Grey-Box Approach for the Prediction of Variable Residence Time Distribution in Continuous Pharmaceutical Manufacturing

A. Elkhashap [*] R. Meier [**] D. Stenger [*] D. Abel [*]

[*] Institute of Automatic Control, RWTH Aachen University,
Aachen, Germany (e-mail: a.elkhashap@irt.rwth-aachen.de).
[**] L.B. Bohle Maschinen + Verfahren GmbH,
Ennigerloh, Germany (e-mail: R.Meier@lbbohle.de)

Abstract: Axial dispersion models are used for the prediction of residence time distribution (RTD) of the flow occurring in various processes. Such models are essential for the understanding of the flow dynamics allowing monitoring, control and material tracing specially in the scope of continuous pharmaceutical manufacturing. However, RTDs are most usually dependent on the process variables (PVs), indicating that a single constant parameter dispersion model would not be capable of capturing this variability. In this contribution a variable parameter axial dispersion model is proposed, where the dependency on the process variables are captured from experimental data using Gaussian Process Regression (GPR) models. The method is illustrated with an example of a Vibrated Fluidized Bed Dryer (VFBD), in which a number of tracer experiments are performed at different values of the drying process air flow rate and vibration acceleration. The axial dispersion model parameter values are identified for each experiment. Manifolds for the axial dispersion model parameters are then constructed by the regression of the GP models on the identified values. Comparisons between the experiments and model predictions for an example validation case are drawn showing that the proposed model is capable of producing accurate RTD predictions and certainty bounds even for points not explicitly included in regression dataset. Insight about the advantages of the method in model based controller design is given.

Keywords: Axial Dispersion Model, Continuous Manufacturing, Residence Time Distribution, Gaussian Process Regression.

1. INTRODUCTION

Analyzing as well as predicting the flow dynamics in processes is essential for design, monitoring and control tasks. Tracer Residence Time Distribution (RTD) experiments (see Danckwerts (1995) and Fogler (2016)) has proven itself to be a very useful tool for experimentally characterizing the flow behavior for fluids as well as powders appearing in processes present in food (e.g. Torres and Oliveira (1998), Ganjyal and Hanna (2002)), chemical or pharmaceutical Industries. In the scope of continuous pharmaceutical manufacturing, studies such as Gao et al. (2012, 2011); Meier et al. (2016, 2017); Martinetz et al. (2018); Engisch and Muzzio (2016) have been focused on the utilization of the RTD theory and tools for the analysis, prediction and monitoring of a wide range of underlying processes. Various models with different characteristics have been proposed for the prediction of the RTD, see Gao et al. (2012) and Martin (2000). Determining the effect of different process variables and parameters on the process RTD has been studied in the scope of process analysis e.g. Sudah et al. (2002). Attempts to capture such dependencies in models have been made mainly for each case separately e.g. by rigorous modeling in de Graaf et al. (1997), through parametric relations in Bongo Njeng et al.

(2015). Black box models have been also proposed for the direct prediction of RTD e.g. in Ganjyal and Hanna (2002). However, the aforementioned literature lacks a generalized approach to consider process variables dependencies in RTD models. In this contribution, a more generalizable approach for considering such dependencies in RTD modeling is proposed, where a Grey-box axial dispersion model is presented. The functional dependency of the RTD on the process variables is captured by introducing Gaussian Process Regression (GPR) models representing the dispersion parameters. The proposed model is specially suited for automatic control, as both the model formulation and its numerical solution are chosen to be compatible with the design framework of model based controllers. The paper is organized as follows. First, a brief introduction about RTD is given. Then, the proposed model is presented illustrating the analytical (white box) as well as the data driven (black box) parts. Afterwards, the model's numerical solution as well as the used approach for parameter identification are introduced and adopted to an application from the field of continuous pharmaceutical manufacturing. The presented application and its experimental setup are outlined. Finally, effectiveness of the proposed method is demonstrated for an example validation case.

2. RESIDENCE TIME DISTRIBUTION

The residence time of a flowing element is defined to be the time elapsed from its point of injection to its appearance at reactor outlet. However, different fragments of the flowing material have different [1] residence times, which are then represented by a distribution, namely the RTD $E(t)$. In a real world experimental setup (Fogler (2016)), a specific amount of an inert tracer material is injected at the inlet, then the concentration $C_{exp}(t)$ of this tracer is measured at the process outlet over time t starting from the time point of injection. Normalizing $C_{exp}(t)$ by the total discharged concentration delivers the experimentally determined RTD

