suitable feeder configurations (e.g., hopper capacity, screws characteristics) for given powder systems and process requirements (Blackshields and Crean, 2018; Cartwright et al., 2013; Engisch and Muzzio, 2012). From the design space description side, García-Muñoz et al. (2017) assessed the impact of the feeders mass flow variability onto the drug product composition in a continuous direct compression process using a mathematical model for the process. Other studies focus on the improvement of feedback and feedforward control. Hanson (2018) demonstrated the advantages of introducing a ratio controller in the feeding system, linking the set-points of the mass flow controllers to the current API powder mass flow. Model predictive control of a feeding-blending system in a continuous pharmaceutical tableting plant was reported by Singh et al. (2014). The authors showed

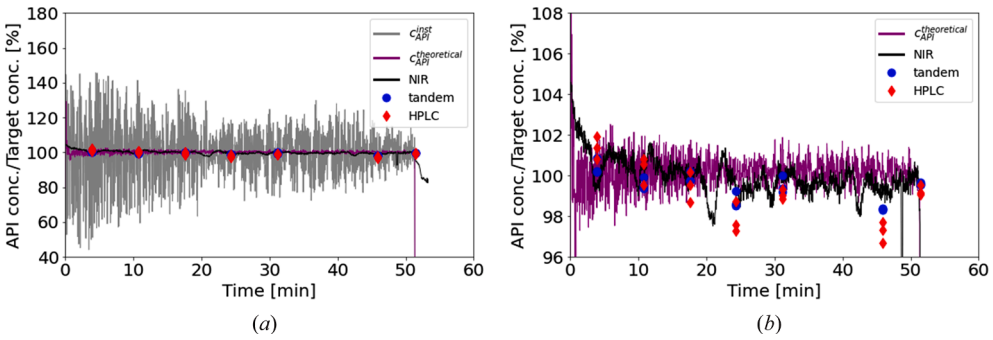

Fig. 1. Time profiles of the ratio between actual API concentrations and target API concentrations from a direct compression line (data from Dataset A discussed in Section 3). (*a*) Full view; (*b*) zoom-in. The concentration values are determined according to different methods: *i*) c_{API}^{inst}; *ii*) $c_{API}^{theoretical}$; *iii*) assessed on the powder mixture that enters the tablet press feed frame using near-infrared (NIR) spectroscopy; *iv*) assessed on the final tablets using a Bruker Tandem IIIA FT-NIR instrument (Bruker Corp. Germany); *v*) assessed on the final tablets using HPLC.

improved performance in the control of the API content in the tablets for servo problems, with respect to traditional proportional-integral-derivative control. Nonetheless, fluctuations in the API concentration in the products were still observed, mostly because of the propagation of measurement inaccuracies into the control loop. In Rehrl et al. (2016), the benefits of model predictive control implementation in a blending system are further outlined through a simulation study. The authors point out that, for applications in physical units, it is necessary to put in place a state estimator, to provide accurate estimations of unmeasured state variables (states), such as mass flows and composition.

State estimation (Ray, 1981) is a family of mathematical techniques that, given a mathematical model of a process and real time measurements from a plant, estimate the unmeasured system states and provide noise-attenuated estimations of the actual measurements. Moving-horizon estimation (Rao, 2000; Rao and Rawlings, 2002) estimates the current system states with an optimization-based strategy. For nonlinear systems, moving-horizon estimation is known to outperform alternative state estimation techniques in many applications (Haseltine and Rawlings, 2005).

Very few studies address state estimation in solid-dosage form manufacturing, despite the great advantages it can bring. Recently, Liu et al. (2018) tested a robust moving-horizon estimator (MHE) on a pharmaceutical feeding-blending system. They conducted an *in-silico* analysis under the ideal assumption of no plant/model mismatch. The state variables relevant to the process (powder hold-ups in feeders and blenders, powder mass flows and powder mixture composition across the system) were successfully reconstructed by the estimator, also in the presence of gross measurement errors, such as outliers and drifts.

In this study, we apply state estimation techniques to obtain a much improved instantaneous estimate of the mass flow of material being dispensed by a screw feeder in a continuous line. We develop a novel MHE-based powder composition monitoring scheme for the feeding and blending section of a real continuous tableting line, following a Quality-by-Design approach rooted in enhanced process understanding. The powder composition monitoring scheme developed encompasses the system from feeder to powder in the die, and does not include the compaction system in the tablet press. The process, detailed elsewhere (García-Muñoz et al., 2017), is made up of three sections: feeding, blending, and compression. An MHE is implemented in the feeding section, based on a first-principles mathematical model describing loss-in-weight feeder operation (Bascone et al., 2020). The MHE estimates the outlet powder mass flow from each feeder using a model of the feeders and measurements of the motor speeds and load cell masses. This in turn enables estimating the API concentration fed downstream, so that a comparison with actual measurements is possible. The intermediate blending effect is accounted for through an additional mathematical model calibrated with tracing experiments or step change experiments. We benchmark the proposed monitoring approach against a typical industrial one that uses API concentration values calculated from the mass flow estimations as obtained from the software embedded in the feeders.

The remainder of the manuscript is organized as follows. The problem is stated formally in Section 2. Section 3 describes the experimental setup and provides mathematical background to MHE and feeders modeling. The workflow for implementing the proposed monitoring system on the direct compression line is discussed in Section 4. Realtime monitoring of the line is demonstrated in Section 5, before of the concluding remarks.

2. The need for a better feeder mass flow estimation

Let us consider a direct compression line made by F feeders, each one providing a different ingredient to the powder mixture. The mass flow estimations are provided to the user by the software interface of feeder i ($i = 1, \ldots, F$). Namely, the following estimation signals are available for our particular case study:

- $\dot{m}_i^{inst}$, the raw instantaneous mass flow calculated as ratio between the difference of two consecutive measurements of powder mass weight in the hopper (loss in weight) and the difference between their respective sampling times;
- $\dot{m}_i^{avg}$, a smoothed version of $\dot{m}_i^{inst}$, calculated using a thirty-second moving average;
- $\dot{m}_i^{smooth}$, an additional smoothed version of $\dot{m}_i^{inst}$, obtained by a strong smoothing of $\dot{m}_i^{inst}$ through proprietary algorithms embedded in the equipment.

The API theoretical fed concentration $c_{API}^{theoretical}$ is calculated for this case study using the moving-average mass flow signals, according to:

$$c_{API}^{theoretical} = \frac{\dot{m}_i^{avg}}{\sum_{i=1}^{F} \dot{m}_i^{avg}} \tag{1}$$

where the API is fed from the first feeder ($i = 1$). For comparison, we define c_{API}^{inst} as the API concentration calculated from the raw instantaneous signals:

$$c_{API}^{inst} = \frac{\dot{m}_i^{inst}}{\sum_{i=1}^{F} \dot{m}_i^{inst}} \tag{2}$$

Additionally, spectroscopic and HPLC measurements are taken downstream, on both the powder mixture entering the tablet press feed frame and the final tablets. It was observed that certain downstream powder composition variations measured with spectroscopic tools and HPLC are not registered by c_{API}^{inst}, which is a quite noisy signal (Fig. 1a). On the other hand, the API concentration calculated from $\dot{m}_i^{smooth}$ (not shown in Fig. 1) cannot fully capture the composition variation dynamics due to oversmoothing. For this reason, a moving average mass flow $\dot{m}_i^{avg}$ can be carefully configured to make this dynamics evident upstream. Nonetheless, the API fed concentration inferred from $\dot{m}_i^{avg}$ (i.e., $c_{API}^{theoretical}$) still shows a low signal-to-noise ratio compared to the direct composition assessment, and lags behind fast-changing variations, as a

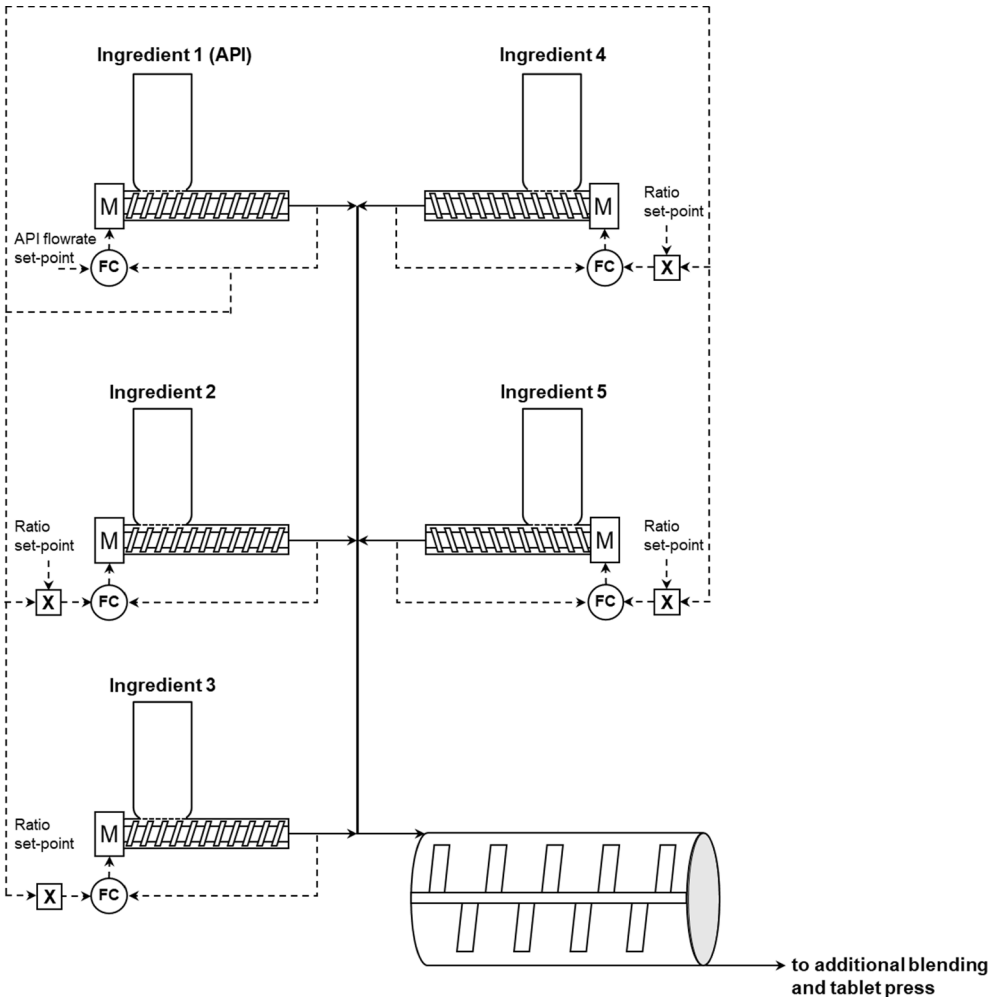

Fig. 2. Schematic of the direct compression line used for data collection. Five loss-in-weight feeders operate in parallel, sending the powder formulation to the horizontal mixer. Additional powder blending occurs in the feed frame of the tablet press, before the powder is fed in the die. The blending model in this work encompasses all the mixing occurring before the powder enters the die.

