Gray-Box Modeling of the Molecular Weight Distribution in a Batch Polymerization Reactor

Balazs Bordas*, Kutup Kurt, Andreas Bamberg, and Sebastian Engell

DOI: 10.1002/cite.202200234

We introduce a gray-box approach for modeling the molecular weight distribution in step-growth polymerization reactions using the aggregation population balance equation. The approach is based on extracting a data-based kernel function from in-process measurements of the molecular weight distribution. The method is applied to historical data from an industrial batch polymerization reactor. The resulting model is used for decision support in production by predicting the reaction endpoint corresponding to a target molecular weight. The accuracy of the predictions proved to be sufficient for the deployment of the method.

Keywords: Aggregation, Decision support, Gray-box, Kernel estimation, Polycondensation, Population balance equation, Step-growth polymerization

Received: December 16, 2022; *revised:* March 23, 2023; *accepted:* May 03, 2023

1 Introduction

Modeling the evolution of the structure, chain-length distribution and composition of polymers during their synthesis is of high interest for specialty chemical companies for optimizing and monitoring the product quality. However, white-box (first-principles) modeling of polymerization reactions involving novel chemistries is a time-consuming process that requires a deep understanding of the specific polymer chemistry and leads to complex mathematical representations and the need for sophisticated numerical methods. An alternative is the application of data-based modeling methods for some parts of the model – called gray-box or hybrid modeling – that can be considered as a compromise between model transparency and resource investment.

The quality of polymer products is typically characterized by nuclear magnetic resonance (NMR) and gel permeation chromatography (GPC) measurements, both of which are time-consuming to perform at high fidelity. GPC is currently the only method that can determine the shape of the molecular weight distribution (MWD) of dissolved polymers, which is a highly important property with a strong influence on the performance of the material in applications.

Due to the inherent time-dependency of batch processes, a dynamic representation is generally preferred for online model applications. Black-box dynamic modeling methods such as nonlinear autoregressive models or recurrent neural networks require a high amount of training data with sufficient variation to reliably predict the evolution of a dynamic system. When the target is to predict polymer quality-related variables such as the MWD, this requirement is rarely fulfilled.

Gray-box modeling methods aim at decreasing the reliance on data by incorporating known physical relationships between process variables into the model. By doing so, faster model development on the one hand and a higher level of interpretability on the other hand can be achieved compared to purely white-box or black-box approaches [1]. Two main configurations of combined physical and data-based models can be identified from past work [2–4]. The parallel gray-box configuration, also known as residual modeling, employs a data-based model to learn and correct the prediction errors of a physical model. This configuration is particularly suitable in situations where a complete but imperfect physical model is available, for example, from the process design phase. In the serial gray-box configuration, either the predictions of a data-based model are fed as inputs to a physical model, or the outputs of a physical model are used as predictor variables in a data-based model. This strategy can be applied when a first-principles model is not available and the physical understanding of the process is incomplete, which is a significant advantage over the parallel configuration. The combination of the two methods is also possible.

Scheffold et al. tested a gray-box approach for a benchmark styrene polymerization problem, using sparse regres-

[1,2]Balazs Bordas https://orcid.org/0000-0002-7050-930X (balazs.bordas@tu-dortmund.de), [2]Dr. Kutup Kurt, [2]Dr. Andreas Bamberg, [1]Prof. Dr.-Ing. Sebastian Engell

[1]Technical University Dortmund, Process Dynamics and Operations Group, Emil-Figge-Straße 70, 44227 Dortmund, Germany.

[2]Merck KGaA, Frankfurter Straße 250, 64293 Darmstadt, Germany.

sion to identify and fit suitable regressors for various subcomponents of the model (reaction kinetics, heat transfer, thermodynamic assumptions) [5]. The model is intended for application in nonlinear model predictive control of the reactor temperature; therefore, the only quality-related output variable is the monomer conversion. Bouaswaig et al. report a gray-box approach for data-driven soft sensing of quality parameters in a radical polymerization reaction using multiway partial least squares regression [6]. Here, the white-box component is a fully parametrized, rigorous process model, therefore the described approach is not readily applicable to cases where a kinetic model is not available. The authors of [7] use a nonlinear finite impulse response technique for the prediction of the polymer extrusion viscosity, where the pool of basis functions is selected based on physical understanding of the properties of the polymer melt and the process. Hayian et al. combine an orthogonal polynomial neural network and a recurrent neural network to model the MWD using data from a rigorous model of a continuous stirred tank styrene polymerization [8]. They derive a relationship between the weights of the orthogonal polynomial network and the moments of the MWD and test the application of the model for model-based control in computer experiments.

In the present work. we apply a previously developed serial gray-box modeling method for step-growth polymerization [9] to provide decision support for the operators of an industrial batch polymerization reactor at Merck KGaA, Darmstadt, Germany. The serial configuration was chosen because there was no model available for this reaction from the engineering phase, and the complexity of branched step-growth polymerization kinetics would make the development of a custom physical model very time-consuming. The method is based on extracting a data-based kernel function of the aggregation population balance equation (PBE) from available GPC measurements using nonlinear optimization. We discuss the effects of scaling in the loss function and introduce an estimation strategy that improves the model accuracy at higher molecular weights and prevents numerical instabilities. Finally, we apply the modeling strategy to predict the reaction endpoint and evaluate the prediction performance on the historical data.

2 Methodology

2.1 Step-Growth Polymerization Kinetics

Step-growth polymerization processes are driven by the reaction of the available functional groups of monomers and polymers of various chain lengths. This results in rapid monomer conversion and a slow increase of the molecular weight in the early phase of the reaction. A typical illustration of the step-growth mechanism is shown in Fig. 1. According to IUPAC recommendations [10], the term "polycondensation" should be used when each reaction between the functional groups forms a low molecular weight condensation byproduct, whereas "polyaddition" describes the case when the reaction forms no byproduct. Our proposed approach applies to both cases; therefore we use the term "step-growth polymerization" in the rest of the paper for simplicity. Another type of differentiation can be made between homopolymerization processes using a single type of monomer, and copolymerization, where multiple monomer types are used as raw materials.

