

Article

Exploiting Domain Partition in Response Function-Based Dynamic Surrogate Modeling: A Continuous Crystallizer Study

Alessandro Di Pretoro ^{*}, Ludovic Montastruc and Stéphane Negny

Laboratoire de Génie Chimique, Université de Toulouse, CNRS/INP/UPS, 31432 Toulouse, France; ludovic.montastruc@ensiacet.fr (L.M.); stephane.negny@ensiacet.fr (S.N.)

* Correspondence: alessandro.dipretoro@ensiacet.fr

Abstract: Given the exponential rise in the amount of data requiring processing in all engineering fields, phenomenological models have become computationally cumbersome. For this reason, more efficient data-driven models have been recently used with the purpose of substantially reducing simulation computational times. However, especially in process engineering, the majority of the proposed surrogate models address steady-state problems, while poor studies refer to dynamic simulation modeling. For this reason, using a response function-based approach, a crystallization unit case study was set up in order to derive a dynamic data-driven model for crystal growth whose characteristic differential parameters are derived via Response Surface Methodology. In particular, multiple independent variables were considered, and a well-established sampling technique was exploited for sample generation. Then, different sample sizes were tested and compared in terms of accuracy indicators. Finally, the domain partition strategy was exploited in order to show its relevant impact on the final model accuracy. In conclusion, the outcome of this study proved that the proposed procedure is a suitable methodology for dynamic system metamodeling, as it shows good compliance and relevant improvement in terms of computational time. In terms of future research perspectives, testing the proposed procedure on different systems and in other research fields would allow for greater improvement and would, eventually, extend its validity.

Keywords: Latin Hypercube Sampling; metamodels; response function; process systems



Citation: Di Pretoro, A.; Montastruc, L.; Negny, S. Exploiting Domain Partition in Response Function-Based Dynamic Surrogate Modeling: A Continuous Crystallizer Study. *Dynamics* **2024**, *4*, 208–221. <https://doi.org/10.3390/dynamics4020012>

Academic Editor: Christos Volos

Received: 4 February 2024

Revised: 3 March 2024

Accepted: 7 March 2024

Published: 26 March 2024



Copyright: © 2024 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/>).

1. Introduction

1.1. An Overview of the Literature on Surrogate Modeling

During the last decade, the digital transition has become, by far, the most impactful innovation in all scientific domains. In the engineering field, the potential of artificial intelligence and, more generally, of approaches based on data have called into question the effectiveness of well-established methodologies and their performance by reopening several challenges related to their continued development. In process engineering applications, the computational challenges of primary concern, particularly those that would considerably benefit from an improvement in effectiveness, involve the problems of superstructure optimization, i.e., the MINLP problem and Real-Time Dynamic Optimization [1]. In order to mitigate the complexity of these problems, several researchers have started exploring the possibility of replacing phenomenological models with data-driven ones. The latter are better known as surrogate models and can vary in terms of structure and derivation procedure [2].

When dealing with surrogate modeling, a methodology where the modeling phase is anticipated by the Design of Experiment (DoE) step and followed by the assessment of model quality indicators is usually implemented. The DoE procedure consists of the selection of the data sampling strategy; two main categories of sampling can be identified, namely, the once-through and iterative approaches. In the former case, a single sampling phase is conducted, and the experimental points are usually selected with the purpose of

providing complete and uniform domain filling. In the latter case, a recursive modeling procedure is performed by updating the first sample according to the outcome of the modeling step until a given condition is satisfied. In terms of computational effort, the second method tends to be more time-consuming due to the recursive need to perform the sampling and modeling steps, while in the first method, the DoE phase occurs only once. However, iterative sampling, also known as adaptive sampling, allows the user to adjust the point selection, which the once-through strategy does not provide; consequently, it usually provides more accurate results at the cost of a longer computational time [3]. For this reason, the sampling approach presented in the literature usually depends on a specific case study and may have a considerable impact on the final results. However, independent of this choice, several sampling methodologies are commonly shared by the two methods. The most established ones are Latin Hypercube Sampling (LHS) and Sobol and Halton sampling, which has the possibility to exploit subsets by means of the domain space repartition [4].