$$E(t) = \frac{C_{exp}(t)}{\int_0^\infty C_{exp}(t)\,dt}. \tag{1}$$

3. PROPOSED GREY BOX AXIAL DISPERSION MODEL

The main motivation for using a hybrid modeling approach is to preserve the physical interpretation of the model parameters and structure while exploiting experimental data for achieving high prediction accuracy and resolving unknown (uncertain) model dependencies. Accordingly, the model structure can be further extended with new modeling knowledge (rigorous and experimental) to describe more complex systems. The clear model structure offered by the approach allows the generalization of the model to be used for several applications. The proposed grey-box model can be divided in an analytical part and a data driven part. The model's analytical part is based on the classical closed boundary dispersion model. The main reason for using an axial dispersion model is that, it is based on a rigorous modeling methodology as its parameters have a clear physical interpretation. This fits the introduced framework as the proposed grey box model would inherit this generic property of the axial dispersion model. Moreover, it has been steadily reported (e.g. Martin (2000) and Bachmann and Tsotsas (2015)) that it delivers the most accurate RTD predictions in comparison with other alternatives (e.g. tank in series see Fogler (2016)) for a wide range of dispersion (Peclet) numbers. The main disadvantage of the dispersion model is complexity and cost of the numerical solution.

3.1 Analytical Model

The model is represented by a one dimensional conservative (i.e. no sources or sinks) convection diffusion partial differential equation (2) for the global representation (Eulerian) of the material concentration $c(t,z)$ in the control volume

$$\frac{\partial c}{\partial t} = -v(\boldsymbol{\vartheta})\frac{\partial c}{\partial z} + D(\boldsymbol{\vartheta})\frac{\partial^2 c}{\partial z^2}, \quad z \in [0, L], \tag{2}$$

the independent variables t and z represent time and axial distance defined over the length L starting from the control volume inlet (injection point). The flow parameters $v(\boldsymbol{\vartheta})$ and $D(\boldsymbol{\vartheta})$ represents the flow's axial velocity and

dispersion rate respectively and are dependent on the vector

$$\boldsymbol{\vartheta}(t) := (pv_1(t), \cdots, pv_n(t))^\mathsf{T}, \quad \boldsymbol{\vartheta} \in \mathbb{R}^{n \times 1}, \tag{3}$$

which includes all the process variables $pv_1(t), \cdots, pv_n(t)$ having influence on the flow (RTD). The RTD is then the concentration at the outlet boundary $E(t) = c(t,L)$ for a unit impulse concentration at the inlet boundary $c_{in}(t) = \delta(t)$ and zero initial condition $c(0,z) = 0$ for the whole control volume. The inlet and outlet boundary conditions depends on the real system configuration, at which each boundary can be considered either closed or open, see Fogler (2016). The closed system boundary conditions given by (4) (also know as Danckwerts boundary conditions Danckwerts (1995)) are used for our case study.

$$c(t,0) = c_{in}(t) + \frac{D(\boldsymbol{\vartheta})}{v(\boldsymbol{\vartheta})}\frac{\partial c}{\partial z}\bigg|_{z=0},$$
$$\frac{\partial c}{\partial z}\bigg|_{z=L} = 0. \tag{4}$$

The two system parameters contained in the parameter vector $\boldsymbol{p}(t) := (v(\boldsymbol{\vartheta}), D(\boldsymbol{\vartheta}))^\mathsf{T}$, $\boldsymbol{p} \in \mathbb{R}^{2 \times 1}$ are assumed to be continuous scalar functions of the vector $\boldsymbol{\vartheta}$. These functions are then represented as Gaussian Processes (GPs). The two GP models are identified in a regression step using experimental data thus forming the data driven part of the model.

3.2 Gaussian Process Regression Models

In principle, any regression model can be used to approximate the unknown functions $v(\boldsymbol{\vartheta})$ and $D(\boldsymbol{\vartheta})$ given some observations. The machine learning community has developed a wide variety of methods reaching from classical linear regression (e.g. polynomial regression) to more advanced methods such as Gaussian Process Regression or Artificial Neural Networks. The Grey-Box model is designed to be used by a model based controller usually employing gradient based optimization methods. Therefore, the used model should allow for smooth differentiable predictions in a closed analytical form. Additionally, due to the cost of experiments the number of observations is limited. Thus, the model to be chosen should ideally be able to incorporate prior knowledge. Thirdly, RTD experiments have an inherent stochastic character. GPR models represent an ideal candidate due to their capability of capturing, quantifying [2], and deriving a generalized uncertainty prediction measure over the required range of inputs. Consequently, the predictive variance delivered by the model (representing the uncertainty) can also be used for a robust controller design. Below, a brief introduction highlighting the main advantages of GPR for the problem at hand is given. For a detailed explanation, the reader is referred to Rasmussen and Williams (2006).