The chemical structure of the monomer has a significant influence on the properties of the polymer and the controllability of the reaction. Step-growth polymerization of certain monomer types for long reaction times can lead to strong cross-linking and the formation of a highly viscous gel phase, where the molecular weight is effectively infinite. Therefore, a practical way to categorize step-growth processes is to look at the number and type of functional groups that belong to the reacting monomer or monomers. The type of functional group is denoted by capital letters, and the number of a specific functional group per monomer is indicated by the lower index. The application of this notation to common categories of step-growth polymerization is summarized in Tab. 1.

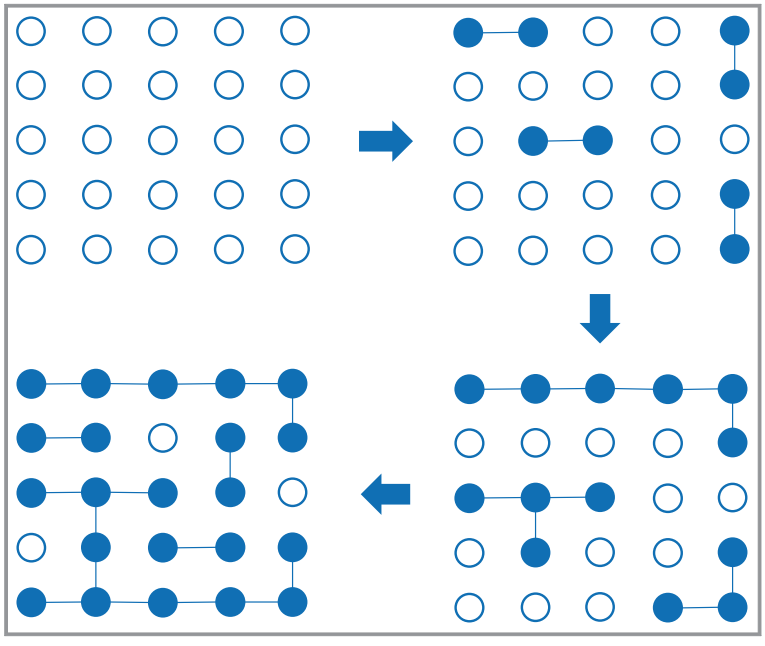

Figure 1. Illustration of the general step-growth mechanism with a three-functional monomer. Empty circles represent monomers, filled circles represent repeating units in a polymer chain.

Table 1. Categorization of common step-growth polymerization systems using the number and type of functional groups of the monomers, based on [11–13].

Monomer type	Type of polymerization	Polymer structure	Gelation
AB or A_2	Homopolymerization	Linear	No gel formation
AB_n	Homopolymerization	Branched	No gel formation
A_n	Homopolymerization	Branched	Gel formation
$A_2 + B_2$	Copolymerization	Linear	No gel formation
$A_n + B_m$	Copolymerization	Branched	Gel formation

2.2 Population Balance for Step-Growth Processes

The applicability of the Smoluchowski equation (the discrete form of the aggregation PBE) to step-growth polymerization kinetics has been established in several articles [14–17]. The discrete model is a system of nonlinear ordinary differential equations (ODE):

$$\frac{\mathrm{d}N_i(t)}{\mathrm{d}t} = \frac{1}{2}\sum_{j=1}^{i-1} K_{j,i-j}N_j(t)N_{i-j}(t) - N_i(t)\sum_{j=1}^{\infty} K_{i,j}N_j(t) \qquad (1)$$

where $i \in \mathbb{N}$, $N_i(t)$ is the number of chains of length i in the system at time t and $K_{i,j}$ denotes the rate of aggregation between chains of lengths i and j. In the context of polymerization reactions, $K_{i,j}$ is a chain length dependent reaction rate parameter.

The continuous aggregation PBE is a partial integro-differential equation:

$$\frac{\partial n(x,t)}{\partial t} = \frac{1}{2}\int_0^x \kappa(x, x-y)n(x-y,t)n(x,t)\mathrm{d}y - n(x,t)\int_0^{\infty} \kappa(x,y)n(y,t)\mathrm{d}y \qquad (2)$$

where x is the continuously distributed property, $n(x,t)$ is the number density function at time t and $\kappa(x,y)$ is the continuous kernel function. Both forms are essentially mass balances on a semi-infinite domain, where the only process that modifies the number of elements of a species is the combination of two species into one, with the assumption that their modeled property is additive in the combination. In most numerical solution methods, the semi-infinite domain is truncated to a maximal considered value to obtain a tractable system. Owing to their generality, these models have found a wide range of applications [18].

The key element of these equations is the kernel function $K_{i,j}$ or $\kappa(x,y)$, which can take various forms depending on the modeled system. Ziff [14] has established three kernels for Eq. (1) on a purely statistical basis. A step-growth polymerization is ideal if the reactivities of all functional groups are equal. When this property is satisfied, the reactivity of a chain is solely defined by how many functional groups are available for the reaction. For ideal homopolymerization processes, the polymerization of bifunctional monomers (AB or A_2) leads to a constant kernel function ($K_{i,j} = C$), the polymerization of monomers of type AB_n to a sum kernel function ($K_{i,j} \propto i + j$) and monomers of type A_n to a product kernel function ($K_{i,j} \propto ij$).

However, real polymerizations deviate from the ideal theory due to several factors. The reactivity of a functional group may change after a neighboring group has reacted, known as the substitution effect. The reactivity may also change with the chain length or time due to mobility constraints and the potential presence of intramolecular reactions (cyclization) can further complicate the situation. In order to use the population balance model for arbitrary nonideal step-growth polymerization reaction, more complicated kernels are needed. These are however very difficult to derive from physico-chemical considerations. A practical option for the case where sufficient data is available is to estimate the kernel from measurement data.

2.3 Gray-Box Model Training

The embedding of the kernel function into the mass balance in (1) or (2) leads to a serial or embedded gray-box approach, where the output of a data-based model that represents the kernel function is used as an input of the physical mass balance. Winz et al. [1] describe a modeling methodology for embedded gray-box systems, which consists of first estimating the outputs of the embedded models from the available data, followed by a decoupled training of the data-based models and finally an integrated parameter estimation step. This decoupling allows for a more informed selection of the embedded model structure since the computational cost of the decoupled estimation and training steps are often lower than repeatedly solving an integrated parameter estimation step with different model candidates. Furthermore, the stability of the training process is significantly improved, as the integrated parameter estimation of the embedded model with randomly initialized parameters can lead to numerical issues. The benefits of the methodology have been recently demonstrated on the nonlinear dynamic modeling of a fermentation process with pilot-scale experimental data [19].

