

## Article

# Holistic Process Models: A Bayesian Predictive Ensemble Method for Single and Coupled Unit Operation Models

Liliana Montano Herrera <sup>1</sup>, Tobias Eilert <sup>2</sup>, I-Ting Ho <sup>3</sup>, Milena Matysik <sup>4</sup>, Michael Laussegger <sup>5</sup>, Ralph Guderlei <sup>4</sup>, Bernhard Schrantz <sup>5</sup>, Alexander Jung <sup>1</sup>, Erich Bluhmki <sup>6,7</sup> and Jens Smiatek <sup>1,8,\*</sup>

<sup>1</sup> Digitalization Development Biologicals CMC, Boehringer Ingelheim Pharma GmbH & Co. KG, D-88397 Biberach, Germany; liliana.montano\_herrera@boehringer-ingelheim.com (L.M.H.); alexander\_1.jung@boehringer-ingelheim.com (A.J.)

<sup>2</sup> Bio Quality Mammalian, Boehringer Ingelheim Pharma GmbH & Co. KG, D-88397 Biberach, Germany; tobias.eilert@boehringer-ingelheim.com

<sup>3</sup> Global Computational Biology and Digital Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, D-88397 Biberach, Germany; i-ting\_2.ho@boehringer-ingelheim.com

<sup>4</sup> IT RDM Services, Boehringer Ingelheim GmbH & Co. KG, D-88397 Biberach, Germany; milena.matysik@boehringer-ingelheim.com (M.M.); ralph\_joachim.guderlei@boehringer-ingelheim.com (R.G.)

<sup>5</sup> IT RDM Services, Boehringer Ingelheim RCV GmbH & Co. KG, A-1121 Vienna, Austria; michael.laussegger.ext@boehringer-ingelheim.com (M.L.); bernhard.schrantz.ext@boehringer-ingelheim.com (B.S.)

<sup>6</sup> Analytical Development Biologicals, Boehringer Ingelheim Pharma GmbH & Co. KG, D-88397 Biberach, Germany; erich.bluhmki@boehringer-ingelheim.com

<sup>7</sup> Biberach University of Applied Sciences, D-88400 Biberach, Germany

<sup>8</sup> Institute for Computational Physics, University of Stuttgart, D-70569 Stuttgart, Germany

\* Correspondence: jens.smiatek@boehringer-ingelheim.com



**Citation:** Montano Herrera, L.; Eilert, T.; Ho, I.-T.; Matysik, M.; Laussegger, M.; Guderlei, R.; Schrantz, B.; Jung, A.; Bluhmki, E.; Smiatek, J. Holistic Process Models: A Bayesian Predictive Ensemble Method for Single and Coupled Unit Operation Models. *Processes* **2022**, *10*, 662. <https://doi.org/10.3390/pr10040662>

Academic Editor: Ján Pitel'

Received: 1 March 2022

Accepted: 23 March 2022

Published: 29 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The coupling of individual models in terms of end-to-end calculations for unit operations in manufacturing processes is a challenging task. We present a probability distribution-based approach for the combined outcomes of parametric and non-parametric models. With this so-called Bayesian predictive ensemble, the statistical moments such as mean value and standard deviation can be accurately computed without any further approximation. It is shown that the ensemble of different model predictions leads to an uninformed prior distribution, which can be transformed into a predictive posterior distribution using Bayesian inference and numerical Markov Chain Monte Carlo calculations. We demonstrate the advantages of our method using several numerical examples. Our approach is not restricted to certain unit operations, and can also be used for the more robust interpretation and assessment of model predictions in general.

**Keywords:** unit operation models; Bayesian inference; machine learning; statistical and mechanistic models; holistic process models; coupled end-to-end process models

## 1. Introduction

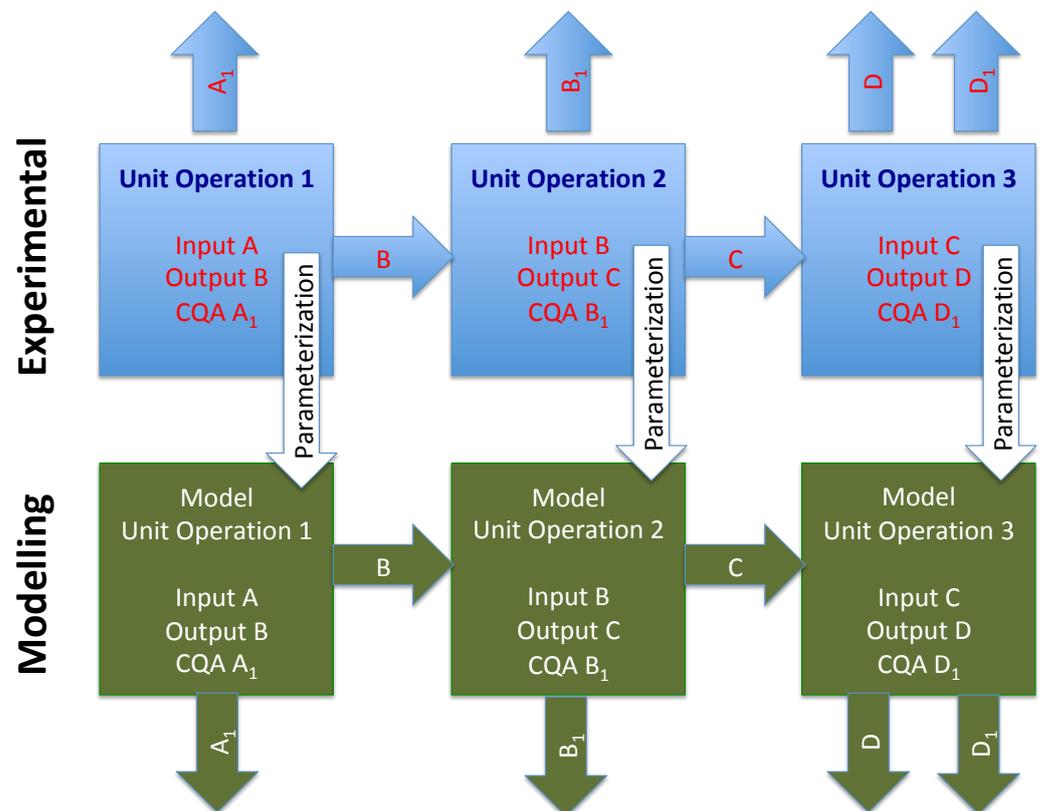
Bioprocess development and manufacturing for therapeutic drugs is often a challenging task. The parameter settings of individual unit operations at upstream, downstream and formulation stages need to be adjusted so that the final drug product is produced at high concentration with a robust and satisfying quality that meets regulatory and patients requirements [1–3]. Due to the coupling of different process steps including cultivation, purification, polishing, filtering and formulation among others, it is known that the dimension of process parameters which form the design space grows significantly with an increasing number of unit operations [1].

Notably, the identification of global optimal process parameter settings for all combined unit operations is not straightforward and provides some challenges for experimental work. Most often, normal operating ranges (NORs) and proven-acceptable ranges (PARs)

are defined for robust and local parameter settings of single unit operations. Hence, bioprocess optimization is mainly performed for all process steps individually while the combined global optimization of parameter settings is often ignored due to the enormous amount of experimental work.

In recent years, several computational modeling approaches were introduced to guide experimental settings and to reduce laboratory work [1,3,4]. Standard methods are mechanistic and hybrid models [5–14] in combination with statistical, data-driven or empirical approaches for upstream and downstream process steps [15–18]. Parametric models are usually calibrated in terms of certain experimental conditions and are often used to study various process parameter settings. Noteworthy, most parametric modeling approaches are computationally cheap and the calculations can be performed on short time scales, which allows to accelerate development times significantly. In addition, the last years have also seen a significant increase of non-parametric models in terms of machine learning [1,4]. Despite the underlying parametric or non-parametric characteristics, an important concept for model development and application is validation, which is usually conducted in comparison to experimental outcomes [19,20].

The straightforward application of modeling approaches further emerged recent interest in the combined evaluation of process outcomes (Figure 1).



**Figure 1.** Top row in blue color: Experimental set-up of three unit operations. The output values of the individual unit operations B and C are used as input values for the next unit operations. In addition, there exist certain CQA values such as  $A_1$ ,  $B_1$ ,  $D_1$ , and the final process outcome D. The corresponding experimental data is used for the parameterization or the training of models for the individual unit operations (bottom row in green color). Three models represent the process behavior at the unit operation stage and can be used for predictions. The results of the model predictions are used as input values for the next models in combination with further CQAs. The corresponding transfer functions with the model outputs are represented as green horizontal arrows.

Hence, over the last years, several combined end-to-end process models have been developed, which are often called integrated process models, end-to-end bioprocess mod-

els, holistic process models, bioprocess replicas, flowsheet models or bioprocess digital twins [1,4,21–26]. Despite the fact that certain differences in the underlying methodology and concepts are evident, all approaches offer the study of process parameter settings across all unit operations in terms of global optimization. However, the coupling of models between the individual unit operations is still a challenging task (Figure 1). A consistent and combined multidimensional approach to estimate parameter ranges for predictions has not been conclusively proposed. Recent approaches suggested a Monte Carlo sampling scheme [22,25], whereas other publications favored a progression of discrete values for further processing steps also in terms of model-predictive control strategies [27–29]. In more detail, it was shown that the detailed evaluation of process parameter ranges crucially affects the outcomes [22], such that the detailed consideration of coupled effects for all models is of utmost importance.