For a given set of inputs $\boldsymbol{\vartheta}$, the unknown functions v and D are not assumed to be correlated a-priori. Therefore, two independent models for v and D are constructed. In the following m observations with inputs $\mathscr{X} = [\boldsymbol{\vartheta}_1, \cdots, \boldsymbol{\vartheta}_m]$ and targets $\mathscr{Y} = [y_1, y_2, \cdots, y_m]$ equal to either v or D are considered for simplicity. Additionally, it is assumed that all samples (of v or D) are corrupted by a noise

---

[1] This holds true for all but strictly plug flow

[2] Uncertainty at the observations can be estimated from redundant experiments as in Gao et al. (2011)

term $\boldsymbol{\varepsilon}$, i.e. $y = f(\boldsymbol{\vartheta}) + \boldsymbol{\varepsilon}$, where $f(\boldsymbol{\vartheta})$ represents the true unknown function value at a given location $\boldsymbol{\vartheta}$. In a GP it is assumed that any finite set of true function values $\{f_1, f_2, \cdots, f_k, \} = \{f(\boldsymbol{\vartheta}_1), \cdots, f(\boldsymbol{\vartheta}_N)\}$ is distributed according to a multivariate Gaussian distribution

$$(f_i) \sim GP(m(\boldsymbol{\vartheta}, \boldsymbol{\theta}_m), k(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}', \boldsymbol{\theta}_k)).$$

The GP is fully characterized by an a-priori mean $m(\boldsymbol{\vartheta}, \boldsymbol{\theta}_m) = \mathrm{E}[f(\boldsymbol{\vartheta})]$ with hyper parameters $\boldsymbol{\theta}_m$ and a covariance function $k(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}', \boldsymbol{\theta}_k) = \mathrm{E}[\{m(\boldsymbol{\vartheta}, \boldsymbol{\theta}_m) - f(\boldsymbol{\vartheta})\}\{m(\boldsymbol{\vartheta}', \boldsymbol{\theta}_m) - f(\boldsymbol{\vartheta}')\}]$ with hyper parameters $\boldsymbol{\theta}_k$. First, assume that $\boldsymbol{\varepsilon} = 0$. In this case the observed values are identical to the true function values at any given location. The predictive normal distribution of f for a query point $\boldsymbol{\vartheta}_q$ can be calculated by conditioning on the observed function values

$$f(\boldsymbol{\vartheta}_q) \sim \mathcal{N}(\mu_f(\boldsymbol{\vartheta}_q | \mathcal{X}, \mathcal{Y}, \boldsymbol{\theta}_m, \boldsymbol{\theta}_k), \sigma_f^2(\boldsymbol{\vartheta}_q | \mathcal{X}, \mathcal{Y}, \boldsymbol{\theta}_m, \boldsymbol{\theta}_k)),$$

the function prediction value at the query point is then a function of past observations $\mathcal{Y}$ at locations $\mathcal{X}$, as well as the hyper parameters of the GP's mean and covariance functions. For a detailed explanation and formulas for $\mu_f(\boldsymbol{\vartheta}_q | \mathcal{X}, \mathcal{Y}, \boldsymbol{\theta}_m, \boldsymbol{\theta}_k)$ and $\sigma_f^2(\boldsymbol{\vartheta}_q | \mathcal{X}, \mathcal{Y}, \boldsymbol{\theta}_m, \boldsymbol{\theta}_k)$ the reader is referred to Rasmussen and Williams (2006). Neither μ_f nor σ_f^2 relies on the parameterization of a fixed basis functions as in the case of parametric methods, e.g. bayesian linear regression. Consequently, GPR can fit arbitrary continues functions and therefore is considered a non-parametric model. By choosing appropriate mean and covariance functions as well as hyper parameters, the GP is adapted to the problem at hand with the goal of making sensible predictions. In this work a squared exponential kernel with automated relevance detection (SEARD-Kernel)

$$k(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}', \boldsymbol{\theta}_k) = \sigma_0^2 \exp[-\frac{1}{2}(\boldsymbol{\vartheta}' - \boldsymbol{\vartheta})^\mathsf{T} \mathrm{diag}(\boldsymbol{l}_{sc})^{-1}(\boldsymbol{\vartheta}' - \boldsymbol{\vartheta})]$$

is chosen for the calculation of the covariance between any two points $\boldsymbol{\vartheta}'$ and $\boldsymbol{\vartheta}$. The kernel is parametrized by hyper parameters $\boldsymbol{\theta}_k = \{\sigma_0, \boldsymbol{l}_{sc}\}$. Note that with the chosen covariance function the predictive mean is infinitely differentiable (see Rasmussen and Williams (2006)). The scaling factor σ_0 influences the predictive variance at unknown locations. The length scales $\boldsymbol{l}_{sc} = [l_1^2, \cdots, l_n^2]$ determine the relative area of influence of each input dimension for the (measured) sample points. Additionally, the behavior of the model far away from the sample points can be adjusted by selecting a suitable a-priori mean function. In this case an affine mean $m(\boldsymbol{\vartheta}, \boldsymbol{\theta}_m) = a + \boldsymbol{b}_m \cdot \boldsymbol{\vartheta}$ with hyper parameters $\boldsymbol{\theta}_m = \{a, \boldsymbol{b}_m\}$, $\boldsymbol{b}_m = [b_1, \cdots, b_n]^\mathsf{T}$ is chosen. Therefore, it is assumed that the approximated function is linear where no measurements are available or far away from the observations. In the more general case where observations are corrupted by homoscedastic gaussian noise [3] $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ representing the observations uncertainty and with the assumed normal distribution describing the true function value $f \sim \mathcal{N}(\mu_f, \sigma_f^2)$, the predictive distribution of the observed value is $y|f \sim \mathcal{N}(f, \sigma_\varepsilon^2) = \mathcal{N}(\mu_f, \sigma_f^2 + \sigma_\varepsilon^2)$. If a gaussian observation model is used the predictive distribution at an unknown location $\boldsymbol{\vartheta}_q$ remains gaussian