We adopt a similar approach in which the estimation of the values of the kernel function is performed first to obtain a training dataset, followed by fitting a data-based model to this data to obtain the kernel function. The final integrated parameter estimation step is not implemented in this work. To facilitate the training data estimation step, we apply an efficient discretization and a kernel estimation method described in [20] and [21]. This method is based on the idea of using separable kernel functions:

$$\kappa(x,y) \approx \sum_{\nu=1}^{k} \alpha_\nu(x)\beta_\nu(y) \tag{3}$$

where αv and βv are suitable functions, and k is called the separation rank. Substituting this expression and discretizing on a uniform grid, the computationally expensive source term in Eq. (2) can be evaluated using fast Fourier transform, reducing the time complexity of evaluation from quadratic to quasi-linear. In the discrete domain, they use piecewise constant approximations $\mathbf{N} \in \mathbb{R}^{n}$ to $n(x,t)$ and $\mathbf{K} \in \mathbb{R}^{n*n}$ to $\kappa(x,y)$, where n is the number of size classes. $\mathbf{K}$ is then decomposed in the following way:

$$\mathbf{K} = \mathbf{U}\mathbf{S}\mathbf{U}^{T} \tag{4}$$

where $\mathbf{U} \in \mathbb{R}^{n*k}$ and $\mathbf{S} = \mathbf{S}^{T} \in \mathbb{R}^{k*k}$. Throughout our work, a separation rank $k = 1$ and $\mathbf{S} = 1$ was used since no improvements were observed at different values.

The kernel estimation problem using MWD data can be formulated in the following general way:

$$\begin{aligned} &\min_{\mathbf{U}} \; J\left(\mathbf{N}_{\text{meas}}, \hat{\mathbf{N}}\right) \\ &s.t. \; \frac{d\mathbf{N}}{dt} = \mathbf{f}(\mathbf{N}, \; \mathbf{U}) \\ &\mathbf{N}(t_0) = \mathbf{N}_{\text{meas}}(t_0) \\ &\hat{\mathbf{N}} = \mathbf{N} \end{aligned} \tag{5}$$

where $\mathbf{N}$ is the discretized number density, $\mathbf{f}$ is given by the discretization method and J denotes the cost function. The choice of the cost function will be discussed in Sect. 4.1. To solve (5), one needs an ODE solver and a nonlinear optimization routine since $\hat{\mathbf{N}}$ is a nonlinear function of $\mathbf{U}$.

Once estimates of $\mathbf{U}$, denoted by $\mathbf{U}_{\text{est}}$, are available, a suitable data-based model $\mathbf{g}$ can be trained to map the discretized molecular weight grid and other potentially relevant process parameters to $\mathbf{U}_{\text{est}}$:

$$\begin{aligned} &\min_{\mathbf{p}} \; J\left(\mathbf{U}_{\text{est}}, \hat{\mathbf{U}}\right) \\ &s.t. \; \hat{\mathbf{U}} = \mathbf{g}\left(M_w, \; \mathbf{x}_{\text{process}}, \; \mathbf{p}\right) \end{aligned} \tag{6}$$

where $\mathbf{x}_{\text{process}}$ denotes the process parameters and $\mathbf{p}$ is the parameter vector of the data-based model. The requirements to solve (6) depend on the nature of $\mathbf{g}$. In terms of its output, $\mathbf{g}$ can be a scalar- or vector valued function. The final gray-box model takes the following form:

$$\frac{d\mathbf{N}}{dt} = \mathbf{f}\left(\mathbf{N}, \; \mathbf{g}\left(M_w, \; \mathbf{x}_{\text{process}}\right)\right) \tag{7}$$

3 Description of the Application

3.1 Process Description

The process under consideration is an industrial-scale, batch step-growth polymerization process that is operated at Merck KGaA, Darmstadt, Germany. The reactor volume is filled with a prepolymer in a solvent, which is mixed, heated, and kept at a constant temperature over several hours in the vessel. The condensation byproduct is continuously removed from the reactor to promote the forward condensation reactions. The resulting polymer structure exhibits a high degree of branching and cyclization due to the presence of multifunctional monomers. We investigate two product grades which differ in the setpoint of the reaction temperature.

3.2 Data Preparation

The available dataset consists of recorded data from 42 batches for product grade A and 62 batches for product grade B. Four in-process GPC chromatograms were available for each batch, taken towards the end of the reaction with a sampling interval of 1 h. The last available sample is taken 1–2 h before the expected endpoint of the reaction, since the GPC measurement has a time delay of approximately 1 h. Polystyrene analytical standards are measured as a reference before each batch to obtain the calibration curve needed to convert the retention time into the molecular weight, and a refractive index detector (RID) is applied to measure the concentration at the outlet of the column.

Although the MWD of a specific polymer is a discrete distribution, the GPC measurement returns a continuous signal which is composed of the residence time distribution of the polymer chains in the sample. If the calibration of the GPC device were perfect, one could potentially deconvolute the underlying discrete distribution, in which case the discrete Eq. (1) could be directly applied. However, the GPC retention time is based on hydrodynamic volume, which not only depends on the molecular weight but also on the chemical composition of the polymer. As a result, GPC measurements with RID can only provide the exact molecular weight distribution if the used calibration standard has the same chemical composition as the measured compound. Since the measured polymer and polystyrene are chemically different, we treat the MWD from GPC as a continuous distribution and apply Eq. (2) using the previously described piecewise constant discretization for the modeling.

The maximum molecular weight was determined based on visually checking the point where the chromatogram signal enters the noise floor, then this value was used to scale the molecular weights to (0,1). Assuming negligible mass loss due to condensation, the mass of each sample was scaled to 1000, as a large value was found to be beneficial in the subsequent estimations. The mass distributions were converted to number distributions using the scaled molecular weight grid. The distributions were resampled to a high-fidelity grid with 2^{16} grid points, $n = 2048$ size classes were defined, and the mean values of each class were assigned to their center points. The high number of size classes is necessary to accurately represent the dynamically changing low

molecular weight range while including the high molecular weight chains as well.