Specifically the determination of confidence intervals is a challenging task for parametric and non-parametric models. Notably, standard techniques like bootstrapping [30–32] or methods that consider Bayesian inference calculations [33,34] often rely on a reasonable amount of model parameterization or training data. Recent articles already showed that the transformation of outcome ranges along certain unit operations is not trivial and crucially affects the specifications [22,25]. Further articles also focused on model-free identification strategies for optimal parameter settings [35,36]. In terms of coupled unit operation models, it is clear that the resulting design space is high dimensional and optimal parameter settings can mainly be identified with stochastic approaches. In consequence, it can be assumed that probabilistic and stochastic approaches for the calculation and optimization of process outcomes will become even more important in the near future.

In this article, we present a probabilistic approach to predict process outcomes using a Bayesian predictive ensemble approach. We show that an ensemble of different model predictions leads to an uninformed prior distribution, which can be transformed into a predictive posterior distribution using Bayesian inference in combination with Markov Chain Monte Carlo (MCMC) calculations. Such approaches are particularly applicable for the modeling of coupled process variables within the framework of end-to-end or holistic process models. Our approach is not restricted to bioprocess models, and can be independently used for all machine learning calculations as well as parametric models.

For the first time, we present a rigorous mathematical framework to describe unit operation models in terms of time evolution operators. The corresponding approach allows us to define meaningful and multivariate transformation functions for the connection of unit operations in accordance with holistic process models. The corresponding introduction of predictive distributions facilitates Bayesian inference calculations for a meaningful evaluation of confidence intervals. We show that point-like discrete connections or simple Gaussian distributions lead to a strong increase of uncertainty after an increasing number of connected unit operations. In contrast, our suggested approach of using conditional probabilities in terms of posterior distribution functions results in a controllable increase and further takes the process memory into full consideration. As a new concept, we introduce the idea of using a collection of different model predictions. This so-called Bayesian predictive ensemble allows us to overcome drawbacks in terms of limited experimental data for model parametrization.

The article is organized as follows. In Section 2, we introduce a mathematical description for time-dependent models and discuss their influence on the underlying probability distribution functions. The corresponding framework is then used to outline the main properties of the Bayesian predictive ensemble in combination with transformation functions in terms of posterior and prior distributions. In Section 3, we present a simplified numerical example to highlight the implications for the sake of clarity. We conclude and summarize in Section 4.

## 2. Theoretical Background

Biotechnological or chemical processes are often dominated by time-dependent behavior. Hence, the corresponding process outcomes show a temporal evolution which hinders the straightforward calculation of statistical moments. In the first subsection, we propose a consistent scheme to compute statistical properties at certain time points for processes with temporal evolution. In more detail, we present a time discretization scheme that can be interpreted as the initial and the end time points of individual process steps. The connections between these time points for the individual probability distribution functions are described in terms of model propagation steps. Explicit evaluation of these distribution functions is performed in terms of Bayesian inference. We show that the outcomes of different models can be used as uninformed prior distributions. The application of numerical Markov Chain Monte Carlo (MCMC) calculations allows us to compute the corresponding posterior distribution for the considered process outcomes. The range of this distribution reflects the impact of model or experimental uncertainty regarding the correct prediction. Moreover, the respective posterior distribution can further be used as initial input for the subsequent model in order to define a new prior distribution for the next unit operation. This iterative scheme based on conditional probabilities represents a new approach for the rational combination of unit operation models.

### 2.1. Non-Stationary Processes

Any non-stationary process is characterized by its explicit time dependency. Thus, the evaluation of mean values  $\langle A \rangle$  and higher moments like the variance  $\sigma^2(A) = \langle (A - \langle A \rangle)^2 \rangle$  for arbitrarily chosen variables or process outcomes  $A(t)$  with temporal evolution is challenging. As a further complication, the ergodic hypothesis [37]

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt A(\Gamma, t) = \int_{\Omega} d\rho(\Gamma) A(\Gamma), \quad (1)$$

is not applicable for time-dependent processes with temporal evolution. The left hand side of Equation (1) denotes a time average whereas the right hand side highlights measurements over different statistical replicas of the system

$$\langle A(\Gamma) \rangle = \int_{\Omega} d\rho(\Gamma) A(\Gamma), \quad (2)$$

where the corresponding parameters are an arbitrarily chosen observable  $A(\Gamma, t) = A(\Gamma(t))|_{\Gamma(0)=\Gamma}$  and the normalized probability density  $\int_{\Omega} d\rho(\Gamma) = 1$ , where  $T$  denotes the measured time interval, and  $\Omega$  the parameter space including all relevant parameters  $\Omega = \{\Gamma\}$  in terms of  $\Gamma = (\mathbf{X})$ , where  $\mathbf{X}$  denotes all variables  $\mathbf{X} = (X_0, X_1, \dots, X_P)^T$  in accordance with vector notation. Notably, the ergodic hypothesis holds for most systems as long as the decorrelation times  $\tau$  from autocorrelation functions

$$\langle A(\Gamma, t=0)A(\Gamma, t) \rangle \sim \exp(-t/\tau)^{\beta} \quad (3)$$

are short in accordance with the Markovian properties including the stretched exponential factor  $\beta \in \{\mathbb{R}^+, 0\}$ .

Despite all challenges, it is possible to average over the values at certain fixed time points  $t_M$ , which results in

$$\int_0^T dt A(\Gamma, t) \delta(t - t_M) = \int_{\Omega} d\rho(\Gamma, t_M) A(\Gamma, t_M) \quad (4)$$

and hence

$$\langle A(\Gamma, t) \rangle_{t_M} = \int_{\Omega} d\rho(\Gamma, t_M) A(\Gamma, t_M) \quad (5)$$

for time dependent values  $\rho(\Gamma, t)$ ,  $A(\Gamma, t)$ , where  $\delta(t - t_M)$  denotes the delta function, which is  $\delta(t - t_M) = 1$  for  $t = t_M$  and 0 otherwise. Hence, if we assume that  $t_M$  is the time point of measurement, one can interpret  $\langle A(\Gamma, t) \rangle_{t_M}$  as the corresponding statistical outcome for process values at identical time points for different replicas of the system. In consequence, the implications from Equation (5) provide the usage of standard statistical approaches for non-stationary processes like cultivation steps [15]. Thus, we introduced a framework to project the outcomes of a set of non-stationary processes on fixed time points in order to define distributions with mean values and standard deviations. With regard to such a concept, we can consider time-dependent and stationary processes as equivalent in terms of identical statistical considerations.

## 2.2. Time Evolution Propagators

In addition to the quasi-stationary approximation for the calculation of mean values that we introduced in the previous subsection, it is also possible to define a similar interpretation for the probability distribution functions in the presence of explicit time evolution. For a constant and time-independent probability density, it follows

$$\frac{d}{dt}\rho(\Gamma, t) = 0. \quad (6)$$

which can be written as a continuity equation in accordance with [38]

$$\frac{\partial \rho(\Gamma, t)}{\partial t} = -\nabla_{\Gamma}(\rho(\Gamma)\dot{\Gamma}) \quad (7)$$

where the dot over the symbol denotes the first derivative in time according to  $\dot{\Gamma} = (\dot{X}_i, \dot{X}_{i+1}, \dots, \dot{X}_N)$  and  $\nabla_{\Gamma}\dot{\Gamma} = \sum_i^N \partial \dot{X}_i / \partial X_i$ . Notably, Equation (7) can also be written as

$$i \frac{\partial \rho(\Gamma, t)}{\partial t} = \mathcal{L}(\Gamma)\rho(\Gamma, t) \quad (8)$$

with the Liouville-like operator [38]

$$\mathcal{L}(\Gamma) = -i \sum_i^N \frac{\partial \dot{X}_i}{\partial X_i} \quad (9)$$

providing the Eigenvalue relation

$$\mathcal{L}(\Gamma)\rho(\Gamma, t) = \lambda_j \rho(\Gamma, t) \quad (10)$$

with the eigenvalues  $\lambda_j$ . Here, we do not explicitly intend to introduce Hamiltonian and thus deterministic energy conservation dynamics [38], but mainly rely on the idea of conserved time evolution through operator formalisms. Thus, the Liouville-like operator does not depend on the momenta and spatial derivatives of the microscopic variables for our purposes. The complex conjugate of Equation (9) reads  $\mathcal{L}^* = -\mathcal{L}$ , such that Equation (10) yields

$$\mathcal{L}^*(\Gamma)\rho(\Gamma, t) = -\lambda_j \rho(\Gamma, t) \quad (11)$$

which is equivalent to a time inversion  $t \rightarrow -t$ . The formal solution for Equation (8) reads

$$\rho(\Gamma, t) = \rho_0(\Gamma)e^{i\mathcal{L}(\Gamma)t}. \quad (12)$$

such that the exponential factor can be further interpreted as a time propagator

$$\mathcal{P}(\Gamma, t) = e^{i\mathcal{L}(\Gamma)t} \quad (13)$$

for the probability distribution with invertible and thus reversible time evolution. In summary, Equation (13) can be considered as an universal time propagator that is used to

evaluate model calculations in a mathematical consistent framework. Herewith, we are able to map process parameter values to the corresponding outcomes. The concept allows us to consider models for non-stationary and stationary processes by means of an identical mathematical description.