---

[3] Here a gaussian observational model (likelihood) is used. Note that non-normal predictive distributions can also be used with gaussian process regression, see e.g. Rasmussen and Williams (2006) or Vanhatalo et al. (2013).

$$y(\boldsymbol{\vartheta}_q) \sim$$
$$\mathcal{N}(\mu_f(\boldsymbol{\vartheta}_q | \mathcal{X}, \mathcal{Y}, \boldsymbol{\theta}_m, \boldsymbol{\theta}_k, \sigma_\varepsilon), \sigma_f^2(\boldsymbol{\vartheta}_q | \mathcal{X}, \mathcal{Y}, \boldsymbol{\theta}_m, \boldsymbol{\theta}_k) + \sigma_\varepsilon^2).$$

In total the model is defined by the hyper parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_m, \boldsymbol{\theta}_k, \sigma_\varepsilon]$. Usually, these parameters are estimated from data by maximizing the likelihood $p(\mathcal{Y} | \mathcal{X}, \boldsymbol{\theta})$ on the observations. Moreover, each of the parameters has a clear distinguishable influence on the prediction. Thus suitable prior distributions for the hyper parameters (called hyper priors) can be found. An alternative to the determination of the hyper parameters by maximizing the likelihood is to minimize the leave one out cross validation loss (see Sammut, Claude and Webb, Geoffrey I (2010)). This can be beneficial in terms of avoiding over-fitting and achieving practical guarantees on the model prediction in the input range of the training dataset.

4. MODEL SOLUTION AND IDENTIFICATION

The presented method is used for the modeling and prediction of the RTD representing the propagation of moist granulate in a fluidized bed drying process. The drying process represents a subprocess of the wet granulation route in a continuous manufacturing line of solid-dosage form pharmaceutics (see Elkhashap et al. (2019)). The state of the art Vibrated Fluidized Bed Dryer (VFBD) unit under study is part of a lab scale unit (commercially named QbCon®1, see Meier and Emanuele (2018)) developed in the R&D facility of the company L.B. Bohle Maschinen+Verfahren GmbH.

4.1 Model Numerical Solution

The PDE representing the axial dispersion model is solved using Method of Lines (MoL), see Schiesser (2012), where the spatial variable is discretized into n_z points over an equidistant grid with segment length

$$\Delta z = \frac{L}{n_z - 1}. \tag{5}$$

The convection term is discretized using a first order upwind scheme, while the diffusion term is discretized using a second order central scheme. Ghost points are employed at the boundaries in order to eliminate algebraic constraints (see LeVeque (1992) and Abgrall and Shu (2017)). Also, a forward difference scheme is employed at the left boundary to avoid any division by $D(\boldsymbol{\vartheta})$ thus allowing for zero values of the parameter, which corresponds to the lower bound for the dispersion in the system. By defining the state vector $x \in \mathbb{R}^{n_z \times 1}$, the input $u \in \mathbb{R}$, and the output $\tilde{y} \in \mathbb{R}$

$$\begin{aligned} \boldsymbol{x} &:= [c(t,0), c(t,\Delta z), c(t,2\Delta z), \cdots, c(t,L)]^\mathsf{T}, \\ u &:= [c_{in}(t)], \\ \tilde{y} &:= [c(t,L)], \end{aligned} \tag{6}$$

the following Linear Parameter Varying (LPV) state space representation of the dispersion model can be formulated

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{A}(p(t))\,\boldsymbol{x} + \boldsymbol{b}(p(t))\,u, \\ \tilde{y} &= \tilde{\boldsymbol{c}}\boldsymbol{x}, \end{aligned} \tag{7}$$