3.3 Model Application

The specific use case considered in the project is to provide decision support for plant personnel by predicting when the reaction should be quenched such that the deviation of the final weight average molecular weight from the specification is minimized. Currently, a heuristic method is used for this purpose in which chromatograms from the running batch are compared to recent chromatograms of finished batches with known final molecular weights, and the endpoint is determined based on deviations between the chromatograms. Although the heuristic method mostly performs well, it reportedly becomes difficult to use in case of varying target molecular weights within product grades, and impossible when the raw chromatograms are shifted due to exchange of the GPC column packing. Furthermore, it is resource-intensive to perform the heuristic calculation, and a fully automated, reliable prediction would free up valuable working time of the plant personnel.

The main source of difficulty in this use case is that the final molecular weight measurements cannot be used for fitting the model. Therefore, the model must be able to extrapolate in time from the four available in-process measurements, which are corrupted by a significant amount of measurement noise. The final molecular weight is measured after two downstream separation steps, which can introduce deviations in the molecular weight that cannot be accounted for at the time of reaction.

4 Model Training and Results

4.1 Estimation of the Training Set

Following the approach discussed in Sect. 2.2., the first step was to estimate the kernel function values to obtain an output dataset for training the data-based model that represents the kernel function. The discretization and kernel estimation method of [21] were implemented using the PyTorch automatic differentiation library [22] and the system of nonlinear ODEs was solved with a Dormand-Prince integrator, using the implementation described in [23]. To solve the nonlinear optimization problems, the L-BFGS optimizer with line search was applied.

The estimation strategy follows the principles described in [9] with some modifications. It was previously observed that the kernel estimates resulting from the measurements of this process are not time-independent, therefore a separate estimation for each interval in the dataset is necessary. An important property of the number density function in aggregation problems is that the function values are distributed across multiple orders of magnitude, decaying exponentially as the distributed property tends to infinity. To accurately estimate the kernel function values corresponding to higher molecular weights, scaling of the loss function across the different size classes is necessary:

$$J = \sum_{j=1}^{n} \left(\frac{N_j^{\text{meas}} - \hat{N}_j}{N_j^{\text{meas}} + \varepsilon} \right)^2 \tag{9}$$

where $j = 1,\ldots n$ is the size class index and ε is a small positive number to prevent division by 0. However, we encountered cost function values of $>10^7$ in the first iteration of the optimizer when this scaled loss function was applied in the interval-wise estimation. This presumably happens when N_i^{meas} is several orders of magnitude smaller than the model prediction $\hat{N}_i$. Therefore, we split the estimation into two phases. First, a time-independent kernel over all intervals is fitted with the following scaled loss:

$$J = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{N_{i,j}^{meas} - \hat{N}_{i,j}}{\frac{1}{m}\sum_{i=1}^{m} N_{i,j}^{meas} + \varepsilon} \right)^2 \tag{10}$$

where $i = 1,\ldots m$ is the observation index. In this loss function, the scaling factor is the average value in a size class taken across the available measurements of the batch. This approach keeps the scaling factor in the denominator in a less sensitive range. In the second phase, the previously obtained kernel estimate is fine-tuned to each interval separately with the unscaled squared error to avoid numerical instabilities and overfitting to measurement noise.

In the present use case, the number of size classes n was chosen to be 2048, and the number of available observations per batch m was 4. The estimation in the first phase was run for five iterations, where convergence could be observer after the first iteration in all cases. The second estimation phase was run for 10 iterations. Different models were fitted to the two product grades. The evolution of the values of the loss function for product grade A can be seen in Fig. 2, and a typical estimated kernel matrix is shown in Fig. 3.

4.2 Data-Based Kernel Model Training

In our preliminary report [9] on the same system, we observed a linear increase of $\mathbf{U}$ with respect to time. One possible explanation for this is that the mobility of the polymer chains in the solution changes with increasing viscosity. The viscosity of polymer solutions and melts is often correlated to the average molecular weight. Therefore, the data-based kernel (7) was parametrized with a linear model, with x_{process} chosen to be the number average molecular weight $\overline{M_n}(t)$. The resulting model structure is shown in Fig. 4.

$$\mathbf{U} = \mathbf{w}\overline{M_n}(t) + \mathbf{b} \tag{11}$$

$$\overline{M_n}(t) = \frac{\sum_{i=1}^{n} M_{w,i} N_i(t)}{\sum_{i=1}^{n} N_i(t)} \tag{12}$$

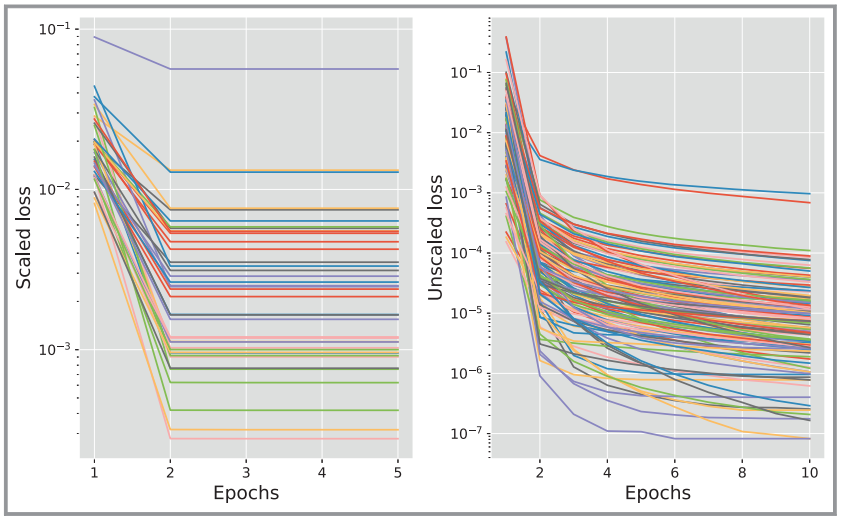

Figure 2. Evolution of the values of the loss function during the two estimation phases of product grade A. Similar trajectories were observed in the estimation of product grade B.