With regard to the metamodeling procedures, two main categories can be distinguished. The first is based on data approximation by means of explicit functions, such as Response Surface Methodology (RSM) [5], Radial Basis Functions [6], or Support Vector Machine (SVM) [7]. In the second category, a probabilistic approach is used, as in the cases of Kriging [8], polynomial chaos expansion [9], or random forest [10]. However, to date, some difficulties have been encountered in the surrogate modeling domain when applied to chemical processes. First, it is difficult to find a general agreement on the optimal modeling strategy since the available research mainly focuses on specific applications carried out by means of selected methodologies; not enough comparative studies are available. Moreover, while publications usually address the positive aspects and the high-quality results, poor studies have focused on relevant discrepancies concerning the output feasibility, which can be nevertheless observed. When modeling unit operations, fundamental conservation laws, such as mass or energy conservation, are sometimes not followed, or some operating points, falling in regions without a physical meaning, can be found. In order to mitigate these issues, alternative approaches based on constrained sampling or subsets have been proposed and have demonstrated good results, but they have not yet achieved thorough and general validation [11]. Moreover, the vast majority of the proposed studies usually refer to steady-state applications, while the process dynamics domain can only be found in a few works, mostly related to process unit or energy system optimization, that are usually based on explicit analytical or statistical functions with case-specific applications [12–14]. If we consider that computationally efficient digital twins for dynamic modeling would substantially increase the effectiveness of tools such as RTDO and optimal control by enabling real-time scheduling and process parameter adjustments, it is evident that the process dynamics field deserves more attention. Furthermore, a wider range of methodologies and case studies concerning multiple fields could be tested in order to provide a better overview of the approaches that could be adapted from steady-state to dynamic applications and of those with validity limited to the former.

1.2. Context and Description of the Proposed Approach

In the light of the preliminary discussion about the current state of the art in process surrogate modeling, the main motivation behind the presented research is the need for greater knowledge and insights into data-driven opportunities for the application domain of process dynamics. Therefore, building upon the derivative-based approach proposed by Di Pretoro et al. (2021) [15], a differential response function is used as the response surface, with its characteristic parameters derived through optimization with respect to the experimental datasets.

Compared to the more common polynomial approaches or those utilizing explicit functions suitable for steady-state systems, the response function differential equation offers significant advantages when addressing dynamics. Specifically, the expression of the derivative-based model remains unchanged for any initial condition, which is not always

the case for the polynomial surrogate coefficients. Additionally, the resulting expression is time-independent, or more precisely, from a numerical perspective, it only depends on the time step size rather than the value of the time variable. Moreover, regression applied to the differential pseudo-models allows the determination of characteristic parameters related to the dynamic behavior of the system, such as the characteristic time, frequency, and dumping factor, which can be exploited for further applications in analyzing system stability and designing process control configurations.

Hence, based on this tool, the first objective of this study is to build a response function-based dynamic surrogate model for a unit operation involving chemical reactions. Specifically, the trend of computational time and accuracy indicators concerning the Design of Experiment sample size is analyzed to ascertain the existence of an optimal compromise solution. The secondary goal is to evaluate the validity limitations of the obtained outcomes across the operation domain and, if necessary, understand the criticalities in terms of model definition. Finally, if required, the third part of this work will address the implementation of adjustment strategies, such as domain partitioning, which has already proven effective in common steady-state approaches.

Concerning the case study application, the aim of this research is to establish a generic correlation between feed properties and crystal growth in crystallization equipment. Crystallizers are solid–liquid separation units designed to increase the concentration of a given solution beyond its saturation limit to induce the precipitation of the desired species and the subsequent formation of high-purity crystals. This unit is of particular interest in process engineering as it can be utilized in the production of various compounds across domains ranging from the pharmaceutical [16] to food industry [17]. Moreover, this separation process is particularly suitable for the study of dynamics due to its multiple operating modes, ranging from batch-discontinuous to continuous [18], including an intermediate fed-batch configuration [19]. As previously outlined, the main objective of the analysis is to take a step forward towards a thorough procedure for leveraging a differential function for dynamic metamodels, progressively enhancing their accuracy and application scope. In particular, given the criticalities encountered during the global modeling approach, the expedient of domain partitioning and domain-specific modeling was successfully explored and tested on the crystallizer unit, resulting in final results validated with an error lower than 0.1%. Further details regarding the proposed methodology and the case study parameters are provided in the following sections.

2. The Continuous Crystallizer Case Study

The selected process for the application of the dynamic surrogate modeling approach is continuous crystallization. Crystallization is a non-conventional unit operation aimed at precipitating a solid phase from a liquid solution based on supersaturation conditions of the selected species. Depending on the final product characteristics, the operation can be conducted using batch, semi-batch, or continuous equipment. Although several studies on crystallization can be found in the literature, most of them rarely address the unit from a macroscopic perspective. While batch and semi-batch configurations are conventionally modeled as dynamic units, continuous crystallization is typically analyzed as a steady-state unit. However, because perturbation can affect the operating parameters, the final properties of the product (especially crystal size) can substantially vary over a non-negligible transient.

A simplified scheme of the conventional continuous crystallizer unit is depicted in Figure 1. As shown, the feed solution is introduced to the unit along with a heating agent (usually steam), providing heat for the partial solvent evaporation. Consequently, the solution concentration exceeds the saturation equilibrium condition, initiating the precipitation of the solid phase. Some relevant aspects of the unit operation layout need to be discussed first in order to accurately define the mathematical model of this unit. Notably, bigger crystals are separated from the finer ones through floatation. Then, the circulating pipe recycles and mixes the finer crystals with the feed. In fact, at the industrial scale,

heterogeneous crystallization (i.e., in the presence of pre-existing nuclei) is greatly preferred over homogeneous crystallization due to its much higher nucleation rate, corresponding to a higher productivity. Another significant aspect of the operation relates to mixing; sustained mixing speeds ensure homogeneous conditions throughout the unit volume and enhance the mass transfer rate [20]. Finally, the barometric condenser is included in the equipment to remove vaporized solvent and dissolved gases from the unit, which typically operates under vacuum to facilitate evaporation. No solid phase is lost from this outlet; solid crystals are instead withdrawn from the product discharge outlet.