For the sake of brevity, the dependence on the parameter vector $p(t)$ as well as time t is omitted hereafter. Defining $D_d := \frac{D(\boldsymbol{\vartheta})}{\Delta z^2}$ and $v_d := \frac{v(\boldsymbol{\vartheta})}{\Delta z}$ the system matrices $\boldsymbol{A} \in \mathbb{R}^{n_z \times n_z}$, $\boldsymbol{b} \in \mathbb{R}^{n_z \times 1}$, and $\tilde{\boldsymbol{c}} \in \mathbb{R}^{1 \times n_z}$ are given by

$$A := \begin{bmatrix} -D_d - v_d & D_d & 0 & \dots & 0 & 0 \\ D_d + v_d & -2D_d - v_d & D_d & \dots & 0 & 0 \\ 0 & D_d + v_d & -2D_d - v_d & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & D_d + v_d & -D_d - v_d \end{bmatrix},$$

$$b := [v_d, \, 0, \, \cdots, \, 0]^{\mathsf{T}}, \, \tilde{c} := [0, \, \cdots, \, 0, \, 1].$$

Different time discretization techniques for LPV systems are studied by van den Hof et al. (2010). Here, the classical complete (exact) zero order hold discretization technique (see Franklin et al. (1998)) exploiting the matrix exponential is chosen for the calculation of the discrete equivalent of (7). The discrete equivalent of (7) is essential for its usage in model based controller and observer design approaches (e.g. Model Predictive Control, Kalman Filter), or in real time tracing cases involving forward prediction (i.e. outlet concentration predictions given arbitrary sampled $c_{in}(t)$). Finally, the predicted RTD function $E_m(t)$ corresponds to the impulse response of (7), namely $E_m(t) = \tilde{y}(t)$ for $c_{in}(t) = \delta(t)$. Also, note that setting the process variables $\boldsymbol{\vartheta}$ constant (as in the identification experiments), produces a constant parameters vector $\boldsymbol{p}$ and consequently a constant matrix pair A and b, which reduces (7) to a Linear Time Invariant (LTI) system.

4.2 Experiments

In order to characterize the flow behavior of the wet granules inside the VFBD a number of experiments are performed. The two main process variables influencing the flow were determined to be the drying air mass flow rate[4] $\dot{m}_{air}$ (Nm³/hr) and the vibration acceleration a_{vib} (m/s²) indicating that $\boldsymbol{\vartheta} := (\dot{m}_{air}, a_{vib})^{\mathsf{T}}$. Nine RTD experiments are preformed at different value combinations of the two process variables. Inert tracer material is injected at the dryer input. The concentration of the tracer was measured at the output using near-infrared spectroscopy while holding the two PVs constant at their chosen values. The nine experimental RTD curves are shown in Fig. 1 in the identification results. Moreover, in Table 1 the set PVs values for each experiment can be found. Next, these experimental RTDs are used to produce sample points of the functions representing the two parameters $v(\boldsymbol{\vartheta})$ and $D(\boldsymbol{\vartheta})$ to be used in the regression step. This is done by identifying the values of the parameters for each RTD experiment.

4.3 Model Parameter Identification

The model parameters v and D are identified for each experiment by minimizing the square of the error between the experimental RTD curve $E(t)$ and the numerically predicted by the model $E_m(t)$. The identification problem is solved using Matlab nonlinear programming solver. For the purpose of a warm start for the solver, estimations for the parameter values are calculated using the method of moments mentioned in Bachmann and Tsotsas (2015). The experimental RTD's first three moments are calculated determining the distribution's mean (consequently the estimate for v) and variance. Estimate for the dispersion coefficient D is calculated from solving the implicit relation given by Levenspiel (1999). The results of the identified

[4] normalized cubic meter per hour

Table 1. The used PVs values and the corresponding identified parameter values for each experiment

Tracer Exp	$\boldsymbol{\vartheta}$		$v(\boldsymbol{\vartheta})$ (m/hr)	$D(\boldsymbol{\vartheta}) \times 10^3$ (m²/hr)
	$\dot{m}_{air}$ (Nm³/hr)	a_{vib} (m/s²)		
Exp1	15	3.7	15.5	7.3
Exp2	15	4.22	26.9	30.8×10^{-8}
Exp3	15	5.2	51.0	43.1
Exp4	20	3.7	22.3	48×10^{-3}
Exp5	20	4.22	37.2	15.7
Exp6	20	5.2	69.3	96.1
Exp7	25	3.7	42.1	72.4
Exp8	25	4.22	57.8	90.6
Exp9	25	5.2	82.4	304.4
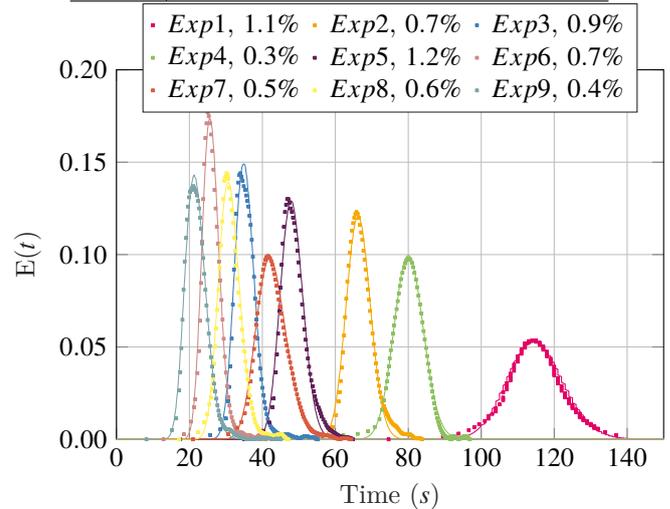
Fig. 1. RTDs determined from the nine tracer experiments (marker) and the predicted RTD (solid) using the identified values. NMSE percentage for each experiment is shown in plot legend