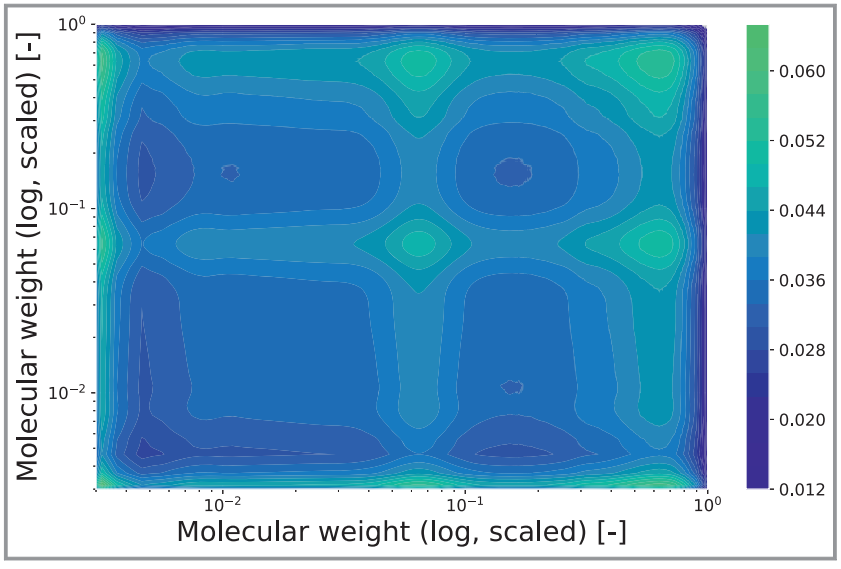

Figure 3. Visualization of a typical estimated kernel matrix.

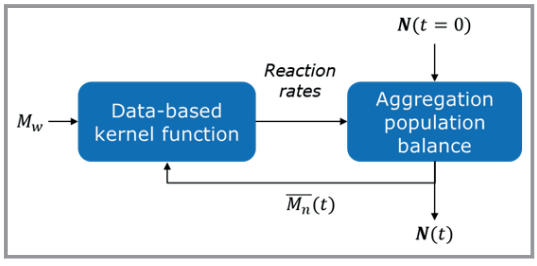

Figure 4. Schematic representation of the gray-box model structure.

where $\mathbf{w}, \mathbf{b} \in \mathbb{R}^{n^*1}$ and $i = 1,\ldots n$ is the size class index. Note that based on (5) this yields a quadratic dependence of the kernel function values $K_{i,j}$ on $\overline{M_n}$:

$$K_{ij} = U_i U_j = \left(w_i \overline{M_n} + b_i\right)\left(w_j \overline{M_n} + b_j\right) \quad (13)$$

For the training dataset of this model, the true number average molecular weights were calculated for each sample and linearly interpolated to the midpoint of each time interval. The linear least squares problems were solved with the *gelsd* solver from LAPACK.

4.3 Model Evaluation

Before applying the model to the end-point prediction, the prediction accuracy was evaluated on the in-process datasets using leave-one-out cross-validation. This can be performed very fast since only the linear least squares problem described in Sect. 4.2 needs to be solved for the different folds, and the parametrized model can be evaluated on the left-out batch in less than a second. Fig. 5 illustrates the prediction accuracy of the molecular weight distribution for two validation batches from the two different product grades. For better visibility, the predictions were resampled on a geometrically spaced grid. A qualitatively very good match can be observed between the predicted and measured distributions.

Fig. 6 and 7 show the relative errors of the prediction for the full molecular weight distribution and for the weight average molecular weight. Fig. 4 reveals a higher prediction inaccuracy at higher molecular weights. In the interval of (0.1, 1) of scaled molecular weight, the number distribution decays exponentially over 3–4 orders of magnitude, as discussed in Sect. 4.1. The main source of error in the distribution measurements is the uncertainty of the baseline determination of the chromatogram, which has a pronounced effect on the relative errors at low values of the number density. Since neither of the figures suggest an accumulation of the prediction error with time, it is assumed that the errors caused by the plant-model mismatch are small compared to those caused by the measurement inaccuracy.

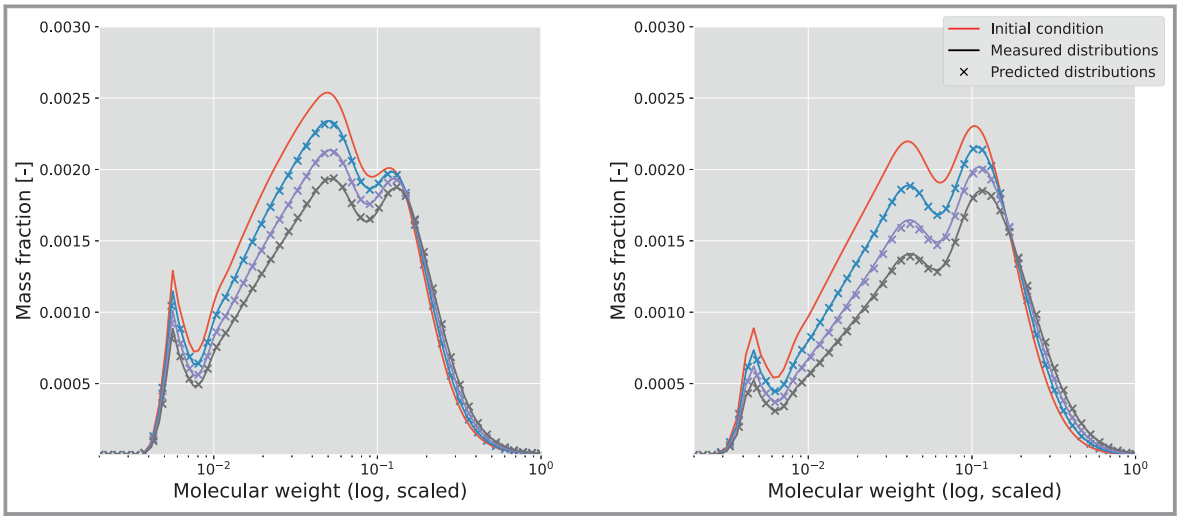

Figure 5. Predicted and measured distributions in two validation batches from the two product grades. The samples were taken with one hour time difference. The relative errors of prediction are quantified in Fig. 6.