### 2.3. Coupled Non-Stationary Processes: Model Propagator Dynamics

In terms of our previous discussions, it was shown that the propagator projects the temporal evolution of the process outcome in terms of statistical quantities onto certain fixed time points. Thus, it becomes immediately clear from Equation (13), that the progression of probability distributions in terms of certain time points  $t_M$  for a set of parameters  $\Gamma_M$  can be written as

$$\rho_{t_M}(A(\Gamma)) = \mathcal{P}(\Gamma, t_M)\rho_{t_0}(A(\Gamma)) \quad (14)$$

with the probability distribution  $\rho_{t_j}(A(\Gamma))$  for observable  $A(\Gamma)$  at time  $t_j$  and  $P^*(\Gamma, t_M) = P(\Gamma, -t_M)$  in accordance with Equation (11). It has to be noted that Equation (14) is valid for all times as long as the process parameter settings  $\Gamma$  remain unchanged.

For the following discussion, we replace the exact time evolution operator  $\mathcal{P}$  with a new model operator  $\mathcal{M}$  including identical time reversal symmetry conditions. In contrast to  $\mathcal{P}(\Gamma)$ , the set of parameters for  $\mathcal{M}(\kappa)$  is incomplete in terms of  $\kappa \subseteq \Gamma$  and  $\kappa \neq \Gamma$ . This can be mainly understood with regard to our incomplete knowledge of underlying correlations and effects for most processes as represented by model artifacts [19]. Hence, any model reflects a certain kind of uncertainty such that the model outcomes after the training or parameterization stage need to be carefully evaluated against experimental data, which can be considered as the ground truth distribution [19]. With regard to a change of parameters from  $\Gamma_0 \rightarrow \kappa_1$  at  $t_0 \rightarrow t_1$  according to

$$\rho_{t_1}(A(\kappa_1)) = \mathcal{M}(\kappa_1, t_1)\rho_{t_0}(A(\Gamma_0)), \quad (15)$$

it follows that  $\rho_{t_1}(A(\kappa_1))$  also differs from the true distribution  $\rho_{t_1}(A(\Gamma_1))$ . In order to correct model parameter settings with the corresponding experimental distributions, one can introduce the Bayes theorem [39]

$$\rho(z_i|y_i) = \frac{\rho(y_i|z_i)}{\rho(y_i)}\rho(z_i) \quad (16)$$

where  $\rho(\cdot|\cdot)$  denotes the conditional probability as defined by  $\rho(y_i|z_i) = \rho(y_i \cap z_i)/\rho(z_i)$  where  $\rho(y_i \cap z_i)$  denotes the joint probability for outcomes  $y_i$  and  $z_i$ . Statistical independence is only evident for  $\rho(y_i|z_i) = \rho(y_i)$  and  $\rho(z_i|y_i) = \rho(z_i)$ . The Bayes theorem thus strongly relies on the prior uninformed probability distribution  $\rho(z_i)$  in combination with the likelihood  $\rho(y_i|z_i)$  for the calculation of the posterior distribution  $\rho(z_i|y_i)$ . Applying the Bayes theorem on Equation (15), one can show that the resulting posterior distribution can be written as

$$\rho_{t_1}(A(\kappa_1)|A(\Gamma_1)) = \frac{\rho_{t_1}(A(\Gamma_1)|A(\kappa_1))}{\rho_{t_1}(A(\Gamma_1))}\rho_{t_1}(A(\kappa_1)) \quad (17)$$

which now explicitly incorporates the true experimental distribution of  $A$  in terms of the full set of parameters  $\Gamma_1$ . Furthermore, the corresponding predictive posterior distribution can be computed by

$$\rho(\hat{A}(\kappa_1)|A(\kappa_1)) = \int \rho_{t_1}(\hat{A}(\kappa_1)|A(\kappa_1), A(\Gamma_1))\rho_{t_1}(A(\kappa_1)|A(\Gamma_1))dA(\kappa_1) \quad (18)$$

where  $\hat{A}(\Gamma_1)$  denotes new data points generated by the model at  $t_1$ . In addition, the posterior distribution  $\rho_{t_1}(A(\kappa_1)|A(\Gamma_1))$  can be combined with a new model operator including new parameters  $\kappa_2$  in order to yield

$$\rho_{t_2}(A(\kappa_2)) = \mathcal{M}(\kappa_2, t_2)\rho_{t_1}(A(\kappa_1)|A(\Gamma_1)) \quad (19)$$

which can be interpreted as the prior distribution for the next step

$$\rho_{t_2}(A(\kappa_2)|A(\Gamma_2)) = \frac{\rho_{t_2}(A(\Gamma_2)|A(\kappa_2))}{\rho_{t_2}(A(\Gamma_2))} \cdot \rho_{t_2}(A(\kappa_2)) \quad (20)$$

In consequence, if we identify the different time points  $t_0, t_1, \dots, t_N$  as initial starting points of  $N$  individual unit operations as described by different model propagators  $\mathcal{M}_1(\kappa_1), \dots, \mathcal{M}_N(\kappa_N)$ , one can interpret the corresponding evolution equation as a Bayesian hierarchical model [1] according to

$$\begin{aligned} \rho_{t_j}(A(\kappa_j)|A(\Gamma_j)) &= \\ &= \frac{\rho_{t_j}(A(\Gamma_j)|A(\kappa_j))}{\rho_{t_j}(A(\Gamma_j))} \mathcal{M}(\kappa_j, t_j)\rho_{t_{j-1}}(A(\kappa_{j-1})|A(\kappa_{j-2}), \dots, A(\kappa_1), A(\Gamma_1)) \end{aligned} \quad (21)$$

for  $j = 4, \dots, N$ . With regard to the previous relations, one is thus able to project and to combine distribution functions on certain time points in terms of Bayesian inference for an improved prediction of model outcomes. Moreover, it becomes clear that the explicit history of the process in terms of previous parameter settings and process outcomes is well-defined and taken into consideration through conditional probabilities. In more detail, Equation (21) represents the center of our proposed approach. Herewith, we can combine individual unit operation models for stationary and non-stationary processes at fixed time points in terms of transformation functions as represented by prior and posterior distribution functions. The corresponding single model predictions can be used to define a predictive distribution with well-defined confidence intervals, which can be used as prior distribution for the next unit operation model. The corresponding mathematical framework is applicable without any further restriction. In the next section we will show how this approach can be used to refine predictions from different models in order to define suitable transformation functions between coupled unit operations.

#### 2.4. Bayesian Predictive Ensemble: Combination of Different Models

In the previous subsection, we identified  $\Gamma$  and the corresponding probability distribution function  $\rho(A(\Gamma))$  as the experimental ground truth without further restrictions. Such an approach is beneficial if a sufficient amount of experimental data for identical process parameter settings are available. Here, we show how the previous relations can be used to assess the uncertainty of model predictions in terms of Bayesian predictive ensembles with regard to coupled model descriptions. It has often been discussed that the accurate calculation of ranges for process outcomes in terms of parametric and non-parametric models is a challenging task. The condition of time inversion for the model propagator introduces a deterministic behavior such that each parameter setting for a fixed model yields a unique model prediction value. Due to these reasons, it is useful to combine different models with reasonable accuracy in order to assess the uncertainty of correct model predictions.

In principle, any modeling approach can be used for the prediction of singular values for the variable  $A$  from the probability distribution function  $\rho(A(\Gamma)) = A(\Gamma) \int_{-\infty}^{+\infty} \delta(x - \Gamma) dx$  in accordance with Equation (15), such that

$$\bar{A}_j(\kappa_j) = \mathcal{M}_j(\kappa_j)A(\Gamma_0) \quad (22)$$

where  $\mathcal{M}_j$  denotes  $m$  individual models with the corresponding model parameters  $\kappa_j$  with  $j \in [0, m]$ . Here, we ignore the explicit time-dependence for the sake of clarity. The consideration of all model predictions  $\bar{A}_j(\kappa_j)$  in terms of the distribution  $\rho(\bar{A}_0(\kappa_0), \dots, \bar{A}_m(\kappa_m))$  can be used to evaluate the mean and standard deviation of model predictions. The corresponding values can be used to define an uninformed prior distribution function  $\rho(\bar{A}(\bar{\kappa}))$  in accordance with

$$\rho(\bar{A}(\bar{\kappa})|\bar{A}_0(\kappa_0), \dots, \bar{A}_m(\kappa_m)) = \frac{\rho(\bar{A}_0(\kappa_0), \dots, \bar{A}_m(\kappa_m)|\rho(\bar{A}(\bar{\kappa}))}{\rho(\bar{A}_0(\kappa_0), \dots, \bar{A}_m(\kappa_m))} \rho(\bar{A}(\bar{\kappa})) \quad (23)$$

which provides the informed posterior distribution  $\rho(\bar{A}(\bar{\kappa})|\rho(\bar{A}_0(\kappa_0), \dots, \bar{A}_m(\kappa_m)|\rho(\bar{A}(\bar{\kappa})))$  as needed for the definition of process parameter ranges for coupled models as well (Equation (21)). The corresponding hierarchy of coupled model predictions and Bayesian inference calculations (Bayesian predictive ensemble) can be summarized as follows:

1. Bayesian predictive ensemble: Application of different models with varying accuracy to estimate the prior distribution for a certain unit operation (mean value and standard deviation) in accordance with Equation (22) based on Equations (5) and (13);
2. Bayesian inference to estimate the posterior probability distribution (MCMC calculations) from the prior distribution of step 1 in accordance with Equation (23) based on Equation (22);
3. Representatives from the posterior distribution: Certain choices from the posterior probability distribution as values for models of the next unit operation in accordance with Equation (22);
4. Repeated application of step 1: Calculation of the prior distribution for the next unit operation with values from step 3.