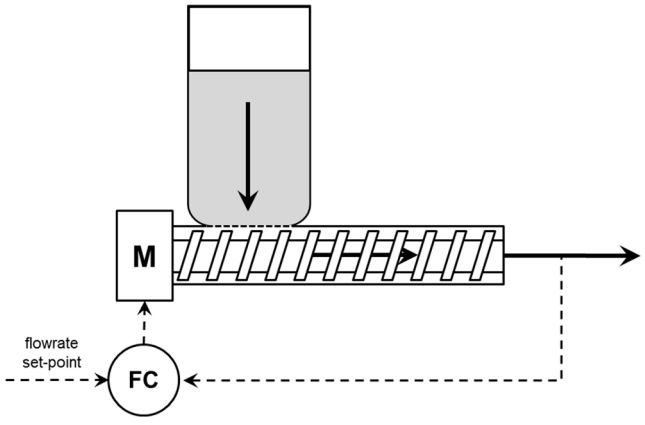

Fig. 3. Schematic of a loss-in-weight feeder. The powder stored in the hopper, aided by an agitator (not shown), flows into the twin-screw system, from where it is pushed out of the feeder. A load cell (not shown) measures the mass of powder in the unit. From the mass measurements, a powder mass flow measurement is inferred and used by the flow controller for adjusting the motor (M) speed, to track the reference mass flow setpoint.

consequence of the moving-average smoothing (Fig. 1b).

Better instantaneous estimates of mass flow for improved quality assurance are therefore desirable, as the feeders are critical elements of the control strategy.

3. Materials and methods

3.1. Process description and experimental data collection

The experimental data for designing and validating the proposed monitoring methodology were collected in a continuous direct compression pilot plant for a molecule in development (García-Muñoz et al., 2017; Fig. 2). Five ($F = 5$) loss-in-weight feeders (Coperion K-Tron KT20) dispense the API and the other four ingredients of the formulation to the downstream continuous horizontal mixer. The powder is then discharged into a vertical hopper and, after passing through a transition piece, enters the feed frame of the rotary tablet press. All operations occurring in between the feeders and the press feed frame form the blending section of the process. Powder blending is necessary for mixing the formulation ingredients, and allows partially compensating for feeder disturbances. Powder blending is known to occur not only in the horizontal mixer but in the tablet press feed frame. The specific details about the formulation are irrelevant for the study and are withheld for confidentiality.

The loss-in-weight feeders (Fig. 3) are made up of a motor-driven

Table 1
List of datasets collected during continuous runs of the direct compression line.

	Final tablet mass [mg]	API target concentration in tablet [%]	Experiment duration [min]
Dataset A	50	40	55
Dataset B	150	40	35
Dataset C	200	40	35

screw system, an agitator, a hopper where the material is stored, and a load cell installed below the unit. The measurements provided by the feeder and used by the proposed monitoring system as if they were available in real time are: *i*) the net weight m_i of material in feeder, and *ii*) the motor drive command DC_i, ranging from 0 to 100%.

The feeders measurements DC_i and m_i and the three aforementioned (Section 2) mass flow estimations $\dot{m}_i^{inst}$, $\dot{m}_i^{avg}$ and $\dot{m}_i^{smooth}$ are updated by the internal feeder software through high-frequency readings, but they are all made available to the user only every 2 s. The feeders can operate at open loop (volumetric mode) or at closed loop (gravimetric mode). When operating in closed-loop mode, the internal mass flow controller uses the high-frequency measurements (not accessible to the user) to keep the mass flow at the relevant setpoint, by adjusting DC_i. A ratio controller (Hanson, 2018) continuously receives the high-frequency API mass flow estimation and, based on that, updates the set points of the other flow controllers, in such a way as to enforce the target concentration for each ingredient.

The API concentration in the powder mixture in the feed frame was estimated through near-infrared (NIR) spectroscopy. Extended iterative optimization technology (EIOT), a recently-developed methodology (Shi et al., 2019) was used to correlate the NIR signal to the API concentration, here denoted as c_{API}^{NIR}, using optimization to enforce a modified Beer-Lambert law to the spectra. Discrete API concentration measurements were also obtained on tablets sampled at the press outlet. HPLC and concentration estimations derived from data collected with a Bruker Tandem IIIA FT-NIR instrument (Bruker Corp. Germany) were used to this purpose, and the relevant measurements are indicated with c_{API}^{HPLC} and c_{API}^{tandem}, respectively. While c_{API}^{NIR} was made available to the user every 2 s, only few measurements of c_{API}^{HPLC} and c_{API}^{tandem} were collected during the operation.

In this study, three datasets (Table 1) were collected during closed-loop continuous runs (without hopper refills) for a molecule in development in a pilot plant facility. This data was used for calibration and validation of the monitoring system. The profiles of $\dot{m}_1^{avg}$ and DC_1 for Dataset A in feeder 1 are shown in Fig. 4 to illustrate how a typical closed-loop operation evolves. As time progresses, powder densification in the screws decreases. To maintain $\dot{m}_1^{avg}$ close to the set-point of 6.5 kg/h (Fig. 4a), the controller progressively increases DC_1 (Fig. 4b), thus increasing the number of revolutions per minute of the feeder screw.

All datasets in Table 1 refer to the same tablet formulation. Although the final mass of the tablets obtained varies across datasets, the feeding and blending sections always operate in the same way. The set of time-dependent variables characterizing each dataset is summarized in Table 2. As will be detailed later, DC_i and m_i are the only measurements required by the proposed monitoring system; the other variables are used for validating and benchmarking it. Additional real-time measurements available through the feeder software interface, but not meaningful for process monitoring, are not discussed here.

3.2. State estimation: moving-horizon estimation of state variables

Let us consider the mathematical model of a process. The model predicts the time evolution of the system states according to:

$$\frac{d\tilde{\mathbf{x}}(t)}{dt} = \mathbf{F}(\tilde{\mathbf{x}}(t),\ \mathbf{u}(t)) \tag{3}$$

where $\tilde{\mathbf{x}} \in \mathbb{R}^{n_x}$ is the vector of the n_x predicted states, $\mathbf{u} \in \mathbb{R}^{n_u}$ is the vector of the n_u system inputs, $\mathrm{F}(\cdot)$: $\mathbb{R}^{n_x+n_u} \mapsto \mathbb{R}^{n_x}$ is the vector function representing the model, and t is the time. The system states may be measured or not. Note that in the remainder of the manuscript, the tilde symbol "~" on top of a variable denotes that the variable has been calculated by means of a model prediction.

The continuous-time model (3) can be expressed in discrete-time

Table 2
List of time-dependent variables characterizing each dataset. A comprehensive list of symbols is reported as a separate section.

Symbol	Type	Variable	Measurement unit
c_{API}^{HPLC}	Measured	API concentration measured with HPLC on tablets	[kg/kg]
c_{API}^{NIR}	Measured	API concentration measured with NIR on powder mixture entering the feed frame	[kg/kg]
c_{API}^{tandem}	Measured	API concentration estimated with spectra collected from tablets with a Bruker Tandem IIIA FT-NIR instrument	[kg/kg]
c_{API}^{inst}	Calculated	Fed concentration of API calculated from $\dot{m}_i^{inst}$ (Equation (2))	[kg/kg]
$c_{API}^{theoretical}$	Calculated	Theoretical fed concentration of API (Equation (1))	[kg/kg]
DC_i	Measured	Motor drive command of feeder i	[%]
m_i	Measured	Powder mass in feeder i	[kg]
$\dot{m}_i^{inst}$	Calculated	Mass flow from feeder i obtained as rate of change of measured mass (raw instantaneous mass flow)	[kg/h]
$\dot{m}_i^{avg}$	Calculated	Mass flow from feeder i obtained as thirty-seconds moving-average of $\dot{m}_i^{inst}$	[kg/h]
$\dot{m}_i^{smooth}$	Calculated	Mass flow from feeder i obtained by smoothing $\dot{m}_i^{inst}$ with a statistical filter	[kg/h]

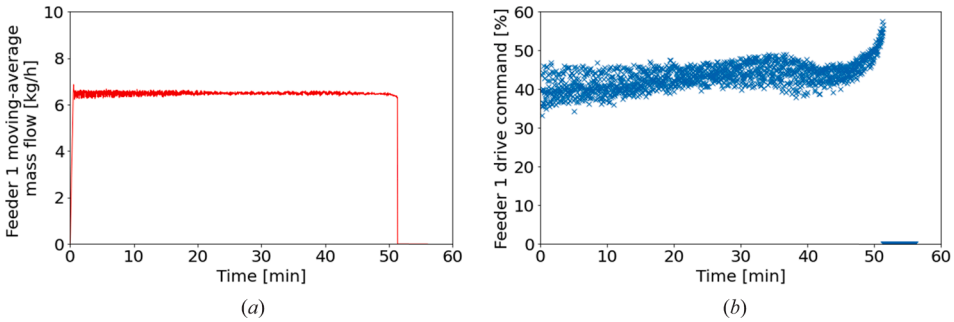

Fig. 4. Dataset A, feeder 1: time profiles of (*a*) $\dot{m}_{1,avg}$ and (*b*) DC_1. The built-in feeder flow controller manipulates DC_1 to keep the mass flow at its set-point. While time progresses during a run, powder densification in the screw decreases, due to the smaller hold-up in the hopper. The controller responds by increasing DC_1 (hence the number of revolutions per minute of the screw).

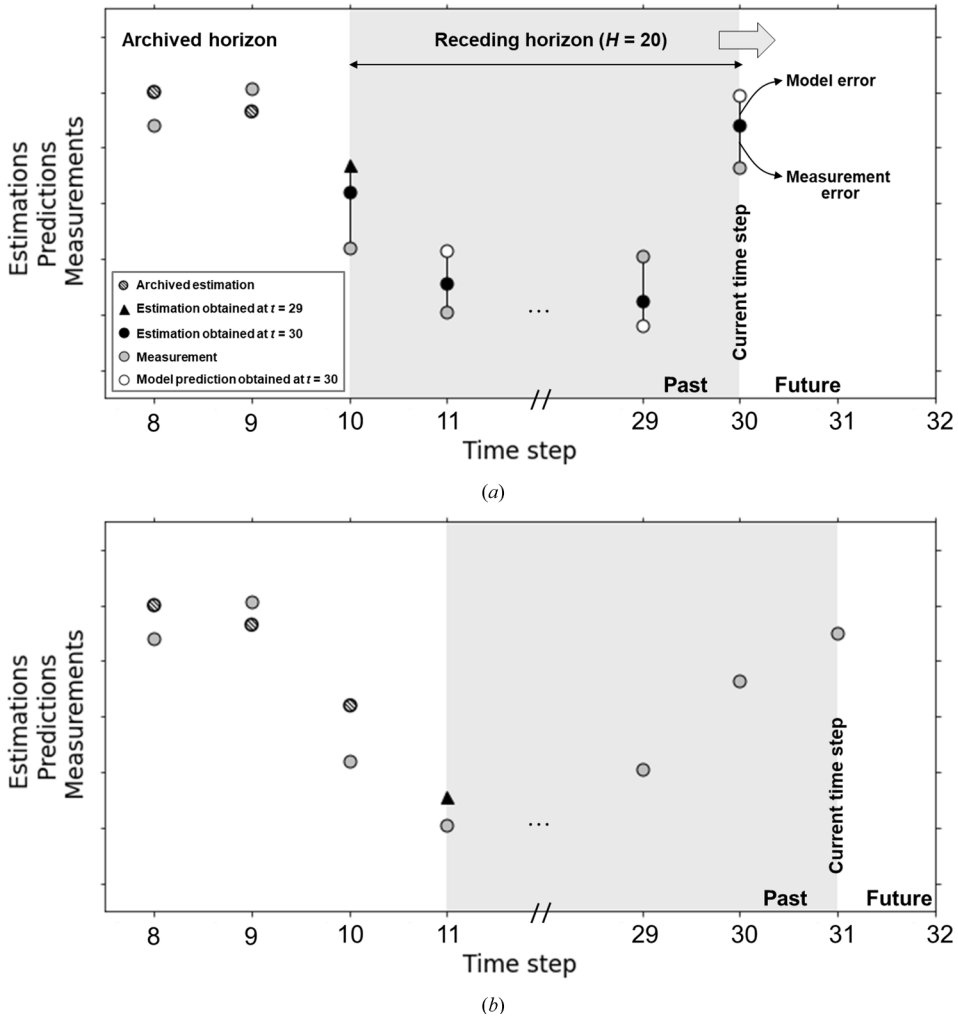

Fig. 5. Schematic view of an MHE. For illustrative purposes, we assume that the measured variables coincide with the system states ($\mathbf{g}(\widehat{\mathbf{x}}(t_l)) = \widehat{\mathbf{x}}(t_l)$in Equation (7). (*a*): State estimation at current time step ($t_k = 30$). (*b*): Window slide at next time step ($t_k = 31$), before of performing the state estimation procedure again.

form as:

$$\widetilde{\mathbf{x}}(t_{k+1}) = \mathbf{f}(\widetilde{\mathbf{x}}(t_k),\ \mathbf{u}(t_k)) \tag{4}$$

in which t_{k+1} is the time step subsequent to t_k, and $\mathbf{f}(\cdot)$: $\mathbb{R}^{n_x+n_u} \mapsto \mathbb{R}^{n_x}$ is the discrete-time form of the model.