parameters at the corresponding process variables are shown in Table 1. A comparison between the experimental RTD and the predicted RTD using the identified values for each experiment is depicted in Fig. 1. The Normalized Mean Square Error (NMSE) between both is listed in the legend. It should be noted that the identified $D(\boldsymbol{\vartheta})$ values are dependent on the spatial discretization (5) due to the numerical diffusion added by the upwind finite difference scheme. This lays restriction on the model used for prediction since it should be with the same spatial discretization as the model used for identification. Otherwise, correction factors for the numerically modified diffusion due to different spatial grid can be derived from relations determining the amount of numerical diffusion (e.g. given by E. Ewing and Wang (2001)). The identified parameter values at the corresponding PVs values are then used in the regression step to characterize the GP models for $v(\boldsymbol{\vartheta})$ and $D(\boldsymbol{\vartheta})$.

5. RESULTS

The available RTD experimental data set is small (nine points regarding two dimensional input), which does not allow a classical cross validation data split. Using all the experimental data in the regression step yields the most accurate GPR models for the prediction of the two parameters of the axial dispersion model. However, to show the

model capability of delivering valid parameter predictions along with accurate uncertainty estimates at points not present in the regression data set, two experiments are excluded from the regression data set and used for validating the GPR models. As an illustrative example, experiment 5 and 7 were chosen to be excluded and the other seven points were used as the regression data. The GPR models have an input dimension of two allowing a visualization of the models. Fig. 2 shows the prediction mean surrounded by two surfaces representing the uncertainty estimate $\pm 2\sigma$ predicted by the GPR models of the two parameters $v(\vartheta)$ and $D(\vartheta)$. Fig. 3 shows the predictions produced by the GPR models against the experimental data, where the predicted mean as well as the uncertainty estimate of the $E(t)$ functions are calculated and compared to the experimental RTD at the two excluded points. The NMSE of the mean predictions corresponding to the excluded experiments $Exp5$ and $Exp7$ are 14.04% and 4.81% respectively. Moreover, it can be seen that for both points the experimental RTD lies within the predicted uncertainty ($\pm 2\sigma$).