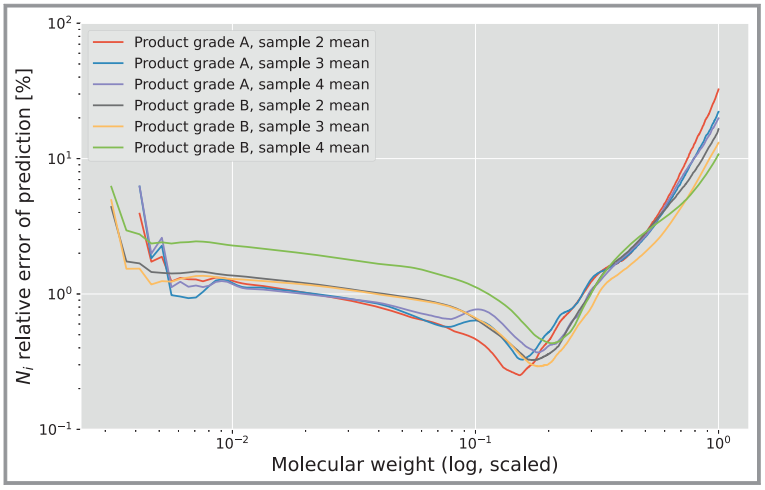

Figure 6. Mean relative errors of prediction of the molecular weight distribution, evaluated using leave-one-out cross-validation. The model was initialized with the first available sample for each batch, as illustrated in Fig. 5.

4.4 Prediction of the Endpoint of the Reaction

The objective in this application of the gray-box model is to estimate the time at which the reaction mixture is expected to reach the target weight average molecular weight. It is not possible to directly evaluate this point in time on the historical dataset since we do not have access to the reaction time that would have yielded the target molecular weight. Therefore, we evaluate the prediction performance of the final measured weight average molecular weight by integrating the model until the final reaction time. As discussed in Sect. 3.2, the final measurements are taken after two downstream separation steps, which is one of the sources of prediction error in this case.

The endpoint prediction is performed based on four measurements which provide inaccurate predictions when used individually, as suggested by Fig. 5. Although the assimilation of noisy measurements into a dynamic model could be approached using state estimation techniques, we here face a low number of available timesteps, complex noise characteristics and a high-dimensional, nonlinear model. Instead, a naive but robust approach was applied: each of the four measurements are used as initial conditions and the resulting predictions are averaged. This approach was also evaluated in a leave-one-out cross-validation loop.

The resulting cross-validation errors are shown in Fig. 8 for the batches of product grade A and B. The average absolute errors are 1.61 % and 1.27 %, respectively. We can also observe a higher prediction offset for product grade A. Since the weight average molecular weight is known to be more sensitive to the higher molecular weight species than the number average, we can attribute some of the random prediction error to the previously highlighted measurement inaccuracy at the higher molecular weights in Fig. 5. However, this form of the model only accounts for batch-to-batch variation in the molecular weight distribution of the initial polymer, and it was suspected that there are other raw material related factors influencing the rate of molecular weight increase.

4.5 Residual Modeling

Similar to the combined gray-box approach of [3], the prediction errors of the serial gray-box model were used to train a statistical model for error correction. In cooperation with the plant engineering team, four quality parameters of the raw polymer (excluding molecular weight related

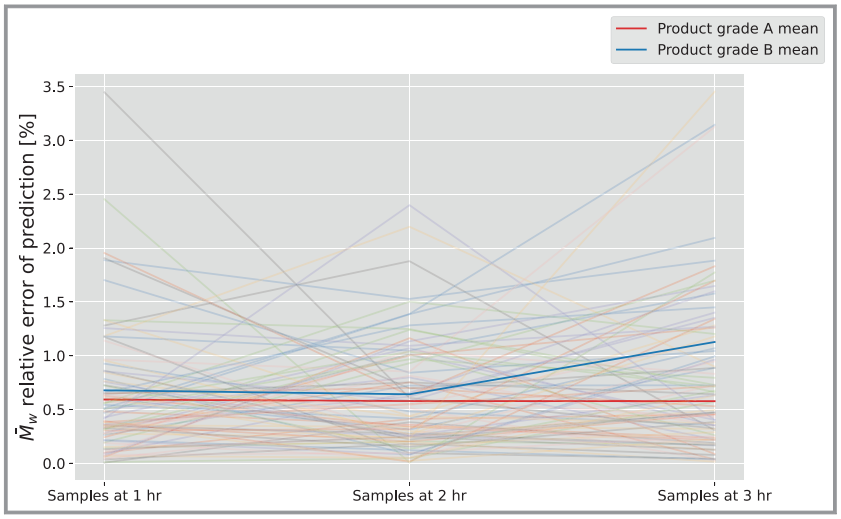

Figure 7. Relative errors of prediction of the weight average molecular weight for 99 historical batches at each sampling point (transparent lines) and their averages within the two product grades (solid lines), evaluated using leave-one-out cross-validation. The model was initialized with the first available sample for each batch.

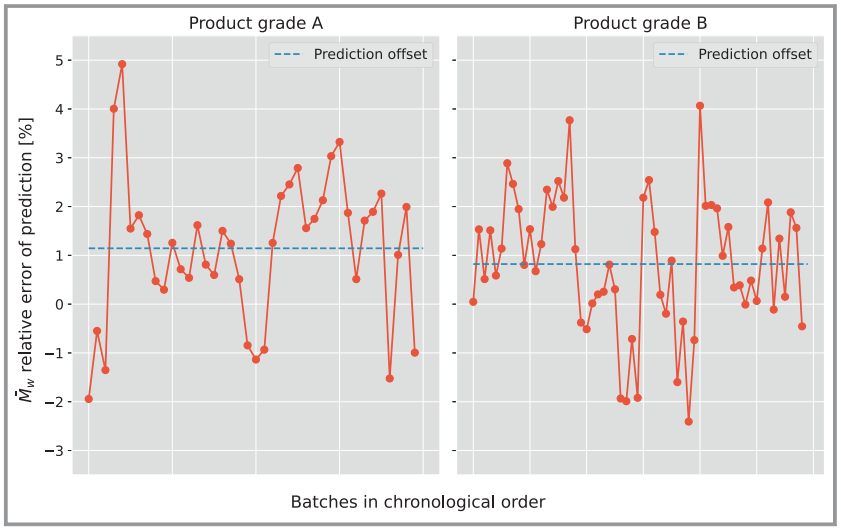

Figure 8. Relative cross-validation errors of predicting the final weight average molecular weight in the historical dataset.

parameters) were selected based on their potential influence on the rate of molecular weight increase. These properties were used as predictor variables in the regression of a linear model without interactions on the prediction error of the serial gray-box model. The selection of a linear model was to mitigate the high risk of overfitting due to a potentially low signal to noise ratio. The residual model regression was incorporated into the previously described leave-one-out cross-validation procedure. Knowing that some of the improvement will simply be the result of learning the prediction offset, the prediction error after offset removal was also quantified.