The corresponding steps 1 to 4 represent the basic implementation of the algorithm. As a prerequisite, one parameterizes a set of models for predictions of a certain unit operation. An alternative approach is to use one model that is trained on different sets of training data in terms of the bootstrapping approach. The corresponding predictions of the models for a certain parameter setting are performed in step 1. The set of predictions can be interpreted as a prior distribution with a certain mean value and standard deviation. The prior distribution is then transformed via MCMC calculations into a posterior distribution in accordance with step 2. The resulting posterior distribution as represented by conditional probabilities is interpreted as transformation function to be used for the next unit operation in terms of the underlying input value distribution. Based on the posterior distribution, certain values are randomly drawn, which are then used as new parameter input values in combination with further process parameter settings as required for the next set of models for the subsequent unit operation (step 3). Finally, steps 1, 2, and 3 are repeated for the considered subsequent unit operations.

It has to be noted that the Bayesian predictive ensemble relies on different parametric or non-parametric models without further restriction in combination with directed Bayesian networks [40] in terms of conditional probability distribution functions. The corresponding variations of the predictions from the different models allow us to draw some conclusions on the uncertainty of the model predictions based on inaccurate parameterization or missing information in the data. Moreover, Bayesian approaches are well-suited to be used for low amounts of predictions [39]. Hence, only a limited number of available models can be used for the Bayesian predictive ensemble. Furthermore, it has to be noted that also Bayesian linear regression approaches [41] can be used as prior and posterior distribution functions. If such models are applicable in terms of reliable model predictions, it becomes clear that the time-consuming step 2 in the previous algorithm is already included in the Bayesian linear regression approach. Finally, it has to be noted that the calculation of Equation (23) is a straightforward task with regard to recent numerical MCMC sampling schemes [42]. Noteworthy, the previous conclusions hold as long as the number of considered variables is low. Although also multidimensional Bayesian predictive ensembles can be defined, it has

to be noted that certain convergence issues of MCMC steps for multivariate distributions were often observed.

### 2.5. Combination of Unit Operation Models: Error Propagation

As already discussed, the Bayesian predictive ensemble provides an unbiased estimate for the uncertainty of different model predictions. Despite certain deviations from the ground truth distribution, one can assume that all models are parameterized or trained to provide the best predictions. In general, the difference between two distributions can be studied in terms of the Kullback–Leibler divergence [43], which is defined as

$$D(\rho||q) = \int_{-\infty}^{\infty} \rho(A_i) \ln\left(\frac{\rho(A_i)}{q(A_i)}\right) dA_i \quad (24)$$

where  $\rho(A)$  denotes the model-predicted and  $q(A)$  the true distribution, respectively. In more detail, the Kullback–Leibler divergence is closely related to information entropy, such that large values of  $D(\rho||q)$  reveal a missing knowledge about the true distribution and thus crucial deviations. For point-like predictions of a model  $\delta(x)$  with vanishing variance, it can be seen that  $D(\rho||q) \rightarrow \infty$  such that the largest deviation from the ground truth distribution with finite variance can be assumed. This becomes even more important after combining a set of models in terms of coupled process model predictions where the output parameter  $\delta(x_1)$  is used as an input parameter for the next unit operation and so forth. Due to the logarithmic properties, the Kullback–Leibler divergence is additive for certain distributions  $D(\rho_1, \rho_2, \dots, \rho_N||q_1, q_2, \dots, q_N) = \sum_{i=1}^N D(\rho_i||q_i)$ , which yields in combination with Equation (24) the following expression for point-like predictions

$$D(\delta_1, \delta_2, \dots, \delta_N||q_1, q_2, \dots, q_N) = \prod_{i=1}^N \int_{-\infty}^{\infty} \delta_i(A_i) \ln\left(\frac{\delta_i(A_i)}{q_i(A_i)}\right) dA_i \quad (25)$$

such that  $D(\rho_1, \rho_2, \dots, \rho_N||q_1, q_2, \dots, q_N) \rightarrow \infty$ . Due to these reasons, it can be concluded that point-like predictions from single as well as coupled unit operation models provide the highest inaccuracy in comparison to the unknown true distributions.

Corresponding conclusions can also be drawn for the consideration of fixed model predicted mean values  $\langle A_i \rangle$  in combination with pre-defined standard deviations  $\sigma(A_i)$  in terms of Gaussian noise. Hence, inserting the corresponding distributions  $\rho_i(A_i) = \mathcal{N}_i(\langle A_i \rangle, \sigma(A_i))$  into Equation (24) yields

$$\begin{aligned} D(\mathcal{N}_1, \mathcal{N}_2, \dots, \mathcal{N}_N||q_1, q_2, \dots, q_N) &= \\ &= \prod_i \int_{-\infty}^{\infty} \mathcal{N}_i(\langle A_i \rangle, \sigma(A_i)) \ln\left(\frac{\mathcal{N}_i(\langle A_i \rangle, \sigma(A_i))}{q_i(A_i)}\right) dA_i, \end{aligned} \quad (26)$$

which implies that the Kullback–Leibler divergence monotonously grows with

$$D(\mathcal{N}_1, \mathcal{N}_2, \dots, \mathcal{N}_N||q_1, q_2, \dots, q_N) \propto N \ln(\mathcal{N}(\langle x \rangle, \sigma(x))). \quad (27)$$

In consequence, the model uncertainty for coupled unit operations also grows linearly with  $\mathcal{O}(N)$ , which results in  $D(\mathcal{N}_1, \mathcal{N}_2, \dots, \mathcal{N}_N||q_1, q_2, \dots, q_N) \rightarrow \infty$  for  $N \rightarrow \infty$ .

In contrast, coupled Bayesian predictive ensemble distributions reveal

$$D(\rho_1, \rho_2, \dots, \rho_N||q_1, q_2, \dots, q_N) \propto \ln \rho_N(A_N|A_{N-1}, A_{N-2}, \dots, A_1) \quad (28)$$

such that  $D(\rho_1, \rho_2, \dots, \rho_N||q_1, q_2, \dots, q_N) \ll \infty$  stays finite for  $N \rightarrow \infty$ . Hence, the accuracy of coupled model predictions grows for point-like predictions via added stochastic noise predictions to conditional probabilities for the Bayesian predictive ensemble method. It is thus of significant importance to consider conditional probabilities for holistic process models in order to reduce the progression of errors and uncertainties to a reasonable level.

By means of holistic process models, it becomes clear that the coupled model design space  $\Omega$  of parameter settings for  $\Gamma$  becomes accessible [1]. If we focus on a certain process outcome like the product concentration  $C(\Gamma)$  as a function of  $\Gamma$ , the corresponding simulations for different parameter settings can be used to find the global optimum value of  $C(\Gamma_{\max})$  in accordance with

$$C(\Gamma_{\max}) = \max_{\Gamma \in \Omega} C(\Gamma). \quad (29)$$

Noteworthy, the search of global optima in a high dimensional space is a challenging task. Besides the fact that it is not clear if one global maximum value exists or even none or many, the search for these optimal conditions is usually performed via stochastic or deterministic optimization algorithms [44–46].

### 3. Numerical Example

#### 3.1. Numerical Details

The Bayesian predictive ensemble, as introduced in the previous section, is used in combination with process steps for further illustration. We consider two simplified unit operations for the sake of clarity, which can be associated with a filtration and a first order chemical reaction step. All of the source code was written in Python 3.9.1 [47] in combination with the modules NumPy 1.19.5 [48], scikit-learn 1.0.1 [49], and PyMC3 3.11.4 [50]. In addition to linear regression (LR), we use random forest (RF) [51], decision tree (DT) [52], extra tree (ET) [53], gradient boosting (GB) [54], and Gaussian process (GP) [55] approaches as non-parametric models for the prediction of process outcomes. The number of estimators in the RF and GB models were chosen as 100 and we used a Gaussian white noise kernel for the GP approach. All Bayesian inference [39] calculations were performed through pyMC3. A standard MCMC approach with an acceptance criterion of 0.95 was used for all calculations. Before effective calculations, we conducted an equilibration phase of 2000 draws for each of the two chains and hereafter the production run of 6000 draws, also each for the two chains. All values are in dimensionless units for the sake of clarity. We mainly focused on ensemble methods like GB, RF, and ET, which are less prone to overfitting.