Assume that n_y variables are measured from the process at a given time step t_k, and that they are collected in vector $\mathbf{y} \in \mathbb{R}^{n_y}$. A prediction of the measured variables can be obtained by:

$$\widetilde{\mathbf{y}}(t_k) = \mathbf{g}(\widetilde{\mathbf{x}}(t_k)) \tag{5}$$

where $\widetilde{\mathbf{y}} \in \mathbb{R}^{n_y}$ is the vector of the measured values as predicted by the model, and $\mathbf{g}(\cdot)$: $\mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_y}$ is the vector function describing the measurement model.

Even though a mathematical model can provide real-time predictions at any time step as a soft-sensor, unfortunately this piece of information cannot be used on its own for process monitoring due to the effect of the accumulation of error. Modeling errors (often referred to as plant/model mismatch) appear since the state variables predicted by Eq. (4) are different from the actual (and unknown) ones, as evidenced by the fact that $\widetilde{\mathbf{y}}(t_k) \neq \mathbf{y}(t_k)$. On the other hand, also process monitoring based on real-time measurements only may be ineffective, due to both lack of measurements and measurement noise.

State estimation (Ray, 1981) is a family of techniques that combines the use of the mathematical model of a system and of real-time measurements from the system to improve the prediction of the system states over the one provided by the model on its own. Namely, at each time step the state estimator calculates an estimation of the states in such a way as to reconcile the values of the measurements predicted by the model at that time step (or within a given time window including it) to the actual measurements at the same time step (or time window). State estimation adjusts the model at each sampling instant using the available measurements, thus improving the model predictions by providing accurate estimates for the states, namely the unmeasured quantities of interest.

In this study, we make use of moving-horizon state estimation (Rao, 2000; Rao and Rawlings, 2002). An intuitive representation of the rationale behind an MHE is illustrated in Fig. 5. At the current time step t_k (Fig. 5a), the MHE estimates the states of the system by finding a trade-off between the predictions from the model and the measurements available in a receding horizon $T(t_k) = \{t | t_{k-H} \leq t \leq t_k\}$ that includes the past H time steps (H is called window size). Instead, older measurements and estimations are retained in the so-called archived horizon (where the estimated states are not updated anymore). This is done practically by solving an optimization problem (Appendix A), namely by finding the set of estimated states in $T(t_k)$ that minimize an objective function that is basically made by the sum of the squared model errors and measurement errors within the receding horizon. At time step t_l inside the receding horizon, the vectors of model error $\mathbf{w} \in \mathbb{R}^{n_x}$ and measurement error $\mathbf{v} \in \mathbb{R}^{n_y}$ are defined as follows:

$$\mathbf{w}(t_l) = \widehat{\mathbf{x}}(t_{l+1}) - \widetilde{\mathbf{x}}(t_{l+1}), \qquad \text{for } l = k-H,\ k-H+1,\ \cdots,\ k-1 \tag{6}$$

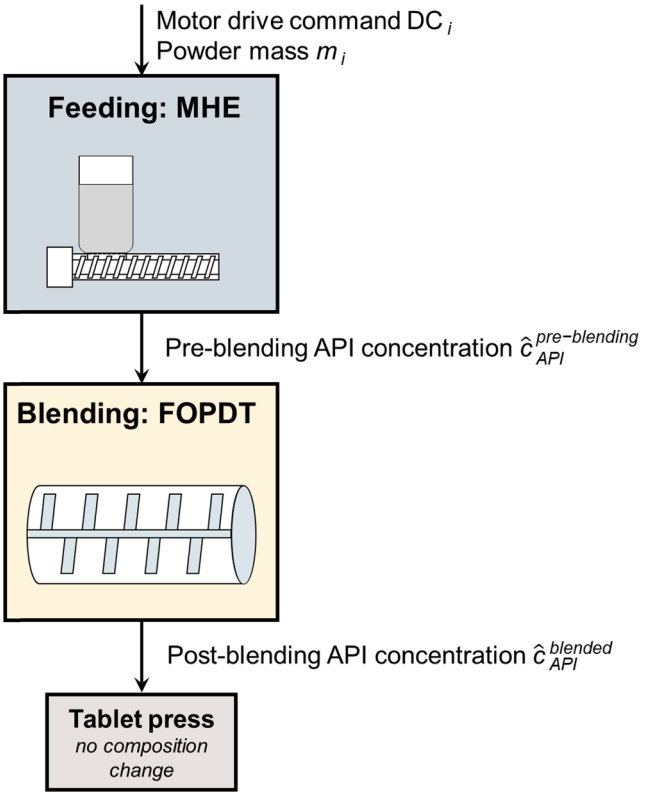

Fig. 6. Scheme of the proposed monitoring system for a direct compression line. The moving-horizon estimator uses real-time measurements of motor drive command and of powder mass from each feeder, and estimates the pre-blending API concentration. Powder blending occurring in the horizontal mixer and (to a lesser extent) in the vertical surge hopper and in the transition piece is accounted for through a first-order-plus-dead-time (FOPDT) dynamics. The API concentration in the blended powder mixture entering the feed frame is eventually calculated.

$$\mathbf{v}(t_l) = \mathbf{y}(t_l) - \mathbf{g}(\hat{\mathbf{x}}(t_l)), \qquad \text{for } l = k-H,\ k-H+1,\ \cdots,\ k \tag{7}$$

where $\hat{\mathbf{x}} \in \mathbb{R}^{n_x}$ is the vector of the estimated states (note that in the remainder of the manuscript, the caret symbol "^" on top of a variable denotes that the variable has been calculated by means of state estimation). At the left boundary of the window, instead of the model error, the arrival cost (which depends on the estimation obtained at time step t_{k-1} instead that on a model prediction) is considered in the objective function of the optimization problem, accounting for all the discarded measurements now laying inside the archived horizon (Appendix A).

At the next time step t_{k+1} (Fig. 5b), when new measurements become available, the window slides one time step forward. The new measurement set is included in the receding horizon, while the oldest one is discarded and enters the archived horizon. The estimation procedure is then iterated. The arrival cost at t_{k+1} is calculated using the estimation at the new left boundary of the window obtained at t_k. Additional details on the optimization problem formulation and on the arrival cost computation are given in Appendix A.

3.3. Feeder model description

Many data-driven and hybrid first-principles/data-driven models for loss-in-weight feeders are described in the literature. In this study, we use a model recently proposed by Bascone et al. (2020). The model is made up of two differential equations:

$$\frac{\mathrm{d}\tilde{m}_i(t)}{\mathrm{d}t} = \dot{m}_i^{refill}(t) - \tilde{\dot{m}}_i(t) \tag{8}$$

$$\tau_i \frac{\mathrm{d}\tilde{\dot{m}}_i(t)}{\mathrm{d}t} + \tilde{\dot{m}}_i(t) = n\ P\ A\ \eta\ \rho_i(t)\ N_i(t) \tag{9}$$

where $\tilde{m}_i$ is the predicted mass of powder in the feeder, $\dot{m}_i^{refill}$ is the powder refill mass flow, $\tilde{\dot{m}}_i$ is the predicted powder mass flow at the feeder outlet, τ_i is the feeder time constant (calibration parameter), n is the number of starts of the screw thread, P is the pitch of the screw, A is the cross-sectional area available for powder flow in the screw, η is the volumetric efficiency of the feeder (ratio between conveyed powder volume and screw volume available during one revolution, calculated from the material and feeder geometrical properties as in Bascone et al., 2020), N_i is the screw rotation speed, and ρ_i is the time-variable effective density of the powder in the screws.

The effective density is given by:

$$\rho_i(t) = \rho_i^0 + k_i \ln(\sigma_V(t)) \tag{10}$$

where ρ_i^0 is the density in the absence of vertical stress due to the powder in the hopper (calibration parameter), k_i is an additional calibration parameter, and σ_V is the vertical stress in the hopper, expressed in [Pa]. The vertical stress σ_V varies with time, and is a nonlinear function of $\tilde{m}_i$; in this study, we calculate σ_V with the set of equations proposed by Bascone et al. (2020) for Coperion K-Tron KT20 feeders.

The screw rotation speed N_i is obtained from DC$_i$ and the maximum screws rotation speed N_{max} as:

$$N_i(t) = \frac{DC_i}{100} N_{max} \tag{11}$$

Overall, the dynamic model described by Eqs. (8)–(11) has (for each feeder) two states ($\tilde{m}_i$ and $\tilde{\dot{m}}_i$), two inputs (DC$_i$ and $\dot{m}_i^{refill}$), and three calibration parameters (k_i, ρ_i^0 and τ_i). Materials physical properties and quantitative geometrical information on the twin-screw system configuration used in the process of interest are not disclosed due to confidentiality reasons. The main geometrical properties of KT20 feeders are reported in Bascone et al. (2020).

4. Implementation of the powder composition monitoring system

In this section, we present the general workflow required for implementing the powder composition monitoring system in a continuous direct compression line. The steps to be carried out are: *i*) feeder model calibration, *ii*) blending model calibration, and *iii*) MHE design. A general overview of the monitoring system is given in Subsection 4.1, together with remarks on the role of the blending model and the procedure for its calibration. For the feeding section, model calibration and state estimator design will be described in detail in Sections 4.2 and 4.3, respectively. We remind that the caret symbol "^" on top of a variable means that the variable has been obtained through state estimation, while variables with the tilde symbol "~" on top have been obtained through model predictions. Concentrations and mass flows without any symbol on top are instead measurements, either direct or indirect (Table 2).

4.1. Monitoring system overview

The proposed monitoring system (Fig. 6) is rooted into the MHE implemented in the feeding section. For each feeder *i*, given m_i and DC$_i$, the MHE estimates the outlet powder mass flow $\hat{\dot{m}}_i$ at each time step. The estimated API concentration in the powder mixture entering the blending section is then calculated as:

$$\hat{c}_{API}^{pre-blending}(t) = \frac{\hat{\dot{m}}_1(t)}{\sum_{i=1}^{5} \hat{\dot{m}}_i(t)} \tag{12}$$

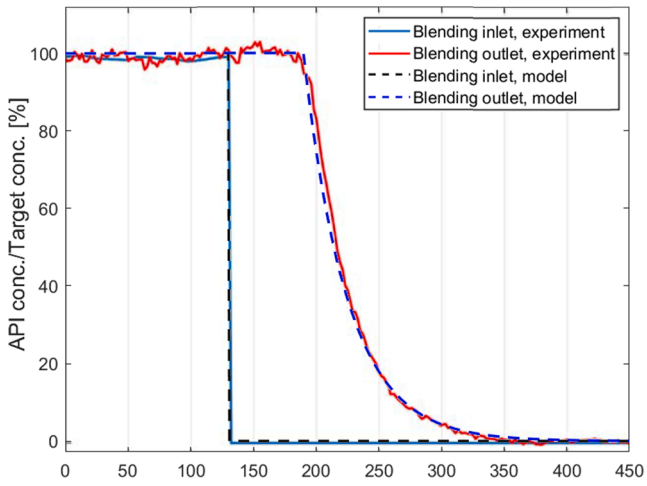

Fig. 7. Step change experiment for blending model calibration. Experimental data are reported as solid lines and are taken from Fig. 5 in García-Muñoz et al. (2017).