6. DISCUSSION AND CONCLUSION

As shown by the previous example, GPR models are capable of giving accurate mean predictions as well as certainty bounds for the axial dispersion model parameters even at points not explicitly included in the experimental data used for regression. More systematic methods (e.g. leave one out cross validation) can be used, in order to heuristically ensure that the GPR models predictive mean error will always lie within the predictive variance for the input range covered by the regression data set. Suitable hyper-parameters adaptation techniques can be formulated accordingly. In comparison to other alternatives, the employed GPR method has numerous advantages such as the high prediction accuracy, conformity with gradient optimization framework, flexibility and online adaptation potential. Moreover, the proposed hybrid approach preserving the axial dispersion model rigorous structure can be further extended with physical modeling knowledge to describe more complex systems without losing the prediction accuracy gained from the experimental data. The
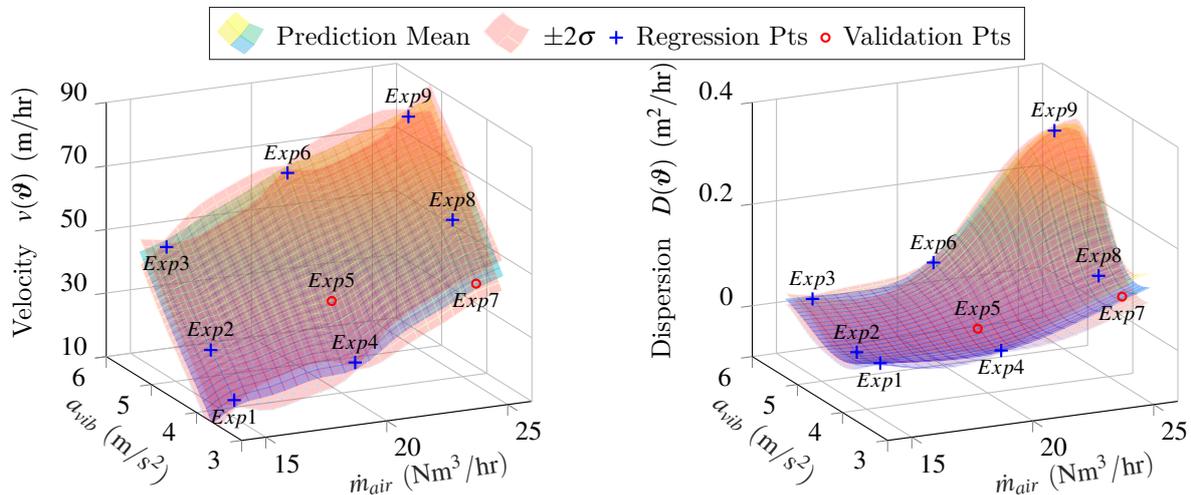
Fig. 2. The two GP models representing the two parameters $v(\vartheta)$ and $D(\vartheta)$. The Prediction mean and variance ($\pm 2\sigma_v$ and $\pm 2\sigma_D$) are shown along with the points used for regression ($+$ marker) and the two excluded points (o marker)
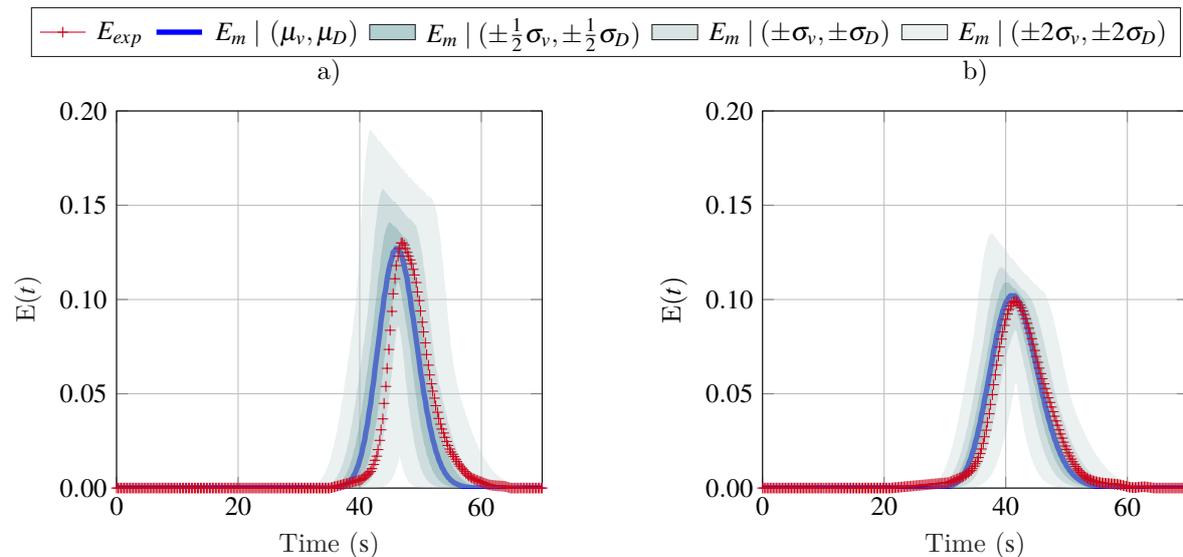
Fig. 3. RTD predictive mean $E_m(t)$ (solid) against experimentally determined RTD $E_{exp}(t)$ ($+$ marker) for the two excluded points a) $Exp5$ and b) $Exp7$. The predicted variance is represented by the shaded areas around the predicted mean

derived GPR models representing parameters in the grey-box model configuration can be cast into differentiable analytic closed form expressions which can be efficiently used in a model based controller design with gradient based optimization. In Elkhashap et al. (2020) the model is extended in order to describe the full transient operation of the VFBD. In Elkhashap et al. (2019) a Nonlinear Model Predictive Controller (NMPC) employing the model is realized and tested in closed loop simulations. For future work, model order reduction techniques are to be investigated to ensure the models computational feasibility and the method is to be tested on different processes also for online tracing configuration.

REFERENCES

Abgrall, R. and Shu, C.W. (2017). Handbook of numerical methods for hyperbolic problems: Applied and modern issues / edited by Rémi Abgrall and Chi-Wang Shu, volume volume 18 of Handbook of numerical analysis. North Holland is an imprint of Elsevier, Amsterdam, Netherlands.

Bachmann, P. and Tsotsas, E. (2015). Analysis of residence time distribution data in horizontal fluidized beds. Procedia Engineering, 102, 790–798.

Bongo Njeng, A.S., Vitu, S., Clausse, M., Dirion, J.L., and Debacq, M. (2015). Effect of lifter shape and operating parameters on the flow of materials in a pilot rotary kiln: Part i. experimental rtd and axial dispersion study. Powder Technology, 269, 554–565.

Danckwerts, P.V. (1995). Continuous flow systems. distribution of residence times. Chemical engineering science, 50(24), 3857–3866.

de Graaf, R.A., Rohde, M., and Janssen, L. (1997). A novel model predicting the residence-time distribution during reactive extrusion. Chemical Engineering Science, 52(23), 4345–4356.