The results can be observed in Tab. 2 and Fig. 9. In both cases, the residual model results in a significant improvement of the mean relative prediction error, but it is also clear from Tab. 2 that product grade A mostly benefits from the offset removal. This is also indicated by the fact that the product grade A error chart in Fig. 9 is largely a shifted version of that of Fig. 8. On the other hand, predictions for product grade B are clearly improved by the error correction model.

4.6 Comparison with Previous Work

Most model-based methods in the literature that have been proposed for the monitoring and control of batch polymerization reactions fall into two main groups. In the absence of a rigorous dynamic process model, the first group of methods leverage online process measurements using principal component analysis, partial least squares or other types of statistical models [24]. These methods compress information contained in the measurement trajectories and facilitate the comparison of the currently running batch with past batches, or directly predict a final property of the running batch. A recent example for batch endpoint prediction can be found in [25]. If a dynamic process model is available, the second group of methods apply state estimation techniques to infer polymer quality properties and monomer conversion from the available online measurements. Due to the nonlinear dynamics of polymerization reactions, extended Kalman filtering [26, 27] or moving horizon estimation [28] may be applied.

The endpoint prediction method developed here uses a dynamic gray-box model to extrapolate from the few available offline measurements of the MWD and a statistical model to improve the endpoint prediction

Table 2. Mean relative cross-validation error of the weight average molecular weight with the baseline model, with offset removal and with the error correction model applied.

Model type	Mean relative error, product grade A [%]	Mean relative error, product grade B [%]
Base model	1.61	1.27
With offset removal	1.16	1.09
With residual model	0.96	0.72

Figure 9. Relative cross-validation errors of predicting the final weight average molecular weight in the historical dataset using the error correction model.

accuracy by considering the batch-to-batch variations in raw polymer quality. As pointed out in Sect. 4.2, the dynamic gray-box model could also be used in a state estimator, but this was not attempted because the frequency of the available molecular weight distribution measurements is quite low. The application of statistical batch process monitoring methods was also hindered by the lack of online measurements that relate to the reaction progress. Therefore, the described endpoint prediction approach is mainly useful in polymerization reactions where the reaction time is sufficiently long to evaluate the MWD of at least 2–3 in-process samples, and disturbances are primarily present in the raw material quality.

5 Conclusion

We have described a combined gray-box approach to model the temporal evolution of the molecular weight distribution in step-growth polymerization reactions. The approach uses the aggregation population balance and nonlinear optimization to extract the values of the discretized kernel function driving the binary reactions between the polymer chains. The obtained dataset is used to fit a simple data-based model that predicts the kernel function values from relevant process parameters. This decoupled training process does not require a priori knowledge about the kernel function. To improve the numerical stability of the training set estimation, it was split into two parts: a time-independent kernel estimation on the entire batch data using a scaled cost function, followed by the refinement of the initial estimate to each time interval without error scaling to avoid numerical instability and overfitting.

The described approach was applied to model a batch polymerization reaction at Merck KGaA using in-process gel permeation chromatography measurements of the molecular weight distribution. The target application of the model is to provide decision support for plant personnel by predicting the reaction endpoint corresponding to a target molecular weight from four in-process measurements. Two separate models were trained and evaluated using leave-one-out cross-validation on the datasets of two different product grades with 40 and 59 batches. The average cross-validation errors of predicting the final the weight average molecular weight in the historical dataset of the two products were 1.61 % and 1.27 %. To improve the predictions, a linear model was trained on the residuals to correct the predictions of the serial gray-box model. This combined gray-box model approach improved the cross-validation errors to 0.96 and 0.72 %. The model performance was found to be acceptable by the engineering team and it will be deployed on premise to replace the currently used time-consuming heuristic endpoint prediction method.

As the described approach relies completely on measurements of the molecular weight distribution to estimate the kinetics, this method certainly has its disadvantages in broader applications. Gel permeation chromatography is an imprecise and time-consuming measurement technique, and in-process sampling of a reaction is not always feasible for practical or safety-related reasons. To efficiently apply the aggregation population balance equation for designing and optimizing step-growth polymerization reactions, the dependency on such measurement data should be decreased. The next steps of this research will therefore involve a deeper look into the kinetics of ideal polymerizations for which physically founded kernel functions can be derived. The kernel estimation techniques described in this paper can then be applied to quantify the deviation from such ideal kinetics, typically caused by different reactivities of functional groups or cyclization reactions. This approach is expected to extend the validity domain of the resulting models and make the method more suitable for applications in process development.

Acknowledgment

The BMWK is acknowledged for funding the KEEN project initiative. (Support code: 01MK20014O). Open access funding enabled and organized by Projekt DEAL.