The blending effect throughout the line up to the point where the powder enters the die in the tablet press is modeled through a first-order-plus-dead-time dynamics (FOPDT; Seborg et al., 2017), accounting for the powder axial mixing and the residence time in this section of the process:

$$\tau_{blending}\frac{d\hat{c}_{API}^{blended}(t)}{dt}+\hat{c}_{API}^{blended}(t)=\hat{c}_{API}^{pre-blending}\left(t-\theta_{blending}\right) \tag{13}$$

where $\hat{c}_{API}^{blended}$ is the estimated API concentration at the blender outlet, and the calibration parameters $\theta_{blending}$ and $\tau_{blending}$ are, respectively, the blending dead time and time constant. As a remark, $\hat{c}_{API}^{blended}$ is the variable of interest for monitoring purposes, representing the API concentration that will also characterize the final tablets.

The blending model is calibrated offline with data from a step experiment (Fig. 7) taken from the literature (García-Muñoz et al., 2017) and carried out in the same blending units used for the experimental activities of this study. The estimated values of the blending parameters are: $\theta_{blending} = 60$ s and $\tau_{blending} = 35$ s. For simplicity, the blending model is used in a purely predictive fashion, instead of being embedded in the state estimation framework.

4.2. Feeder model calibration

Dataset A (Table 2) is used for calibrating the feeder model (Section 3.3) to each of the five feeders supplying the ingredients of the formulation. Maximum-likelihood estimation (MLE) under the assumptions discussed by Rawlings et al. (1993) is used for model parameter estimation, minimizing the objective function:

$$\Phi_{\mathrm{MLE}}(\theta_i)=\frac{n_{exp}}{2}\ln\left(\sum_{t_k=0}^{t_f}\left(\widetilde{m}_i(t_k,\theta_i)-m_i(t_k)\right)\right) \tag{14}$$

where n_{exp} is the number of experimental points in the dataset, t_f is the final time step of the experiment, and θ_i is the set vector of estimated parameters for feeder i. Note that, for a given feeder i, θ_i needs to be estimated again if the ingredient, the type of feeder, or the feeder characteristics (e.g., hopper or screws) are changed during the process life cycle. Bascone et al. (2020) indicate a linear dependence on N_i for all calibration parameters. From preliminary parameter estimation results, we conclude that, for the process of interest, the calibration parameters are constant and do not depend on N_i, except for ρ_i^0, which is calculated according to:

$$\rho_i^0=\rho_i^{0,intercept}+\rho_i^{0,slope}N_i \tag{15}$$

where $\rho_i^{0,intercept}$ and $\rho_i^{0,slope}$ respectively are the intercept and the slope of the equation representing the linear dependency of ρ_i^0 on N_i. Hence, θ_i defined as:

$$\theta_i=\begin{bmatrix}k_i & \rho_i^{0,intercept} & \rho_i^{0,slope} & \tau_i\end{bmatrix} \tag{16}$$

We solve the optimization problem for parameter estimation in the Pyomo environment, based on Python (Hart et al., 2017). Details on the procedure followed for reaching the solution are reported in Appendix B. The estimated parameters, together with the estimation uncertainty, are reported in Table 3. All parameters are estimated with sufficiently small uncertainty, except $\rho_4^{0,slope}$. However, we verified that this parameter has a low sensitivity on the API concentration in the tablet.

The model performance (no state estimation) is assessed in Fig. 8 for Dataset A (calibration) and Dataset B (validation). The plots in the figure refer to Ingredients 1 (API) and 2 of the formulation, making up for more than 80% of the tablet. Fig. 8a, and its zoomed-in version Fig. 8b, show that, in Dataset A, the experimental powder mass profile (i.e., the only measurement used in the objective function of Equation (14)) is fitted well by the model for material 1. The model predicts API mass flow values ($\widetilde{m}_1$) that agree well with $\dot{m}_1^{avg}$ (Fig. 8c), resulting in a composition signal that is much less noisy than $\dot{m}_1^{inst}$. Analogous considerations can be drawn for Ingredient 2 in Dataset A (Fig. 8d), and Ingredients 1–2 in Dataset B (Fig. 8e–f).

4.3. MHE design

Based on the calibrated feeder model, the MHE is implemented in the Pyomo environment with the procedure detailed in Appendix B. At each time step, the estimated state vector includes 10 elements, while the input and measurement vectors have 5 elements each, namely:

$$\hat{\mathbf{x}}=\begin{bmatrix}\hat{m}_1 & \hat{\dot{m}}_1 & \hat{m}_2 & \hat{\dot{m}}_2 & \cdots & \hat{m}_5 & \hat{\dot{m}}_5\end{bmatrix}\in\mathbb{R}^{10} \tag{17}$$

$$\mathbf{u}=[\mathrm{DC}_1 \quad \mathrm{DC}_2 \quad \cdots \quad \mathrm{DC}_5]\in\mathbb{R}^5 \tag{18}$$

$$\mathbf{y}=[m_1 \quad m_2 \quad \cdots \quad m_5]\in\mathbb{R}^5 \tag{19}$$

Implementation of the MHE requires designing the tuning matrices **R** (measurement noise covariance matrix) and **Q** (model error covariance matrix), as outlined in Appendix A. Matrix **R** is designed as a diagonal matrix:

Table 3
Feeder model: parameter estimation results. For each parameter, the estimation uncertainty is reported as standard deviation from the estimated value.

Parameter	Feeder 1		Feeder 2		Feeder 3		Feeder 4		Feeder 5	
	Estimation	Std. dev.	Estimation	Std. dev.	Estimation	Std. dev.	Estimation	Std. dev.	Estimation	Std. dev.
k_i [kg/m^3]	−3.16	1.34E-2	6.66	2.40E-2	−1.85	3.44E-2	10.46	1.81	−2.21	1.64E-1
$\rho_i^{0,intercept}$ [kg/m^3]	403.12	8.52E-1	265.16	1.64	274.61	1.08	133.19	1.44E1	249.36	2.01
$\rho_i^{0,slope}$ [kg/m^3/rpm]	−3.28	1.11E-1	−1.25	8.54E-2	−4.71	2.18E-1	−4.34	1.14E1	−10.05	3.50E-1
τ_i[s]	0.72	1.28E-1	2.10	5.09E-7	0.52	6.42E-2	0	1.10E-4	0.98	1.27E-1

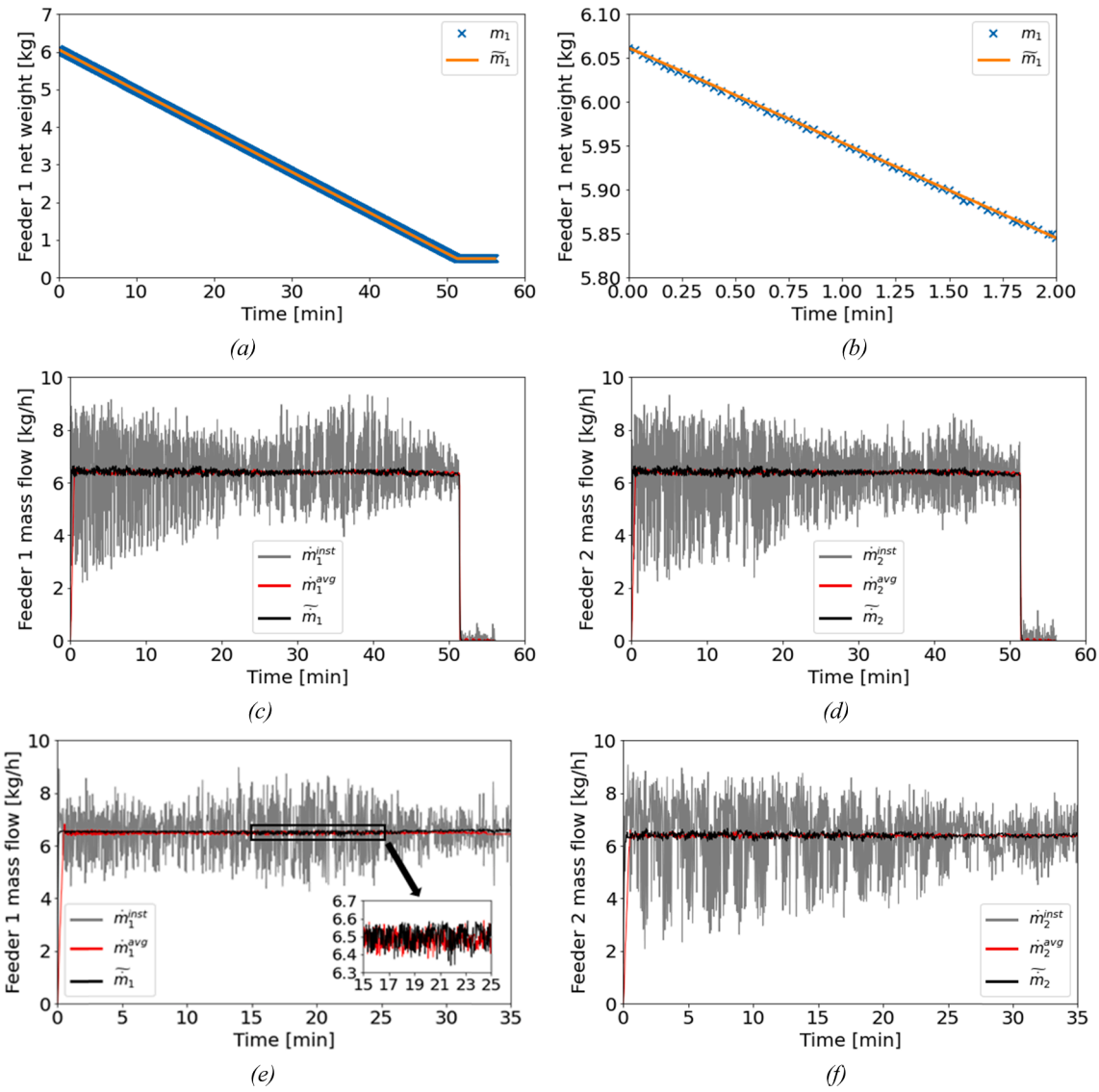

Fig. 8. Experimental measurements compared against the predictions of the calibrated model. (*a*) Dataset A, m_1 vs $\widetilde{m}_1$, (*b*) zoom-in of Fig. 8**a**, (*c*) Dataset A, $\dot{m}_1^{inst}$ vs $\dot{m}_1^{avg}$ vs $\widetilde{\dot{m}}_1$, (*d*) Dataset A, $\dot{m}_2^{inst}$ vs $\dot{m}_2^{avg}$vs $\widetilde{\dot{m}}_2$, (*e*) Dataset B, $\dot{m}_1^{inst}$ vs $\dot{m}_1^{avg}$ vs $\widetilde{\dot{m}}_1$, (*f*) Dataset B, $\dot{m}_2^{inst}$ vs. $\dot{m}_2^{avg}$ vs $\widetilde{\dot{m}}_2$. Fig. 8**e** also reports a zoom-in of $\dot{m}_1^{avg}$ vs $\widetilde{\dot{m}}_1$.