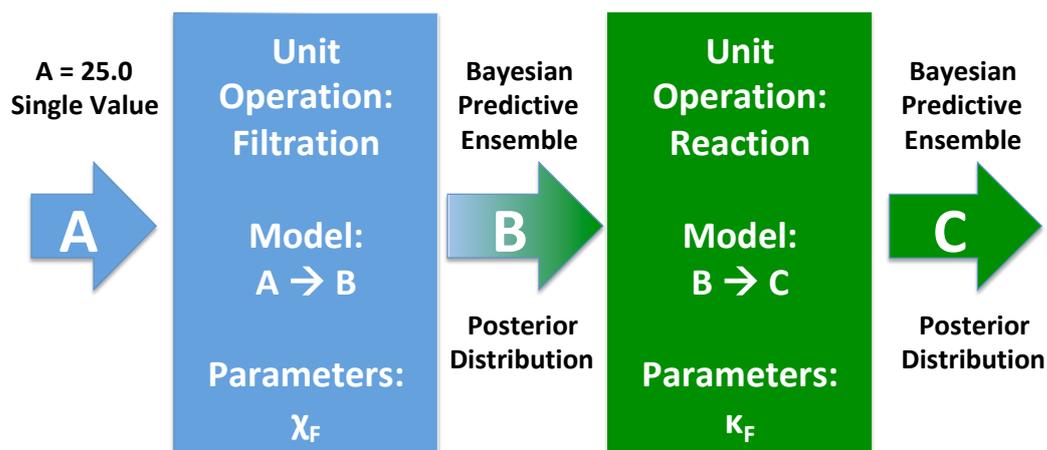
The first unit operation can be regarded as a filtration step where a concentration  $A$  is filtered to achieve a final concentration  $B$ . The following linear relation is used

$$B = \chi A + \eta_1 \quad (30)$$

where the filter constant was chosen as  $\chi = 0.5$ , including the measurement uncertainty  $\langle \eta_1(t) \rangle = 0$  and  $\langle \eta_1(t)\eta_1(t') \rangle = 0.2 \delta(t - t')$ . The second unit operation can be regarded as a chemical reaction step in which the concentration  $B$  transforms into the final product concentration  $C$  according to the first order chemical reaction

$$C = B e^{-k\tau} + \eta_2 \quad (31)$$

with the rate constant  $k = 1$ ,  $\langle \eta_2(t) \rangle = 0$  and  $\langle \eta_2(t)\eta_2(t') \rangle = 0.1 \delta(t - t')$ . The choice of such simple equations for unit operations allows us to calculate values of  $B$  and  $C$  analytically and to compare the outcomes with the model predictions accordingly. Furthermore, we can discuss basic features of the method without having to focus on specifics of more complex unit operations. In terms of a combined approach, the individual models were coupled to predict the outcomes of a holistic process model for the filtration and the reaction step. The coupled unit operation steps are schematically shown in Figure 2. The first unit operation model predicts the concentration  $B$  from a fixed value of the concentration  $A$  in accordance with Equation (30). From the corresponding posterior distribution of  $B$  as calculated by the Bayesian predictive ensemble, certain values were randomly drawn, which are then used for the calculation of  $C$  in accordance with Equation (31). Noteworthy, we limit the calculations to only six training data points for each unit operation in order to mimic realistic settings for parameterization purposes.



**Figure 2.** Schematic representation of the coupled unit operations in terms of a Bayesian network [40]: (i) filtering and (ii) chemical reaction. The models predict the concentration  $B$  from a fixed value of the concentration  $A = 25.0$  in accordance with Equation (30). From the corresponding posterior distribution of  $B$  as calculated by the Bayesian predictive ensemble, 100 values were randomly drawn, which were then used for the calculation of  $C$  in accordance with Equation (31). The species  $B$  react to species  $C$  with a certain time constant  $\kappa_F$ .

### 3.2. Training Phase: Unit Operation Models

For reliable estimates of single unit operation outcomes, we first studied the predictive accuracy of several non-parametric models for each process step individually. As training data, we chose six values of  $A_{tr}$ , which were used for the calculation of  $B_{tr}$  in accordance with Equation (30). The values of  $A_{tr}$  were uniformly distributed between 49 and 51 to mimic certain designs of the experiment conditions [56]. The low amount of training or parameterization data is a standard problem for biotechnological model development as experimental work is time-consuming and costly. The corresponding six data points were used to train individual LR, GP, GB, RF, DT, and ET models. Hereafter, we again draw six randomly chosen values of  $A$  as testing data [19] for the calculation of the corresponding  $B_{Exp}$  values (Equation (30)) and for predictions of the models  $B_{pred}$ . The standard values [19], like the root-mean squared error (RMSE), the normalized root-mean squared error (nRMSE), and the mean absolute error (MSE) between predicted  $B_{pred}$  and analytical values  $B_{Exp}$  for the individual models are summarized in Table 1. It can clearly be seen that the predictive accuracy of all models is rather low. This can be rationalized by the low amount of training data and the corresponding influence of the measurement noise on  $B_{Exp}$  from Equation (30). However, it comes out that the GP model reveals the highest predictive accuracy despite certain deviations from the predicted values. The corresponding nRMSE value of 0.54 is quite high, but is still acceptable for the sake of our simple example. Based on these values, we chose the GP model as the corresponding representative approach for the filtration unit operation. For the development of different models in accordance with the Bayesian predictive ensemble, we split the previous training data set into four different data sets with a test:training ratio of 20:80. Based on the individual data, we trained four different GP models with each of these reduced training sets in accordance with the concept of bootstrapping [31,32].

A comparable procedure for estimating the model with the highest predictive accuracy was also performed for the chemical reaction unit operation.

**Table 1.** Root-mean squared errors (RMSE), normalized root-mean squared errors (nRMSE), mean absolute errors (MAE), and Pearson correlation coefficients  $R^2$  for the individual models regarding the test data between predicted  $B_{pred}$  and observed values  $B_{Exp}$  for the filtration step.

Approach	RMSE	nRMSE	MAE
GP	0.19	0.54	0.03
LR	0.19	0.57	0.04
RF	0.34	0.99	0.11
ET	0.36	1.06	0.13
DT	0.39	1.14	0.15
GB	0.39	1.14	0.15

We thus first prepared the training data set by choosing six values of  $B$ , which were inserted into Equation (31) for the calculation of  $C_{Exp}$ . The values of  $B$  were uniformly distributed between 24 and 26 to mimic a certain design of experiment conditions [56] and the values of  $\tau$  were uniformly distributed between 0.4 and 0.6. These data points were used to train individual LR, GP, GB, RF, DT, and ET models. Hereafter, we repeated the same procedure such that we draw six random values from a uniform distribution of  $B$  and inserted them into Equation (31) for the calculation of  $C_{Exp}$  and for  $C_{pred}$  from model predictions. The corresponding RMSE, nRMSE, and MAE values for the model predictions are shown in Table 2.

**Table 2.** Root-mean squared errors (RMSE), normalized root-mean squared errors (nRMSE), mean absolute errors (MAE), and Pearson correlation coefficients  $R^2$  for the individual models regarding the test data between predicted  $C_{pred}$  and observed values  $C_{Exp}$  for the chemical reaction unit operation step.

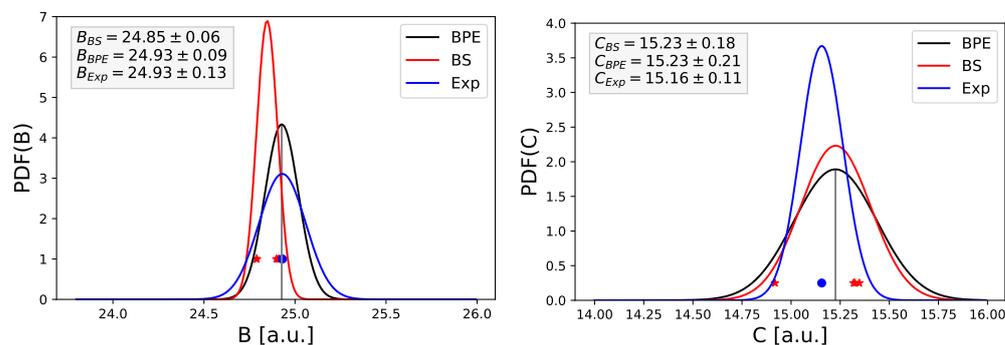
Approach	RMSE	nRMSE	MAE
LR	0.25	0.25	0.06
ET	0.57	0.55	0.32
RF	0.71	0.69	0.51
GB	0.79	0.77	0.63
DT	0.83	0.81	0.69
GP	1.27	1.24	1.62

It can clearly be seen that the LR models reveal the highest predictive accuracy. Nevertheless, the nRMSE values are still rather high, which highlights certain deviations between the predicted and the observed values. For reasons of simplicity, we ignore a further improvement of the models and choose the pre-trained LR model as a corresponding representative for the second unit operation. In agreement with the previous unit operation, we split the corresponding training data set into four different data sets with a test:training ratio of 20:80. Based on the individual data, we trained four reduced LR models that are used for the Bayesian predictive ensemble in accordance with bootstrapping calculations [31,32] in the next subsection.