Symbols used

b	[–]	Linear model bias vector
K	[s^{-1}]	Kernel matrix
$\overline{M_n}$	[g mol^{-1}]	Number average molecular weight
$\overline{M_w}$	[g mol^{-1}]	Weight average molecular weight
N	[m^{-3}]	Discrete number density
n	[m^{-3}]	Continuous number density
S	[–]	Kernel coefficient matrix
U	[$s^{-0.5}$]	Kernel basis matrix
w	[–]	Linear model weight vector

Greek letters

α, β	[$s^{-0.5}$]	Kernel decomposition vectors
κ	[s^{-1}]	Number average molecular weight

Sub- and Superscripts

est	Estimated
meas	Measured

Abbreviations

GPC	Gel permeation chromatography
IUPAC	International Union of Pure and Applied Chemistry
MWD	Molecular weight distribution
NMR	Nuclear magnetic resonance
ODE	Ordinary differential equation
PBE	Population balance equation
RID	Refractive index detector

References

[1] J. Winz, S. Engell, *IFAC-Pap.* **2022**, *55 (7)*, 86–93. DOI: https://doi.org/10.1016/j.ifacol.2022.07.426

[2] J. Willard, X. Jia, S. Xu, M. Steinbach, V. Kumar, *ACM Comput. Surv.* **2023**, *55 (4)*, 1–37. DOI: https://doi.org/10.1145/3514228

[3] I. Ahmad, M. Kano, S. Hasebe, H. Kitada, N. Murata, *J. Process Control* **2014**, *24 (4)*, 375–382. DOI: https://doi.org/10.1016/j.jprocont.2014.01.018

[4] Z. Yang, T. Hagedorn, D. Eddy, S. Krishnamurty, I. Grosse, P. Denno, Y. Lu, P. Witherell, in *Proc. ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, New York **2017**. DOI: https://doi.org/10.1115/DETC2017-67794

[5] L. Scheffold, T. Finkler, U. Piechottka, *Comput. Chem. Eng.* **2021**, *146*, 107204. DOI: https://doi.org/10.1016/j.compchemeng.2020.107204

[6] A. E. F. Bouaswaig, K. Rahimi-Adli, M. Roth, A. Hosseini, H. Vale, S. Engell, J. Birk, *at - Autom.* **2020**, *68 (7)*, 582–598. DOI: https://doi.org/10.1515/auto-2020-0038

[7] X. Liu, K. Li, M. McAfee, B. K. Nguyen, G. M. McNally, *Polym. Eng. Sci.* **2012**, *52 (6)*, 1332–1341. DOI: https://doi.org/10.1002/pen.23080

[8] H. Wu, L. Cao, J. Wang, *J. Process Control.* **2012**, *22 (9)*, 1624–1636. DOI: https://doi.org/10.1016/j.jprocont.2012.06.014

[9] B. Bordas, K. Kurt, A. Bamberg, S. Engell, in *Proc. 2022 33rd Annu. SEMI Adv. Semicond. Manuf. Conf. ASMC*, IEEE, Piscataway, NJ **2022**. DOI: https://doi.org/10.1109/ASMC54647.2022.9792518

[10] I. Mita, R. F. T. Stepto, U. W. Suter, *Pure Appl. Chem.* **1994**, *66 (12)*, 2483–2486. DOI: https://doi.org/10.1351/pac199466122483

[11] P. J. Flory, *Principles of Polymer Chemistry*, 19th ed., Cornell Univ. Press, Ithaca, NY **2006**.

[12] D. Schmaljohann, J. G. Barratt, H. Komber, B. I. Voit, *Macromolecules* **2000**, *33 (17)*, 6284–6294. DOI: https://doi.org/10.1021/ma000065m

[13] D. Schmaljohann, B. Voit, *Macromol. Theory Simul.* **2003**, *12 (9)*, 679–689. DOI: https://doi.org/10.1002/mats.200350032

[14] R. M. Ziff, *J. Stat. Phys.* **1980**, *23 (2)*, 241–263. DOI: https://doi.org/10.1007/BF01012594

[15] J. Spouge, *Macromolecules* **1983**, *16 (1)*, 121–127. DOI: https://doi.org/10.1021/ma00235a024

[16] P. G. J. van Dongen, M. H. Ernst, *J. Stat. Phys.* **1984**, *37 (3–4)*, 301–324. DOI: https://doi.org/10.1007/BF01011836

[17] H. Galina, J. B. Lechowicz, in *Grafting/Characterization Techniques/Kinetic Modeling* (Eds: H. Galina, Y. Ikada, K. Kato, R. Kitamaru, J. Lechowicz, Y. Uyama, C. Wu), Springer, Heidelberg **1998**, 135–172.

[18] J. A. D. Wattis, *Phys. Nonlinear Phenom.* **2006**, *222 (1–2)*, 1–20. DOI: https://doi.org/10.1016/j.physd.2006.07.024

[19] J. Winz, S. Assawajaruwan, S. Engell, *Chem. Ing. Tech.* **2023**, *95*, *(7)*, in press. DOI: https://doi.org/10.1002/cite.202200237

[20] S. Le Borne, L. Shahmuradyan, K. Sundmacher, *Comput. Chem. Eng.* **2015**, *74*, 115–127. DOI: https://doi.org/10.1016/j.compchemeng.2014.12.011

[21] R. Ahrens, S. Le Borne, *Adv. Comput. Math.* **2021**, *47 (3)*, 39. DOI: https://doi.org/10.1007/s10444-021-09871-w

[22] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, arXiv.1912.01703, **2019**. DOI: https://doi.org/10.48550/arXiv.1912.01703

[23] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud, *Neural Ordinary Differential Equations*, arXiv.1806.07366, **2018**. DOI: https://doi.org/10.48550/arXiv.1806.07366

[24] P. Nomikos, J. F. MacGregor, *Chemom. Intell. Lab. Syst.* **1995**, *30 (1)*, 97–108. DOI: https://doi.org/10.1016/0169-7439(95)00043-7

[25] F. Souza, T. Offermans, J. Jansen, in *Proc. 2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, IEEE, Piscataway, NJ **2022**. DOI: https://doi.org/10.1109/MetroXRAINE54828.2022.9967672

[26] P. Appelhaus, S. Engell, *IFAC Proc. Vol.* **1996**, *29 (1)*, 5971–5976. DOI: https://doi.org/10.1016/S1474-6670(17)58637-1

[27] M. F. Ellis, T. W. Taylor, V. Gonzalez, K. F. Jensen, *AIChE J.* **1988**, *34 (8)*, 1341–1353. DOI: https://doi.org/10.1002/aic.690340812

[28] J. D. Hedengren, K. V. Allsford, J. Ramlal, in *Proc. 2007 American Control Conference*, IEEE, Piscataway, NJ **2007**. DOI: https://doi.org/10.1109/ACC.2007.4282820