Table 4
MHE tuning comparison (Dataset A). Values of the tuning parameters of **Q** (Equation (21)) used for obtaining the estimations reported in Fig. 9.

	$\sigma^2_{\widetilde{m}_1}$	$\sigma^2_{\widetilde{m}_2}$	$\sigma^2_{\widetilde{m}_3}$	$\sigma^2_{\widetilde{m}_4}$	$\sigma^2_{\widetilde{m}_5}$
Tuning Fig. 9a	0	0	0	0	0
Tuning Fig. 9b	7.5	7.5	1.5	7.5	15
Tuning Fig. 9c	300	7.5	1.5	7.5	15
Tuning Fig. 9d	3000	7.5	1.5	7.5	15

$$\mathbf{R} = \mathrm{diag}\left(\begin{bmatrix}\sigma^2_{m_1} & \sigma^2_{m_2} & \cdots & \sigma^2_{m_5}\end{bmatrix}\right) \tag{20}$$

where $\sigma^2_{m_i}$ (variance of the mass measurement of feeder *i*) is assigned the value 1×10^{-6} for $i = 1, \ldots, 5$. Matrix **Q** is tuned as a diagonal matrix, too:

$$\mathbf{Q} = \mathrm{diag}\left(\begin{bmatrix}\sigma^2_{m_1} & \sigma^2_{\dot{m}_1} & \sigma^2_{m_2} & \sigma^2_{\dot{m}_2} & \cdots & \sigma^2_{m_5} & \sigma^2_{\dot{m}_5}\end{bmatrix}\right) \tag{21}$$

in which, for each feeder *i*, $\sigma^2_{m_i}$ and $\sigma^2_{\dot{m}_i}$ are the variances of the element of $\mathbf{w}(t_l)$ corresponding to the discretized version of (respectively) Eqs. (8) and (9). Note that Eq. (8) has no modeling error, as it represents the material balance of powder in the feeder. Hence, the elements of $\mathbf{w}(t_l)$ corresponding to Eq. (8) are set to 0. For the same reason, $\sigma^2_{m_i}$ is set to 0, for $i = 1, \ldots, 5$. On the other hand, Eq. (9) is subject to error, and the variances $\sigma^2_{\dot{m}_i}$ are assigned by trial and error. Namely, we used Dataset A and adjusted the diagonal elements until the variability interval observed for $\hat{c}^{blended}_{API}$ was roughly comparable to the one observed experimentally for c^{NIR}_{API} (i.e., for the only concentration measurement available at sufficiently high rate from downstream). The tuning parameters that were found to have the strongest impact on the state estimation are $\sigma^2_{\dot{m}_1}$ and $\sigma^2_{\dot{m}_2}$, which is not surprising because Ingredients 1 and 2 are the main components of the tablet formulation.

The $\hat{c}^{blended}_{API}$ values obtained with four different tunings of the MHE (Table 4) are illustrated in Fig. 9, with increasing values of $\sigma^2_{\dot{m}_i}$ from Fig. 9a–d. As expected, the lowest $\hat{c}^{blended}_{API}$ fluctuations are registered in Fig. 9a, where the MHE response is practically equivalent to the model prediction. By increasing $\sigma^2_{\dot{m}_i}$, the MHE is tuned to trust the model less and less, and to give increasingly more importance to the measurements

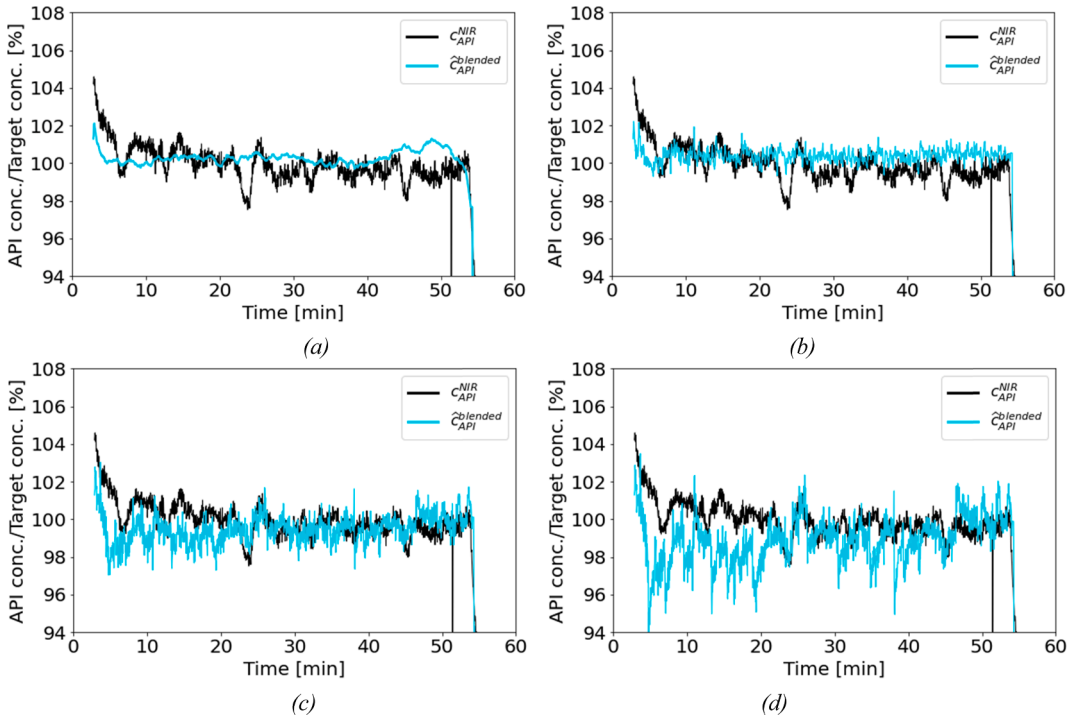

Fig. 9. MHE tuning comparison (Dataset A). Time profile of $\hat{c}_{API}^{blended}$ obtained with decreasing confidence in the model from *(a)* to *(d)*. The values of the different tuning parameters are reported in Table 4. The time profile of c_{API}^{NIR} is reported for comparison.

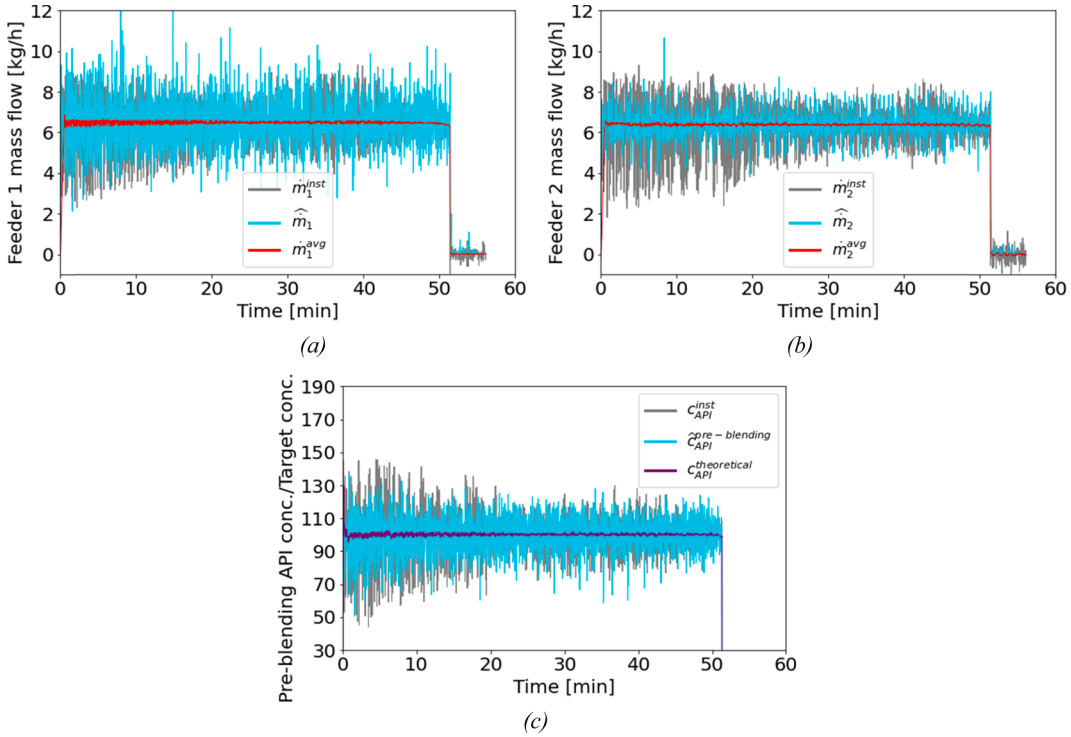

Fig. 10. Comparison of pre-blending estimations and measurements (Dataset A). *(a)* $\dot{m}_1^{inst}$ vs $\hat{\dot{m}}_1$ vs $\dot{m}_1^{avg}$; *(b)* $\dot{m}_2^{inst}$ vs $\hat{\dot{m}}_2$ vs $\dot{m}_2^{avg}$; *(c)* c_{API}^{inst} vs $\hat{c}_{API}^{pre-blending}$ vs $c_{API}^{theoretical}$.

(see Eq. (A.1) in Appendix A). When measurements are weighted very strongly (Fig. 9d), the $\hat{c}_{API}^{blended}$ profile becomes qualitatively similar to the very noisy API concentration calculated from the raw instantaneous mass flows (Fig. 1a). The parameters used for the estimation shown in Fig. 9c are eventually selected, as they lead to a $\hat{c}_{API}^{blended}$ profile that more closely resembles the c_{API}^{NIR} one.

5. Real-time monitoring: Proof of concept

We test the proposed monitoring system on Datasets A-C. Real-time monitoring is mimicked by using the data stored in the datasets as if they were coming from the sensors of the running plant. For Dataset A, the obtained profiles of the estimated mass flows $\hat{\dot{m}}_1$ and $\hat{\dot{m}}_2$ at the feeder

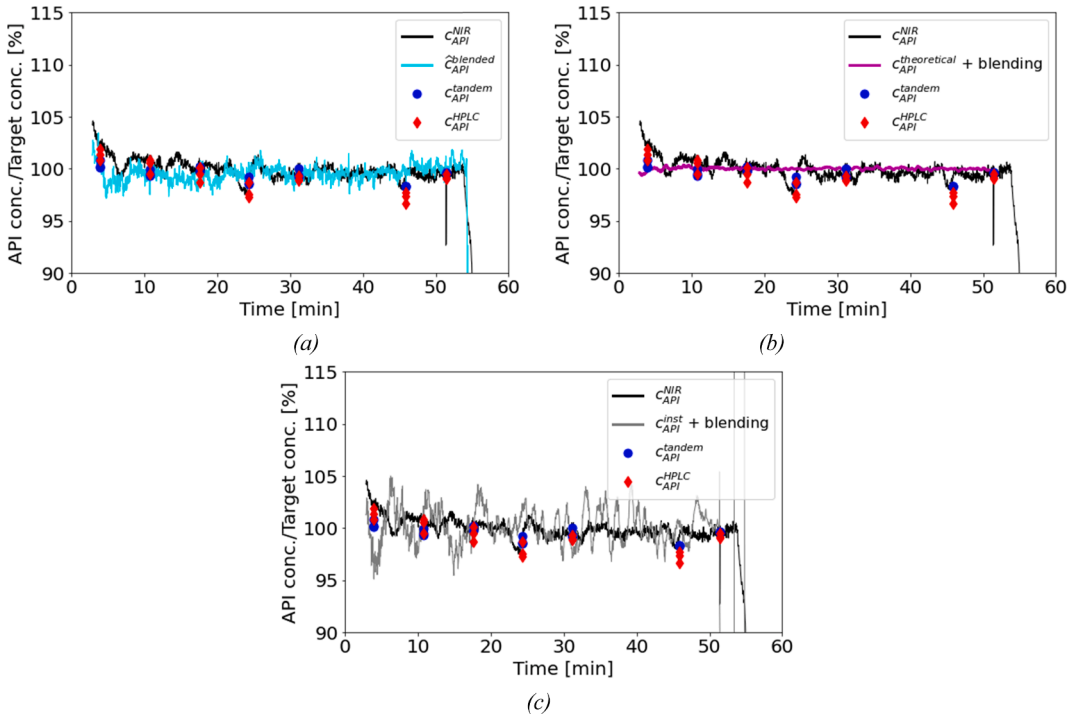

Fig. 11. Comparison of post-blending estimations and measurements (Dataset A). *(a)* $\hat{c}_{API}^{blended}$, *(b)* $c_{API}^{theoretical}$ + blending effect (Equation (13)) and *(c)* c_{API}^{inst} + blending effect (Equation (13)). We include c_{API}^{HPLC}, c_{API}^{NIR} and c_{API}^{tandem} for reference.