### 3.3. Bayesian Predictive Ensemble: Single Unit Operations

The corresponding pre-trained GP and LR models are used for the prediction of process outcomes for the filtration and chemical reaction unit operation steps. Hence, we chose an initial value of  $A = 50.0$ , which was inserted into the bootstrapped GP models and the GP model trained with the full data set. The corresponding predictions of  $B_{pred}$  from the four bootstrapped GP models are shown as red stars in Figure 3. The prediction of  $B_{pred}$  from the GP model with the full training set is used as mean value for the prior distribution function in combination with the corresponding standard deviation from the bootstrapped GP model predictions in accordance with step 1 of the algorithm. Furthermore, the values of  $B_{pred}$  from the bootstrapped models were used as observed values for the corresponding

MCMC calculations (step 2 of the algorithm). The outcomes of these calculations for the posterior distribution  $B_{BPE}$  are shown in Figure 3.



**Figure 3.** (Left side): Individual model predictions of  $B$  for a value of  $A = 50.0$  from the bootstrapped GP models (red stars) in combination with the mean value from the observed values (blue circle). The blue curve shows the observed experimental distribution, the red curve the simplified estimate from the bootstrapped GP model predictions, and the black curve the corresponding distribution of  $B$  from the Bayesian predictive ensemble calculations. (Right side): Individual model predictions of  $C$  for values of  $B = 24.87$  and  $\tau = 0.5$  from the bootstrapped GP models (red stars) in combination with the mean value from the observed values (blue circle). The blue curve shows the observed experimental distribution, the red curve the simplified estimate from the bootstrapped GP model predictions, and the black curve the corresponding distribution of  $C$  from the Bayesian predictive ensemble calculations. All units are dimensionless.

In addition to the predictions from the Bayesian predictive ensemble, we also visualized the corresponding distribution of observed values  $B_{Exp}$  as reference and the distribution of predictions  $B_{BS}$  from the bootstrapping calculations as simplified estimates. One can clearly see that the distribution of  $B_{BS}$  is different from  $B_{BSE}$  in terms of the mean value and the standard deviation. In comparison to the observed values, one can conclude that the distribution of  $B_{BPE}$  shows more similarities than  $B_{BS}$  in terms of similar mean values and standard deviations. Although the stochastic nature of the experimental values  $B_{Exp}$  from six randomly chosen values of  $A$  has to be noted, one can assume that the Bayesian predictive ensemble calculations provides a better description than the simplified bootstrapping estimates. This can be partly understood by the low amount of training data, which restricts the number of bootstrapped models such that these predictions are of minor accuracy.

Corresponding conclusions can also be drawn for the predictions of the chemical reaction unit operation as shown in the right side of Figure 3. We chose initial values of  $B = 24.87$  and  $\tau = 0.5$  that were inserted into the bootstrapped GP models and the full GP model. The prediction of the full GP model in combination with the standard deviation from the predictions of the bootstrapped GP models were used as mean value and standard deviation, respectively, for the prior distribution function of  $C$  (in accordance with step 1 of the algorithm). The corresponding predictions from the bootstrapped LR models were used as observed values for the calculation of the posterior distribution function in accordance with step 2 of the algorithm. The posterior distribution of  $C_{BPE}$  from the Bayesian predictive ensemble calculations after the MCMC steps in combination with the bootstrapping distribution  $C_{BS}$  are shown as black and red curves, respectively. It can be seen that the mean values  $C_{BS}$  and  $C_{BPE}$  coincide, but the standard deviations differ. Hence, the standard deviation from the BPE calculations is slightly broader and thus overlaps with the corresponding observed experimental distribution function  $C_{Exp}$ . Although the accuracy is not that high when compared to the previous unit operation, it can be concluded that the combined BPE calculations provide reliable estimates for the distribution of process outcomes with higher accuracy than simple Bootstrapping predictions. The corresponding

mean values and standard deviations for all distributions and unit operations are shown in Table 3.

**Table 3.** Mean values  $\langle B \rangle$ ,  $\langle c \rangle$ , standard deviations  $\sigma B$ ,  $\sigma C$  for experimental values (Exp), and calculations from the Bayesian predictive ensemble (BPE) and the simple estimates from the bootstrapped models (BS). The results of the best ML model for a singular prediction of  $\langle B \rangle$  (full GP model) and  $\langle C \rangle$  (full LR model) are also shown.

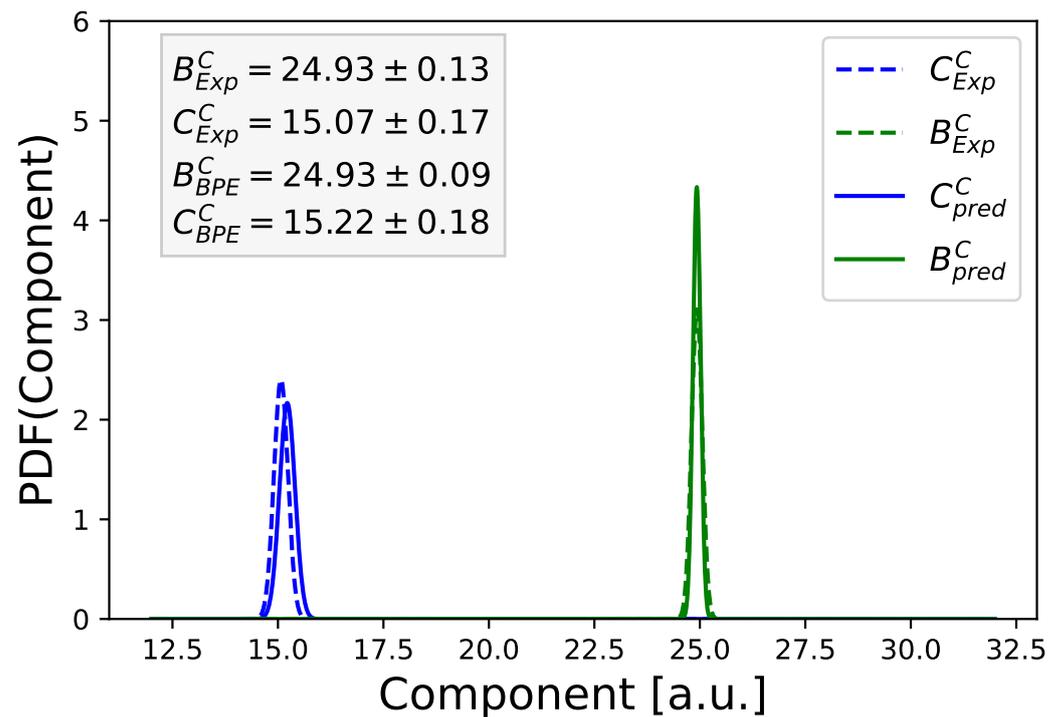
Method	$\langle B \rangle$	$\sigma(B)$	$\langle C \rangle$	$\sigma(C)$
Exp	24.93	0.13	15.16	0.11
BS	24.85	0.06	15.23	0.18
BPE	24.93	0.09	15.23	0.21
ML	25.04	0.00	15.64	0.00

Moreover, we included point-like predictions for  $B$  and  $C$  for the corresponding values from the full GP and the full LR model. As can be seen, the best ML models provide good estimates but no reasonable standard deviation for estimating the model uncertainty. However, their usage as mean value for the prior distribution function allows us to improve the corresponding values for the posterior distribution function. Hence, it can be concluded that the combination of both aspects in terms of the Bayesian predictive ensemble calculations provides good estimates for model outcome ranges and mean values.

### 3.4. Coupled Unit Operations: Holistic Process Models

In this subsection, we evaluate the corresponding model predictions for the coupled unit operation steps of filtration and chemical reaction. For all predictions, the pre-trained models from the previous subsection were used. The initial starting value for the first unit operation (filtration) was chosen as  $A = 25.0$ . Thus, the calculations as well as the values for the distribution of  $B_{pred}^C$  were identical to the procedure as described in the previous subsection. The corresponding posterior distribution  $B_{BPE}^C$  when considered as transformation function was used as starting distribution for the second unit operation in accordance with step 2 of the algorithm. We draw six random samples from the posterior distribution  $B_{pred}^C$  in combination with a fixed value of  $\tau = 0.5$  (Equation (31)) which are used as input values for the corresponding pre-trained bootstrapped LR models for the chemical reaction step in accordance with step 3 of the algorithm. The outcomes of these bootstrapped GP models are used as observed values for the calculation of the posterior distribution function  $C_{BPE}^C$ . Moreover, we used the six random values of  $B_{BPE}^C$  as initial values for the full LR model. The resulting distribution is used as prior distribution function for the MCMC calculations in accordance with step 2 for the calculation of  $C_{pred}^C$ . The corresponding posterior distributions for the coupled outcomes  $B_{pred}^C$  and  $C_{pred}^C$  are shown in Figure 4 in combination with the observed coupled values  $B_{Exp}^C$  and  $C_{Exp}^C$ .