E. Ewing, R. and Wang, H. (2001). A summary of numerical methods for time-dependent advection-dominated partial differential equations. Journal of Computational and Applied Mathematics, 128(1), 423–445.

Elkhashap, A., Meier, R., and Abel, D. (2019). Modelling and control of a continuous vibrated fluidized bed dryer in pharmaceutical production. In VDI Kongress Automation, VDI-Berichte, 159–174. VDI-Verlag GmbH, Duesseldorf.

Elkhashap, A., Meier, R., and Abel, D. (2020). A grey box distributed parameter model for a continuous vibrated fluidized bed dryer in pharmaceutical manufacturing. European Control Conference (ECC) 2020. (Accepted).

Engisch, W. and Muzzio, F. (2016). Using residence time distributions (rtds) to address the traceability of raw materials in continuous pharmaceutical manufacturing. Journal of Pharmaceutical Innovation, 11(1), 64–81.

Fogler, H.S. (2016). Elements of chemical reaction engineering. Prentice Hall international series in the physical and chemical engineering sciences. Prentice Hall, Boston, fifth edition edition.

Franklin, G.F., Powell, J.D., and Workman, M.L. (1998). Digital control of dynamic systems. Addison-Wesley, Menlo Park, Calif. and Reading, 3rd ed. edition.

Ganjyal, G. and Hanna, M. (2002). A review on residence time distribution (rtd) in food extruders and study on the potential of neural networks in rtd modeling. Journal of Food Science, 67(6), 1996–2002.

Gao, Y., Muzzio, F.J., and Ierapetritou, M.G. (2012). A review of the residence time distribution (rtd) applications in solid unit operations. Powder Technology, 228, 416–423.

Gao, Y., Vanarase, A., Muzzio, F., and Ierapetritou, M. (2011). Characterizing continuous powder mixing using residence time distribution. Chemical Engineering Science, 66(3), 417–425.

Levenspiel, O. (1999). Chemical reaction engineering. Wiley, New York and Chichester, 3rd ed. edition.

LeVeque, R.J. (1992). Numerical Methods for Conservation Laws. Lectures in Mathematics ETH Zürich, Department of Mathematics Research Institute of Mathematics. Birkhäuser Basel, Basel, second edition edition.

Martin, A.D. (2000). Interpretation of residence time distribution data. Chemical Engineering Science, 55(23), 5907–5917.

Martinetz, M.C., Karttunen, A.P., Sacher, S., Wahl, P., Ketolainen, J., Khinast, J.G., and Korhonen, O. (2018). Rtd-based material tracking in a fully-continuous dry granulation tableting line. International journal of pharmaceutics, 547(1-2), 469–479.

Meier, R. and Emanuele, D. (2018). Kontinuierliche feuchtgranulierung und wirbelschichttrocknung: Experimentelle untersuchung eines neuartigen, revolutionaeren systems prozesstechnik. TechnoPharm, 8(3), 124.

Meier, R., Moll, K.P., Krumme, M., and Kleinebudde, P. (2017). Impact of fill-level in twin-screw granulation on critical quality attributes of granules and tablets. European Journal of Pharmaceutics and Biopharmaceutics, 115, 102–112.

Meier, R., Thommes, M., Rasenack, N., Moll, K.P., Krumme, M., and Kleinebudde, P. (2016). Granule size distributions after twin-screw granulation – do not forget the feeding systems. European Journal of Pharmaceutics and Biopharmaceutics, 106, 59–69.

Rasmussen, C.E. and Williams, C.K.I. (2006). Gaussian processes for machine learning. Adaptive computation and machine learning. MIT, Cambridge, Mass. and London.

Sammut, Claude and Webb, Geoffrey I (ed.) (2010). Leave-One-Out Cross-Validation, 600–601. Springer US, Boston, MA. doi:10.1007/978-0-387-30164-8_469.

Schiesser, W.E. (2012). The Numerical Method of Lines: Integration of Partial Differential Equations.

Sudah, O.S., Chester, A., Kowalski, J., Beeckman, J., and Muzzio, F. (2002). Quantitative characterization of mixing processes in rotary calciners. Powder Technology, 126(2), 166–173.

Torres, A. and Oliveira, F. (1998). Residence time distribution studies in continuous thermal processing of liquid foods: a review. Journal of Food Engineering, 36(1), 1–30.

van den Hof, P., Heuberger, P., and Tóth, R. (2010). Discretisation of linear parameter-varying state-space representations. IET Control Theory & Applications, 4(10), 2082–2096.

Vanhatalo, J., Riihimäki, J., Hartikainen, J., Jylänki, P., Tolvanen, V., and Vehtari, A. (2013). Gpstuff: Bayesian modeling with gaussian processes. Journal of Machine Learning Research, 14(Apr), 1175–1179.