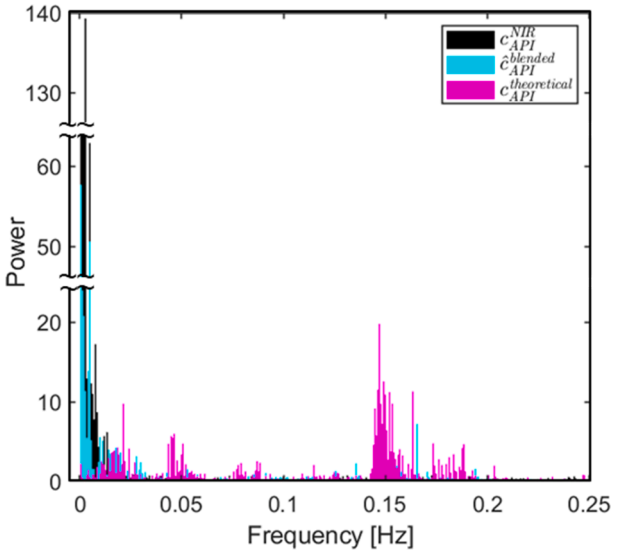

Fig. 12. Single-sided Fourier transform derived power spectrum of c_{API}^{NIR}, $\hat{c}_{API}^{blended}$ and $c_{API}^{theoretical}$ (Dataset A). Note that the sampling time of the datasets is 2 s, corresponding to a sampling frequency of 0.50 Hz.

outlet (Fig. 10a and b, respectively) show much greater variability than the corresponding $\dot{m}_i^{avg}$, and slightly smaller than $\dot{m}_i^{inst}$. The same considerations apply to $\hat{c}_{API}^{pre-blending}$, which is compared to c_{API}^{inst} and $c_{API}^{theoretical}$ in Fig. 10c.

The resulting $\hat{c}_{API}^{blended}$ profile (downstream the blending section) is validated against c_{API}^{HPLC}, c_{API}^{NIR} and c_{API}^{tandem} measurements in Fig. 11a (in the industrial practice, c_{API}^{HPLC} is considered the most reliable API concentration measurement across the whole process line). We remark that none of these validation measurements are used by the monitoring system for estimating $\hat{c}_{API}^{blended}$. Visual inspection shows that $\hat{c}_{API}^{blended}$ is aligned well with the HPLC and NIR-derived measurements in the final tablet. Quantitively speaking, the sum of squared errors (SSE) between c_{API}^{HPLC} and $\hat{c}_{API}^{blended}$ is ~ 20% smaller than the SSE between c_{API}^{HPLC} and the blending outlet concentration calculated applying the blending dynamics (Equation (13)) to $c_{API}^{theoretical}$ (Fig. 11b). The SSE reduction is even greater (~40%) when c_{API}^{tandem} is considered, instead of c_{API}^{HPLC}. When $\hat{c}_{API}^{blended}$ is benchmarked against the blending outlet concentration calculated applying the blending dynamics to c_{API}^{inst} (Fig. 11c), the SSE reduction with respect to c_{API}^{HPLC} is greater than 40%, while the one with respect to c_{API}^{tandem} is greater than 60%.

Although $\hat{c}_{API}^{blended}$ and c_{API}^{NIR} do not match perfectly in Fig. 11a, it appears from the same figure that $\hat{c}_{API}^{blended}$ captures the underlying low-frequency dynamics that is apparent in the c_{API}^{NIR} signal. Instead, the blending outlet concentration calculated from $c_{API}^{theoretical}$ (Fig. 11b), essentially, does not display any dynamics, due to the low signal-to-noise ratio of $c_{API}^{theoretical}$ (Fig. 1). Application of fast Fourier transform (a mathematical technique capable of identifying the frequencies of the underlying dynamics of a signal) to $\hat{c}_{API}^{blended}$, c_{API}^{NIR} and $c_{API}^{theoretical}$ confirms these observations. Fig. 12 shows the Fourier transform-derived power spectra (not to be confused with the spectra obtained through PAT in the plant) of the $\hat{c}_{API}^{blended}$, c_{API}^{NIR} and $c_{API}^{theoretical}$signals. Both $\hat{c}_{API}^{blended}$ and c_{API}^{NIR} have their main components laying in the low frequencies band of 0.001–0.005 Hz, corresponding to periods of about 3–15 min (coherently with the plots displayed in Fig. 11a). Instead, $c_{API}^{theoretical}$ mostly presents higher-frequency components that are associated with noise, and, after the blending effect is applied to $c_{API}^{theoretical}$, almost no low-frequency dynamics is left (Fig. 11b).

The partial misalignment between $\hat{c}_{API}^{blended}$ and c_{API}^{NIR} in Fig. 11a might be due to unmodelled stochastic phenomena occurring in the blending section, or to the need of improving the feeder model (on the structural side and/or on the parametric one). In particular, the datasets used in this study come from a pilot-plant experimental campaign on a drug product that was not analyzed in further experiments. Hence, limited

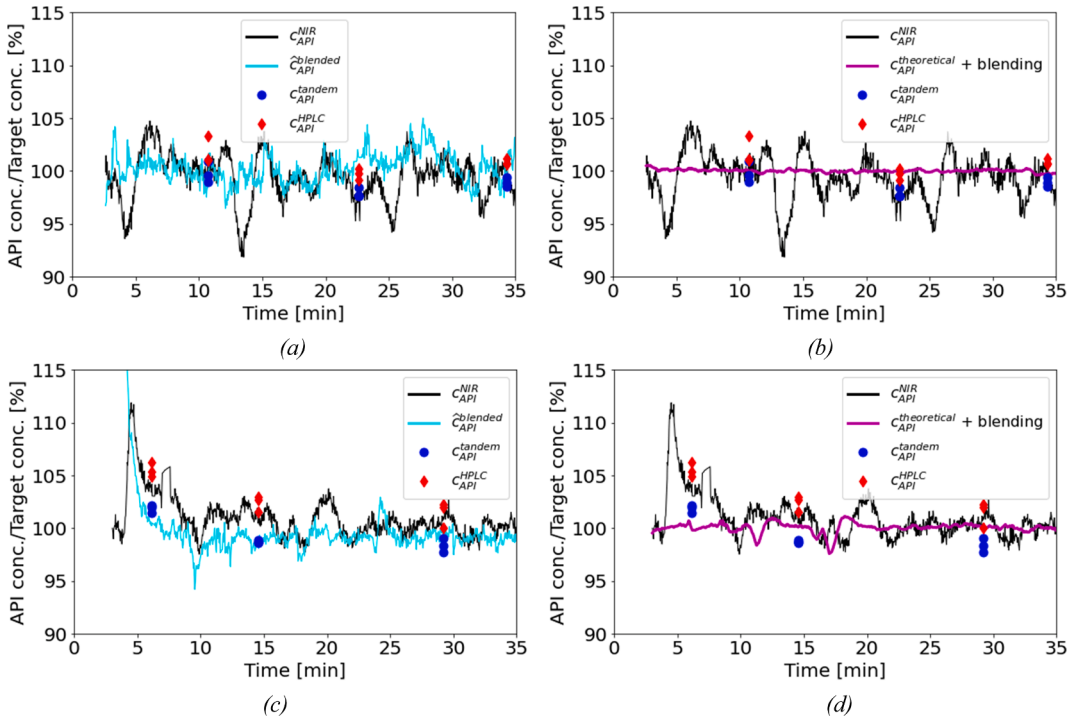

Fig. 13. Comparison of post-blending estimations and measurements (Datasets B and C). *(a)* $\hat{c}_{API}^{blended}$, Dataset B, *(b)* $c_{API}^{theoretical}$ + blending effect (Equation (13)), Dataset B, *(c)* $\hat{c}_{API}^{blended}$, Dataset C, *(d)* $c_{API}^{theoretical}$ + blending effect (Equation (13)), Dataset C.c_{API}^{HPLC}, c_{API}^{NIR} and c_{API}^{tandem} are reported in all plots for comparison.

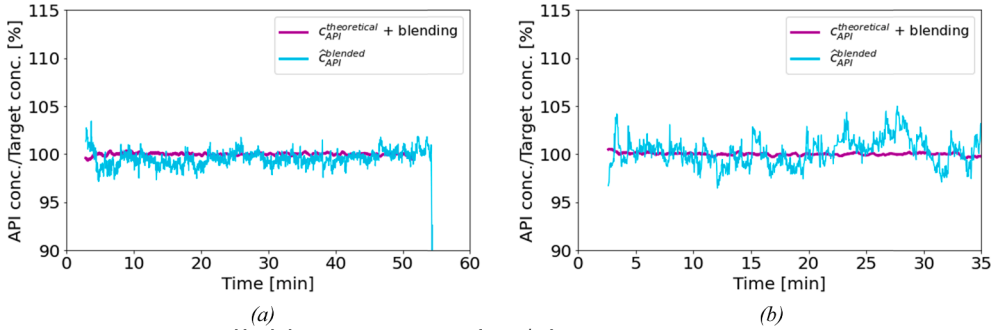

Fig. 14. $\hat{c}_{API}^{blended}$ compared against $c_{API}^{theoretical}$ + blending effect (Equation (13)) for *(a)* Dataset A and *(b)* Dataset B.

data were available for calibrating the chemometric model relating the NIR spectra to the API concentration, and complete trust in c_{API}^{NIR} cannot be argued.

The results demonstrate that the proposed monitoring framework provides better estimations of the API concentration across the process than done by a conventional approach based on monitoring the API concentration by $c_{API}^{theoretical}$. Notice that the variability of the API concentration upstream the blending section as estimated by the MHE is greater than the one represented by $c_{API}^{theoretical}$ (see Fig. 10c). This is expected given that the latter is calculated with a moving average, which is a dampened signal by definition.

To further appreciate the benefits of including the state estimator in the monitoring system, consider again Fig. 11c. It can be seen that the API concentration at the blending outlet as obtained by applying the blending dynamics of Equation (13) to c_{API}^{inst} still displays high-frequency fluctuations (e.g. from ~ 30 to ~ 45 min), not resembling any of the measured c_{API}^{HPLC}, c_{API}^{NIR} and c_{API}^{tandem} profiles. No practical improvement is achieved by either using a discrete three-point derivative for calculating $\dot{m}_i^{inst}$, or applying a low-pass filter to c_{API}^{inst}. This means that, despite $\hat{c}_{API}^{pre-blending}$ variability is similar to the one of c_{API}^{inst} (Fig. 10c), the fluctuations observed in the MHE estimations reflect the underlying process dynamics, instead of noise as in c_{API}^{inst}; re-configuring the number of points of the moving-average for calculating $\dot{m}_i^{avg}$ would not suffice for achieving the good performance reached by the MHE.