It can clearly be seen that the distribution for  $C_{BPE}^C$  is roughly identical when compared to  $C_{BPE}$  for single unit operation models (Table 3). For single unit operation models, we already used as input values a comparable value of  $B = 24.87$  in order to estimate the accuracy of the approaches after coupling. Hence, it can be concluded that the coupled approaches reveal a high accuracy with reliable estimates for the predicted distributions. Thus, all corresponding results reveal that the Bayesian predictive ensemble can be regarded as a fast and straightforward approach to estimate the mean outcomes for coupled and individual unit operation models in good agreement with the observed distribution.



**Figure 4.** Posterior distributions of  $B_{pred}^C$  (green solid curve) and  $C_{pred}^C$  (blue solid curve) from Bayesian predictive ensemble calculations for coupled unit operations. The corresponding observed values are shown as dashed curves. More details can be found in the main text. All units are dimensionless.

#### 4. Summary and Conclusions

In this study, we presented a new approach in accordance with Bayesian inference and various model predictions for the advanced calculation of single unit as well as coupled end-to-end process outcomes in combination with the corresponding uncertainty ranges. The Bayesian predictive ensemble method provides a more robust interpretation of model predictions and parameter ranges when compared to discrete or stochastic connections. Due to the coupling of individual unit operations, a consistent consideration of conditional probability distributions is introduced, which crucially relies on pre-defined parameter uncertainty ranges. Hence, the standard deviations reflect the general model uncertainty in combination with the variability of the observed data which was used for parameterization or training purposes.

It has to be noted that this concept is also applicable for high-dimensional design spaces as well as multivariate parameter distribution functions with certain limitations in terms of MCMC convergence issues. Furthermore, the prior distributions for the Bayesian inference calculations can include as much information as possible, including experimental data as well, which thus increases the accuracy of the posterior distribution functions. In addition to using different model types like neural network- or decision tree-based approaches, it is also possible to focus on a single type of model with slight changes in the hyper parameter settings or in combination with bootstrapping concepts. It has to be noted that the Bayesian framework reveals its full advantages for a small number of model predictions. Herewith, we reweigh the corresponding model predictions under the assumption that the majority of models shows a sufficient predictive accuracy.

In summary, we presented a robust, extensible, and consistent approach for single and coupled unit operation models. The main advantages are the probabilistic interpretation of outcome ranges in terms of uncertainty intervals. In consequence, we explicitly use different models for the prediction of process outcomes. In terms of Bayesian inference and the consideration of conditional probabilities, we incorporate the outcomes of previous unit operations as implicit knowledge into the resulting posterior distributions. Thereby, we present a consistent approach to transfer the resulting distributions with meaningful and

interpretable ranges between the individual unit operations. In contrast to other approaches, the proposed transfer functions including prior and posterior distributions have well defined and interpretable ranges. Furthermore, we have shown that the deviations from the true distribution deviate only marginally after certain process steps in contrast to other approaches. Hence, our approach provides a controllable and finite error progression for any holistic process description in terms of parametric or non-parametric models. Noteworthy, the method is also applicable for limited experimental data sets in order to obtain reliable distributions from MCMC calculations. Despite all benefits, it has to be noted that the MCMC calculations are rather time consuming and may show some convergence issues for poorly behaving model predictions. The computational costs for these calculations increase linearly with the amount of considered unit operations and the number of dimensions in the model design space. It has to be noted that our concept is not restricted to chemical or biotechnological process models, as it also can be used for other machine learning applications. As an outlook, we assume that the presented approach provides a robust basis for the development of future holistic process models to support the development of optimized and robust manufacturing processes.

**Author Contributions:** Conceptualization, J.S.; methodology, J.S.; formal analysis, J.S.; validation, J.S., T.E., I.-T.H., B.S., M.M. and E.B.; supervision, J.S., L.M.H., R.G., M.L., E.B. and A.J.; project administration, J.S., R.G., M.L., A.J. and E.B.; funding acquisition: J.S., L.M.H., A.J. and E.B.; writing—original draft preparation: J.S., L.M.H., T.E. and E.B., writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Boehringer Ingelheim Pharma GmbH & Co. KG, Development Biologicals and the Digital Innovation Unit.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We thank Hermann Schuchnigg for valuable discussions. We acknowledge funding from the Digital Innovation Unit and Boehringer Ingelheim Pharma GmbH & Co. KG.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

NOR	Normal operation range
PAR	Proven acceptable range
PAT	Process Analytical Technology
BS	Bootstrapping
MCMC	Markov Chain Monte Carlo
LR	Linear Regression
GB	Gradient Boosting
DT	Decision Trees
ET	Extra Trees
GP	Gaussian Processes
RF	Random Forest
RMSE	Root-Mean Squared Error
nRMSE	Normalized Root-Mean Squared Error
MAE	Mean Absolute Error
BPE	Bayesian Predictive Ensemble

## References

1. Smiatek, J.; Jung, A.; Bluhmki, E. Towards a Digital Bioprocess Replica: Computational Approaches in Biopharmaceutical Development and Manufacturing. *Trends Biotechnol.* **2020**, *38*, 1141–1153. [[CrossRef](#)] [[PubMed](#)]
2. Brass, J.M.; Hoeks, F.W.J.M.M.; Rohner, M. Application of Modelling Techniques for the Improvement of Industrial Bioprocesses. *J. Biotechnol.* **1997**, *59*, 63–72. [[CrossRef](#)]
3. Narayanan, H.; Luna, M.F.; von Stosch, M.; Cruz Bournazou, M.N.; Polotti, G.; Morbidelli, M.; Butté, A.; Sokolov, M. Bioprocessing in the Digital Age: The Role of Process Models. *Biotechnol. J.* **2020**, *15*, 1900172. [[CrossRef](#)] [[PubMed](#)]
4. Herwig, C.; Pörtner, R.; Möller, J. (Eds.) *Digital Twins: Applications to the Design and Optimization of Bioprocesses*; Advances in Biochemical Engineering and Biotechnology; Springer Nature: Berlin/Heidelberg, Germany, 2021; Volume 177.
5. Narayanan, H.; Sokolov, M.; Morbidelli, M.; Butté, A. A New Generation of Predictive Models—The added Value of Hybrid Models for Manufacturing Processes of Therapeutic Proteins. *Biotechnol. Bioeng.* **2019**, *116*, 2540–2549. [[CrossRef](#)]
6. von Stosch, M.; Davy, S.; Francois, K.; Galvanauskas, V.; Hamelink, J.M.; Luebbert, A.; Mayer, M.; Oliveira, R.; O’Kennedy, R.; Rice, P.; et al. Hybrid Modeling for Quality by Design and PAT—Benefits and Challenges of Applications in Biopharmaceutical Industry. *Biotechnol. J.* **2014**, *9*, 719–726. [[CrossRef](#)] [[PubMed](#)]
7. Bayer, B.; von Stosch, M.; Striedner, G.; Duerkop, M. Comparison of Modeling Methods for DoE-Based Holistic Upstream Process Characterization. *Biotechnol. J.* **2020**, *15*, 1900551. [[CrossRef](#)]
8. Sokolov, M.; Soos, M.; Neunstoeklin, B.; Morbidelli, M.; Butté, A.; Leardi, R.; Solacroup, T.; Stettler, M.; Broly, H. Fingerprint Detection and Process Prediction by Multivariate Analysis of Fed-Batch Monoclonal Antibody Cell Culture Data. *Biotechnol. Prog.* **2015**, *31*, 1633–1644. [[CrossRef](#)]
9. Rischawy, F.; Saleh, D.; Hahn, T.; Oelmeier, S.; Spitz, J.; Kluters, S. Good Modeling Practice for Industrial Chromatography: Mechanistic Modeling of Ion Exchange Chromatography of a Bispecific Antibody. *Comput. Chem. Eng.* **2019**, *130*, 106532. [[CrossRef](#)]
10. Briskot, T.; Stückler, F.; Wittkopp, F.; Williams, C.; Yang, J.; Konrad, S.; Doninger, K.; Griesbach, J.; Bennecke, M.; Hepbildikler, S.; et al. Prediction Uncertainty Assessment of Chromatography Models using Bayesian Inference. *J. Chromatogr. A* **2019**, *1587*, 101–110. [[CrossRef](#)]
11. Saleh, D.; Wang, G.; Müller, B.; Rischawy, F.; Kluters, S.; Studts, J.; Hubbuch, J. Straightforward Method for Calibration of Mechanistic Cation Exchange Chromatography Models for Industrial Applications. *Biotechnol. Prog.* **2020**, *36*, e2984. [[CrossRef](#)]
12. Ulonska, S.; Kroll, P.; Fricke, J.; Clemens, C.; Voges, R.; Müller, M.M.; Herwig, C. Workflow for Target-Oriented Parametrization of an Enhanced Mechanistic Cell Culture Model. *Biotechnol. J.* **2018**, *13*, 1700395. [[CrossRef](#)] [[PubMed](#)]
13. Moser, A.; Appl, C.; Brüning, S.; Hass, V.C. Mechanistic Mathematical Models as a Basis for Digital Twins. In *Digital Twins*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 133–180.
14. Narayanan, H.; Seidler, T.; Luna, M.F.; Sokolov, M.; Morbidelli, M.; Butté, A. Hybrid Models for the Simulation and Prediction of Chromatographic Processes for Protein Capture. *J. Chromatogr. A* **2021**, *1650*, 462248. [[CrossRef](#)]
15. Smiatek, J.; Clemens, C.; Herrera, L.M.; Arnold, S.; Knapp, B.; Presser, B.; Jung, A.; Wucherpennig, T.; Bluhmki, E. Generic and Specific Recurrent Neural Network Models: Applications for Large and Small Scale Biopharmaceutical Upstream Processes. *Biotechnol. Rep.* **2021**, *31*, e00640. [[CrossRef](#)]
16. Narayanan, H.; Dingfelder, F.; Condado Morales, I.; Patel, B.; Heding, K.E.; Bjelke, J.R.; Egebjerg, T.; Butté, A.; Sokolov, M.; Lorenzen, N.; et al. Design of Biopharmaceutical Formulations Accelerated by Machine Learning. *Mol. Pharm.* **2021**, *18*, 3843–3853. [[CrossRef](#)] [[PubMed](#)]
17. Narayanan, H.; Dingfelder, F.; Butté, A.; Lorenzen, N.; Sokolov, M.; Arosio, P. Machine Learning for Biologics: Opportunities for Protein Engineering, Developability, and Formulation. *Trends Pharmacol. Sci.* **2021**, *42*, 151–165. [[CrossRef](#)] [[PubMed](#)]
18. Sokolov, M. Decision Making and Risk Management in Biopharmaceutical Engineering—Opportunities in the Age of Covid-19 and Digitalization. *Ind. Eng. Chem. Res.* **2020**, *59*, 17587–17592. [[CrossRef](#)]
19. Smiatek, J.; Jung, A.; Bluhmki, E. Validation Is Not Verification: Precise Terminology and Scientific Methods in Bioprocess Modeling. *Trends Biotechnol.* **2021**, *39*, 1117–1119. [[CrossRef](#)]
20. Rajamanickam, V.; Babel, H.; Montano-Herrera, L.; Ehsani, A.; Stiefel, F.; Haider, S.; Presser, B.; Knapp, B. About Model Validation in Bioprocessing. *Processes* **2021**, *9*, 961. [[CrossRef](#)]
21. Nargund, S.; Guenther, K.; Mauch, K. The Move toward Biopharma 4.0: Insilico Biotechnology Develops "Smart" Processes that Benefit Biomanufacturing through Digital Twins. *Gen. Eng. Biotechnol.* **2019**, *39*, 53–55. [[CrossRef](#)]
22. Zahel, T.; Hauer, S.; Mueller, E.; Murphy, P.; Abad, S.; Vasilieva, E.; Maurer, D.; Brocard, C.; Reinisch, D.; Sagmeister, P.; et al. Integrated Process Modeling—A Process Validation Life Cycle Companion. *Bioengineering* **2017**, *4*, 86. [[CrossRef](#)]
23. Chen, Y.; Yang, O.; Sampat, C.; Bhalode, P.; Ramachandran, R.; Ierapetritou, M. Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review. *Processes* **2020**, *8*, 1088. [[CrossRef](#)]
24. Park, S.Y.; Park, C.H.; Choi, D.H.; Hong, J.K.; Lee, D.Y. Bioprocess Digital Twins of Mammalian Cell Culture for Advanced Biomanufacturing. *Curr. Opin. Chem. Eng.* **2021**, *33*, 100702. [[CrossRef](#)]
25. Taylor, C.; Marschall, L.; Kunzelmann, M.; Richter, M.; Rudolph, F.; Vajda, J.; Presser, B.; Zahel, T.; Studts, J.; Herwig, C. Integrated Process Model Applications Linking Bioprocess Development to Quality by Design Milestones. *Bioengineering* **2021**, *8*, 156. [[CrossRef](#)] [[PubMed](#)]