We test the MHE-based monitoring system also for Datasets B and C of Table 1. The manufactured tablets have a different mass than those related to Dataset A, and this would suggest to consider retuning both the model and the MHE. Nevertheless, for simplicity we consider a worst-case scenario where the tuning is not adjusted. The results shown in Fig. 13 are still satisfactory. In fact, $\hat{c}_{API}^{blended}$ is coherent with c_{API}^{HPLC} and c_{API}^{tandem}, while the blending outlet concentration calculated by applying the blending dynamics to $c_{API}^{theoretical}$ essentially does not display any dynamics.

By overlapping $\hat{c}_{API}^{blended}$ with $c_{API}^{theoretical}$ after blending (Fig. 14), it is once again confirmed that $c_{API}^{theoretical}$ is an over-smoothed signal mainly including high-frequency noise components, while $\hat{c}_{API}^{blended}$ includes lower-frequency components representing the underlying dynamics of the process. The misalignment between $\hat{c}_{API}^{blended}$ and c_{API}^{NIR} in Fig. 13a and

13c might be due to the aforementioned reasons, such as disturbances in the blending or to the need to improve the feeders model. However, for these datasets an even more important role in the mismatch might be played by the unreliability of c_{API}^{NIR}, possibly due to poor sensor readings. For instance, referring to Fig. 13e, the occurrence of 10% variations of API concentrations in less than 2 min (e.g., from ~12 to ~14 min) with no likewise variations in the feeders mass measurements seems unlikely. Finally, since the tablets produced in the experiments of Datasets B and C have increasingly larger mass, another reason for the mismatch might be related to the need of tuning the MHE for experiments producing tablets of a fixed mass. These possibilities will be further inquired in future work.

As a concluding remark, the computational burden of the monitoring framework is compatible with real-time implementation. The state estimation results were obtained in approximately 50% of the process time with an Intel® Core™ i7-8565U CPU @1.80 GHz processor and total memory of 16.0 GB RAM. The computational time can be further reduced by using smaller values for H as in Liu et al. (2018) or with fast MHE approaches (Zavala et al., 2008).

6. Conclusions

We presented a novel approach to monitoring powder feeding in continuous solid-dosage forms manufacturing, and successfully validated it against experimental datasets collected on a direct compression line. The monitoring system is based on a state estimator (MHE), which effectively reconciles the mass measurements coming from loss-in-weight feeders with downstream measurements potentially coming from a PAT instrument, and estimates the delivered powder mass flows by means of a model-based optimization strategy. The monitoring system exploits a detailed mathematical model of the process for state estimation purposes, meeting the Quality-by-Design framework invitation to develop control strategies rooted on enhanced process understanding. The powder mass flows estimated with the proposed monitoring approach (and, in turn, the estimated API concentration in the final dosage forms) are highly consistent with downstream HPLC and spectroscopic measurements, when compared with traditional approaches to feeding monitoring. Additionally, powder concentration estimations are provided in a practically continuous way, which is a significant advantage over sampled measurements. We showed that using statistical filters instead of the state estimator for calculating the powder mass flows from the mass measurements, as done by default by loss-in-weight feeder software, can lead to inaccuracies. Finally, the required computational time is compatible with real time implementation.

Future work on a different molecule will involve additional validation activities on extended datasets, for which a reliable chemometric model relating the NIR spectra to the API concentration in the powder mixture is available. The integration of the state estimator with latent-variable modeling for fault detection and diagnosis purposes is also envisioned, following a recently proposed hybrid monitoring approach (Destro et al., 2020).

We believe state estimation technology is the appropriate way to obtain maximum synergy from the resources invested in the development of a deterministic model for a process, and the resources dedicated to the implementation of PAT solutions. State estimation puts the data from PAT solutions in the context of a the wholistic view of the process that is represented in a deterministic model. This is, in the opinion of the authors, the first step in the journey to adapt model-based closed loop control in the pharmaceutical industry.

CRediT authorship contribution statement

Francesco Destro: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Visualization, Formal analysis. **Salvador García Muñoz:** Conceptualization, Validation, Writing – review & editing. **Fabrizio Bezzo:** Conceptualization, Validation, Writing – review & editing. **Massimiliano Barolo:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge financial support from the University of Padova under project BIRD194889-SID 2019 “Augmenting data-driven models with knowledge-driven information to enhance process monitoring in the Industry 4.0 era (AUGH)”. F.D. gratefully acknowledges the CARIPARO Foundation for his PhD scholarship. We also thank Adam Butterbaugh from Eli Lilly for his kind assistance in collecting the data sets used in this work.

Appendix A -. Mathematical formulation of moving-horizon estimation

In this Appendix, we complement the description of the MHE algorithm given in Section 3.2 with a formal mathematical formulation.

Let us consider a time step t_l lying inside the receding horizon $T(t_k)$. We further detail the notation introduced in Section 3.2 by denoting with $\widehat{\mathbf{x}}(t_l|t_k)$ the estimation of the system states at time t_l when the system is at time t_k (i.e., the estimation is based on all measurements within the receding horizon). At the current time step t_k, the optimal estimation of the states inside $T(t_k)$, namely $\{\widehat{\mathbf{x}}(t_{k-H}|t_k), \cdots, \widehat{\mathbf{x}}(t_k|t_k)\}$, is obtained by solving the following nonlinear programming problem:

$$\min_{\widehat{\mathbf{x}}(t_{k-H}|t_k),\mathbf{w}(t_{k-H}),\cdots,\mathbf{w}(t_{k-1})} \left\{ \Theta(\widehat{\mathbf{x}}(t_{k-H}|t_k)) + \frac{1}{2}\sum_{t=t_{k-H}}^{t_k} \mathbf{v}(t)^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{v}(t) + \frac{1}{2}\sum_{t=t_{k-H}}^{t_{k-1}} \mathbf{w}(t)^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{w}(t) \right\} \tag{A.1}$$

$$\text{subjectto}: \widetilde{\mathbf{x}}(t_{l+1}|t_k) = \mathbf{f}(\widehat{\mathbf{x}}(t_l|t_k), \mathbf{u}(t_l)) \tag{A.2a}$$

$$\widehat{\mathbf{x}}(t_{l+1}|t_k) = \widetilde{\mathbf{x}}(t_{l+1}|t_k) + \mathbf{w}(t_l) \tag{A.2b}$$

$$\mathbf{y}(t_l) = \mathbf{g}(\widehat{\mathbf{x}}(t_l|t_k)) + \mathbf{v}(t_l) \tag{A.2c}$$

$$\mathbf{w}(t_l) \in \mathbf{W}_l \tag{A.2d}$$

$$\widehat{\mathbf{x}}(t_{l+1}|t_k) \in \mathbf{X}_l \tag{A.2e}$$

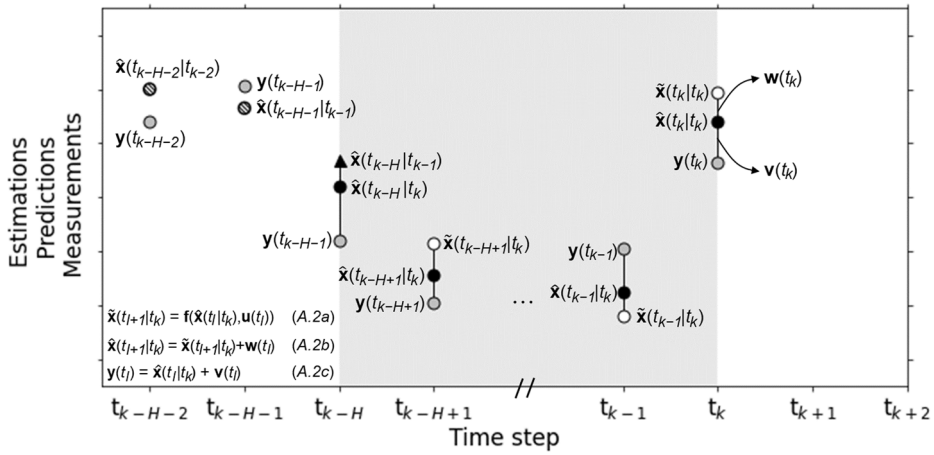

Fig. A1. Schematic view of the MHE of Fig. 5, proposed again using the notation for estimations, predictions and measurements introduced in this Appendix.

where, $\Theta(\tilde{\mathbf{x}}(t_{k-H}))$ is referred to as the arrival cost, $\mathbf{R} \in \mathbb{R}^{n_y \times n_y}$ is the measurement error covariance matrix, $\mathbf{Q} \in \mathbb{R}^{n_x \times n_x}$ is the model error covariance matrix, $\tilde{\mathbf{x}}(t_{l+1}|t_k)$ is the vector of predicted states at time step t_{l+1}, obtained at t_k from initial conditions $\hat{\mathbf{x}}(t_l|t_k)$, and $\mathbf{W}_l$ and $\mathbf{X}_l$, (which define the boundaries of, respectively, $\mathbf{w}(t_l)$ and $\hat{\mathbf{x}}(t_l)$, are taken as polyhedral convex sets.

The objective function in Equation (A.1) represents in mathematical form the request of finding the optimal set of estimated states that minimize the sum of squared model errors and measurement errors within the receding horizon. Note that Equations A.2b and A.2c are, respectively, equivalent to Equations (6) and (7). The arrival cost accounts for all the data collected in the archived horizon ($0 \leq t < t_{k-H}$), not directly considered in the objective function (Equation (A.1)). It is defined as:

$$\Theta(\hat{\mathbf{x}}(t_{k-H}|t_k)) = \frac{1}{2}\left((\hat{\mathbf{x}}(t_{k-H}|t_{k-1}) - \hat{\mathbf{x}}(t_{k-H}|t_k))^{\mathrm{T}}\mathbf{P}^{-1}(t_{k-H}|t_{k-1})(\hat{\mathbf{x}}(t_{k-H}|t_{k-1}) - \hat{\mathbf{x}}(t_{k-H}|t_k))\right) \tag{A.3}$$

in which $\mathbf{P}(t_{k-H}|t_{k-1})$ is the covariance of the estimation of $\hat{\mathbf{x}}(t_{k-H}|t_{k-1})$.

The goodness of the MHE estimation increases with both H and the arrival cost computation accuracy. Hence, the computational burden for solving an MHE problem can be reduced by using a smaller H, and compensating with an improved approximation of the arrival cost. In this study, we follow a recently proposed approach for the calculation of the arrival cost (López-Negrete and Biegler, 2012), which computes $\mathbf{P}(t_{k-H}|t_{k-1})$ from the reduced Hessian of Problem A.1-A.3 at the solution. The mathematical notation introduced to formalize the MHE problem is summarized in Fig. A.1, to be compared with the qualitative schematic view of the MHE algorithm presented in Fig. 5.

In the MHE formulation of Equations (A.1)–(3), $\mathbf{v}(t_l)$ and $\mathbf{w}(t_l)$ are assumed to be zero-mean Gaussian processes of covariances $\mathbf{R}$ and $\mathbf{Q}$, respectively. In the presence of gross errors, robust implementations of MHE can be resorted to (Nicholson et al., 2014). Matrices $\mathbf{Q}$ and $\mathbf{R}$ are regarded as tuning parameters, respectively representing the confidence in the model and in the measurements. If the model is trusted more than the measurements, due to significant noise, larger elements for the matrix $\mathbf{R}$ should be selected than for $\mathbf{Q}$, and vice versa.

Appendix B. - Additional information on parameter estimation and MHE implementation

In this Appendix we provide further details on the procedures adopted for solving the parameter estimation problem (Section 4.2) and for implementing the MHE (Section 4.3). Both activities have been carried out in the Pyomo environment (Hart et al., 2017), based on Python.

For parameter estimation, the feeders model is first coded into the Pyomo environment in the time-continuous form of Equations (8)–(11). Then, the set of differential–algebraic equations are transformed in time-discrete form using a simultaneous strategy (Biegler, 2007), exploiting the Pyomo.dae toolbox (Nicholson et al., 2018).The differential equations are approximated with polynomials using Radau orthogonal collocation (Biegler, 2010), with 1500 finite elements (corresponding to n_{exp}) and 3 collocation points. The obtained time-discrete feeder model equations are used as constraints in the optimization problem for minimizing Φ_{MLE} (Equation (14)). The solution is reached using the large-scale nonlinear solver IPOPT (Wächter and Biegler, 2006), providing the optimal estimation of parameters $\hat{\theta}_i$ for $i = 1, 2, \ldots, 5$ (Table 3). The corresponding estimation uncertainty is given by the variance–covariance matrix (whose diagonal is reported in standard deviation form in Table 3), approximated as the inverse of the Hessian of the objective function (Equation (14)) at the solution (Bard, 1974). For computing the Hessian at the solution, we use sIPOPT (Pirnay et al., 2012), an extension of IPOPT.

On the state estimation side, the MHE framework is implemented starting from the continuous form of the feeders model already coded in the Pyomo environment for parameter estimation. A time step of 2 s is used (corresponding to the sampling time of the actual measurements), and a value of $H = 30$ is selected, since it is found to be a good compromise between computational burden and estimation accuracy. The MHE framework is defined by the optimization problem of Equations A.1-A.3. The discrete-time model $\mathbf{f}(\hat{\mathbf{x}}(t_{l+1}|t_k), \mathbf{u}(t_k))$ of Equation A.2a is obtained using the same approach adopted for parameter estimation, namely through discretization with Radau orthogonal collocation of Equations (8)–(11). Also in this case, 3 collocation points are used, but the number of finite elements is set to 30, corresponding to the selected value of H. The measurement model $\mathbf{g}(\hat{\mathbf{x}}(t_l|t_k))$ appearing in Equation A.2c is defined as:

$$\mathbf{g}(\hat{\mathbf{x}}(t_l|t_k)): \quad \hat{\mathbf{x}}(t_l|t_k) \quad \mapsto \quad \begin{bmatrix} \hat{m}_1(t_l|t_k) & \hat{m}_2(t_l|t_k) & \cdots & \hat{m}_5(t_l|t_k) \end{bmatrix} \tag{B.1}$$

At each t_k, IPOPT (Wächter and Biegler, 2006) is used for solving the MHE optimization problem, and the estimated hold-ups and mass flows at the current time step (respectively, $\hat{m}_i(t_k)$ and $\hat{m}_i(t_k)$ for $i = 1, \ldots, 5$) are stored. Then, sIPOPT (Pirnay et al., 2012) is resorted to for calculating the reduced

Hessian of Problem A.1-A.3 at the solution with respect to $\widehat{\mathbf{x}}(t_{k-H+1}|t_k)$. Following the approach proposed by López-Negrete and Biegler (2012), from the reduced Hessian we calculate $\mathbf{P}(t_{k-H+1}|t_k)$, which is used at t_{k+1} for updating the arrival cost (Equation (A.3)).

References

Bard, Y., 1974. Nonlinear parameter estimation. Academic Press, New York (U.S.A.).

Bascone, D., Galvanin, F., Shah, N., Garcia-Munoz, S., 2020. A hybrid mechanistic-empirical approach to the modelling of twin screw feeders for continuous tablet manufacturing. Ind. Eng. Chem. Res. https://doi.org/10.1021/acs.iecr.0c00420.

Biegler, L.T., 2010. Nonlinear programming: concepts, algorithms, and applications to chemical processes. Society for Industrial and Applied Mathematics.

Biegler, L.T., 2007. An overview of simultaneous strategies for dynamic optimization. Chem. Eng. Process. Process Intensif. 46, 1043–1053. https://doi.org/10.1016/j.cep.2006.06.021.

Blackshields, C.A., Crean, A.M., 2018. Continuous powder feeding for pharmaceutical solid dosage form manufacture: a short review. Pharm. Dev. Technol. 23, 554–560. https://doi.org/10.1080/10837450.2017.1339197.

Cartwright, J.J., Robertson, J., D'Haene, D., Burke, M.D., Hennenkamp, J.R., 2013. Twin screw wet granulation: Loss in weight feeding of a poorly flowing active pharmaceutical ingredient. Powder Technol. 238, 116–121. https://doi.org/10.1016/j.powtec.2012.04.034.

De Beer, T., Burggraeve, A., Fonteyne, M., Saerens, L., Remon, J.P., Vervaet, C., 2011. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. Int. J. Pharm. 417, 32–47. https://doi.org/10.1016/j.ijpharm.2010.12.012.

Destro, F., Facco, P., Muñoz, S.G., Bezzo, F., Barolo, M., 2020. A hybrid framework for process monitoring: Enhancing data-driven methodologies with state and parameter estimation. J. Process Control 92, 333–351. https://doi.org/10.1016/j.jprocont.2020.06.002.

Engisch, W.E., Muzzio, F.J., 2012. Method for characterization of loss-in-weight feeder equipment. Powder Technol. 228, 395–403. https://doi.org/10.1016/j.powtec.2012.05.058.

Facco, P., Santomaso, A.C., Barolo, M., 2017. Artificial vision system for particle size characterization from bulk materials. Chem. Eng. Sci. 164, 246–257. https://doi.org/10.1016/j.ces.2017.01.053.

Fisher, A.C., Lee, S.L., Harris, D.P., Buhse, L., Kozlowski, S., Yu, L., Kopcha, M., Woodcock, J., 2016. Advancing pharmaceutical quality: An overview of science and research in the U.S. FDA's Office of Pharmaceutical Quality. Int. J. Pharm. 515, 390–402. https://doi.org/10.1016/j.ijpharm.2016.10.038.

Food & Drug Administration, 2004. PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing. and Quality Assurance 1–32.

Galata, D.L., Mészáros, L.A., Kállai-Szabó, N., Szabó, E., Pataki, H., Marosi, G., Nagy, Z. K., 2021. Applications of machine vision in pharmaceutical technology: a review. Eur. J. Pharm. Sci. 159, 105717 https://doi.org/10.1016/j.ejps.2021.105717.

García-Muñoz, S., Butterbaugh, A., Leavesley, I., Manley, L.F., Slade, D., Bermingham, S., 2017. A flowsheet model for the development of a continuous process for pharmaceutical tablets: An industrial perspective. AIChE J. 64, 511–525. https://doi.org/10.1002/aic.15967.

Gyürkés, M., Madarász, L., Köte, Á., Domokos, A., Mészáros, D., Beke, Á.K., Nagy, B., Marosi, G., Pataki, H., Nagy, Z.K., Farkas, A., 2020. Process design of continuous powder blending using residence time distribution and feeding models. Pharmaceutics 12, 1–20. https://doi.org/10.3390/pharmaceutics12111119.

Hanson, J., 2018. Control of a system of loss-in-weight feeders for drug product continuous manufacturing. Powder Technol. 331, 236–243. https://doi.org/10.1016/j.powtec.2018.03.027.

Hart, W.E., Laird, C.D., Watson, J.-P., Woodruff, D.L., Hackebeil, G.A., Nicholson, B.L., Siirola, J.D., 2017. Pyomo-optimization modeling in python. Springer, New York (U.S.A.).

Haseltine, E.L., Rawlings, J.B., 2005. Critical evaluation of Extended Kalman Filtering and Moving-Horizon Estimation 2451–2460. https://doi.org/10.1021/ie034308l.

Lee, S.L., O'Connor, T.F., Yang, X., Cruz, C.N., Chatterjee, S., Madurawe, R.D., Moore, C. M.V., Yu, L.X., Woodcock, J., 2015. Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production. J. Pharm. Innov. 10, 191–199. https://doi.org/10.1007/s12247-015-9215-8.

Liu, J., Su, Q., Moreno, M., Laird, C., Nagy, Z., Reklaitis, G., 2018. Robust state estimation of feeding–blending systems in continuous pharmaceutical manufacturing. Chem. Eng. Res. Des. 134, 140–153. https://doi.org/10.1016/j.cherd.2018.03.017.

López-Negrete, R., Biegler, L.T., 2012. A Moving Horizon Estimator for processes with multi-rate measurements: A Nonlinear Programming sensitivity approach. J. Process Control 22, 677–688. https://doi.org/10.1016/j.jprocont.2012.01.013.

Nicholson, B., López-Negrete, R., Biegler, L.T., 2014. On-line state estimation of nonlinear dynamic systems with gross errors. Comput. Chem. Eng. 70, 149–159. https://doi.org/10.1016/j.compchemeng.2013.11.018.

Nicholson, B., Siirola, J.D., Watson, J.-P., Zavala, V.M., Biegler, L.T., 2018. pyomo. dae: A modeling and automatic discretization framework for optimization with differential and algebraic equations. Math. Program. Comput. 10, 187–223. https://doi.org/10.1007/s12532-017-0127-0.

Pirnay, H., López-Negrete, R., Biegler, L.T., 2012. Optimal sensitivity based on IPOPT. Math. Program. Comput. 4, 307–331. https://doi.org/10.1007/s12532-012-0043-2.

Plumb, K., 2005. Continuous processing in the pharmaceutical industry: Changing the mind set. Chem. Eng. Res. Des. 83, 730–738. https://doi.org/10.1205/cherd.04359.

Rao, C. V., 2000. Moving horizon strategies for the constrained monitoring and control of nonlinear discrete-time systems.

Rao, C.V., Rawlings, J.B., 2002. Constrained process monitoring: moving-horizon approach. AIChE J. 48 (1), 97–102. https://doi.org/10.1002/aic.690480111.

Rawlings, J.B., Miller, S.M., Witkowski, W.R., 1993. Model Identification and Control of Solution Crystallization Processes: A Review. Ind. Eng. Chem. Res. 32, 1275–1296. https://doi.org/10.1021/ie00019a002.

Ray, W.H., 1981. Advanced process control. McGraw-Hill, New York (U.S.A.).

Rehrl, J., Kruisz, J., Sacher, S., Khinast, J., Horn, M., 2016. Optimized continuous pharmaceutical manufacturing via model-predictive control. Int. J. Pharm. 510, 100–115. https://doi.org/10.1016/j.ijpharm.2016.06.024.

Seborg, D.E., Edgar, T.F., Mellichamp, D.A., Doyle, F.J., 2017. Process Dynamics and Control, 4th ed. John Wiley & Sons, New York (U.S.A.).

Shi, Z., Hermiller, J., Muñoz, S.G., 2019. Estimation of mass-based composition in powder mixtures using Extended Iterative Optimization Technology (EIOT). AIChE J. 65, 87–98. https://doi.org/10.1002/aic.16417.

Singh, R., Sahay, A., Muzzio, F., Ierapetritou, M., Ramachandran, R., 2014. A systematic framework for onsite design and implementation of a control system in a continuous tablet manufacturing process. Comput. Chem. Eng. 66, 186–200. https://doi.org/10.1016/j.compchemeng.2014.02.029.

Su, Q., Reklaitis, G.V., Nagy, Z.K., 2020. Continuous Feeding-Blending in Pharmaceutical Continuous Manufacturing. Continuous Pharm. Process. Springer 193–226.

Wächter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. 106, 25–57. https://doi.org/10.1007/s10107-004-0559-y.

Zavala, V.M., Laird, C.D., Biegler, L.T., 2008. A fast moving horizon estimation algorithm based on nonlinear programming sensitivity. J. Process Control 18, 876–884. https://doi.org/10.1016/j.jprocont.2008.06.003.