26. Zobel-Roos, S.; Schmidt, A.; Uhlenbrock, L.; Ditz, R.; Köster, D.; Strube, J. Digital Twins in Biomanufacturing. In *Digital Twins*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 181–262.
27. Gomis-Fons, J.; Schwarz, H.; Zhang, L.; Andersson, N.; Nilsson, B.; Castan, A.; Solbrand, A.; Stevenson, J.; Chotteau, V. Model-Based Design and Control of a Small-Scale Integrated Continuous End-to-End mAb Platform. *Biotechnol. Prog.* **2020**, *36*, e2995. [[CrossRef](#)] [[PubMed](#)]
28. Appl, C.; Moser, A.; Baganz, F.; Hass, V.C. Digital Twins for Bioprocess Control Strategy Development and Realisation. In *Digital Twins*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 63–94.
29. Maloney, A.J.; Icten, E.; Capellades, G.; Beaver, M.G.; Zhu, X.; Graham, L.R.; Brown, D.B.; Griffin, D.J.; Sangodkar, R.; Allian, A.; et al. A Virtual Plant for Integrated Continuous Manufacturing of a Carfilzomib Drug Substance Intermediate, part 3: Manganese-Catalyzed Asymmetric Epoxidation, Crystallization, and Filtration. *Organ. Proc. Res. Dev.* **2020**, *24*, 1891–1908. [[CrossRef](#)]
30. DiCiccio, T.J.; Efron, B. Bootstrap Confidence Intervals. *Statist. Sci.* **1996**, *11*, 189–228. [[CrossRef](#)]
31. Efron, B. Bootstrap Methods: Another Look at the Jackknife. In *Breakthroughs in Statistics: Methodology and Distribution*; Kotz, S., Johnson, N.L., Eds.; Springer: New York, NY, USA, 1992; pp. 569–593.
32. Efron, B. Second Thoughts on the Bootstrap. *Stat. Sci.* **2003**, *18*, 135–140. [[CrossRef](#)]
33. Efron, B. Bayes, Oracle Bayes and Empirical Bayes. *Stat. Sci.* **2019**, *34*, 177–201. [[CrossRef](#)]
34. Efron, B. Bayes' Theorem in the 21st Century. *Science* **2013**, *340*, 1177–1178. [[CrossRef](#)]
35. Roman, R.C.; Precup, R.E.; Petriu, E.M. Hybrid Data-Driven Fuzzy Active Disturbance Rejection Control for Tower Crane Systems. *Eur. J. Control* **2021**, *58*, 373–387. [[CrossRef](#)]
36. Zhu, Z.; Pan, Y.; Zhou, Q.; Lu, C. Event-Triggered Adaptive Fuzzy Control for Stochastic Nonlinear Systems with Unmeasured States and Unknown Backlash-Like Hysteresis. *IEEE Transact. Fuzzy Sys.* **2020**, *29*, 1273–1283. [[CrossRef](#)]
37. Walters, P. *An Introduction to Ergodic Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2000.
38. Prigogine, I. *Non-Equilibrium Statistical Mechanics*; Courier Dover Publications: Mineola, NY, USA, 2017.
39. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
40. Jensen, F.V. *Bayesian Networks and Decision Graphs*; Springer: Berlin/Heidelberg, Germany, 2001.
41. Carlin, B.P.; Louis, T.A. *Bayesian Methods for Data Analysis*; Chapman and Hall/CRC: London, UK, 2008.
42. Geyer, C.J. Introduction to Markov Chain Monte Carlo. In *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*; CRC Press: Boca Raton, FL, USA, 2011.
43. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
44. Homem-de Mello, T.; Bayraksan, G. Monte Carlo Sampling-Based Methods for Stochastic Optimization. *Surv. Oper. Res. Manag. Sci.* **2014**, *19*, 56–85. [[CrossRef](#)]
45. Fouskakis, D.; Draper, D. Stochastic Optimization: A Review. *Int. Stat. Rev.* **2002**, *70*, 315–349. [[CrossRef](#)]
46. Ruder, S. An Overview of Gradient Descent Optimization Algorithms. *arXiv* **2016**, arXiv:1609.04747.
47. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009.
48. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)]
49. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
50. Salvatier, J.; Wiecki, T.V.; Fonnesbeck, C. Probabilistic Programming in Python using PyMC3. *PeerJ Comput. Sci.* **2016**, *2*, e55. [[CrossRef](#)]
51. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
52. Kamiński, B.; Jakubczyk, M.; Szufel, P. A Framework for Sensitivity Analysis of Decision Trees. *Cent. Eur. J. Oper. Res.* **2018**, *26*, 135–159. [[CrossRef](#)]
53. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach. Learn.* **2006**, *63*, 3–42. [[CrossRef](#)]
54. Mason, L.; Baxter, J.; Bartlett, P.; Frean, M. Boosting Algorithms as Gradient Descent. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 1999; Volume 12, pp. 512–518.
55. Williams, C.K.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006.
56. Politis, S.N.; Colombo, P.; Colombo, G.; Rekkas, D.M. Design of Experiments (DoE) in Pharmaceutical Development. *Drug Dev. Ind. Pharm.* **2017**, *43*, 889–901. [[CrossRef](#)] [[PubMed](#)]