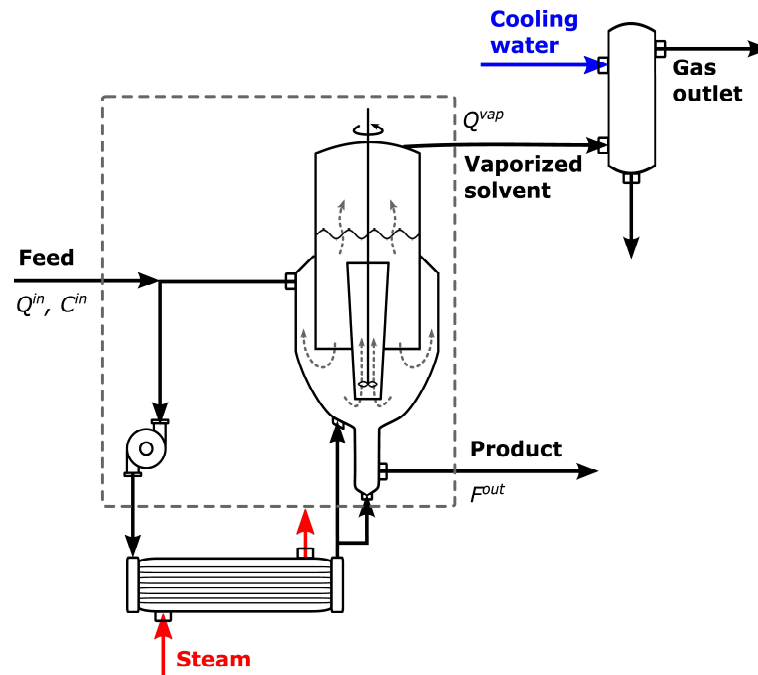


Figure 1. Schematic representation of evaporative crystallizer.

Based on the given process description, the dynamic mass balances in the liquid and solid phases for the solute species can be respectively written as follows:

$$\frac{dM_S^l}{dt} = Q^{in} \cdot C_S^{in} - Q^{out} \cdot C_S^{out} - r_c \cdot V \quad (1)$$

$$\frac{dM_S^s}{dt} = F_S^{in} - F_S^{out} + r_c \cdot V \quad (2)$$

where M_S is the mass of the solute in the liquid (l) and solid (s) phases, F_S is the crystal mass flowrate for the inlet (in) and the outlet (out) solid streams, C_S is the solute concentration in [kg/m^3] in the liquid streams, Q is the volumetric flowrate in [m^3/s], and r_c is the rate of crystallization per unit volume. Finally, V is the liquid volume inside the crystallization unit, and, assuming that the presence of the solute species does not affect the volume occupied by the solvent, it can be calculated with a mass balance on the liquid solvent phase:

$$\frac{d(\rho_L \cdot V)_S^l}{dt} = \rho_L^{in} \cdot Q^{in} - Q^{vap} \quad (3)$$

where Q^{vap} is the evaporation rate.

Based on these equations, several observations can be made to define each parameter and reshape the proposed model into a more concise formulation. Firstly, due to the presence of recycling, no outlet liquid stream is necessary, and no inlet crystals need to be considered except for the recirculated fines. Second, owing to the homogeneous conditions

within the unit, the properties of the outlet streams match those inside the equipment (e.g., $C_S^{out} = C_S$). Thirdly, as the outlet solid flowrate, indicative of the process productivity, includes all grown crystals removed from the bottom section of the unit, \dot{F}_S^{out} can be assumed to be equal to M_S per hour, except for the fraction α of seeds recycled, which corresponds to 10% of the steady-state mass value. Finally, the density of the liquid phase can be treated as constant, and the evaporation rate is a function of the heating rate (W) expressed by the equation:

$$Q^{vap} = \frac{W}{\rho_L \cdot \Delta H_L^{ev}} \quad (4)$$

where ΔH_L^{ev} is the solvent evaporation enthalpy and where the duty is controlled in order to keep the same evaporation rate as nominal steady-state conditions, i.e., $Q^{vap} = Q^{in,ss}$.

Regarding the interaction between the liquid and solid phases, mass transfer can be neglected due to the sustained mixing rate of the equipment [21], and the solute precipitation can be modeled as a second-order reaction, according to the example provided by Aghajanian and Koiranen (2020) [22]. The reaction rate can then be described by the equation:

$$r_c = k_r \cdot C_S^2 \quad (5)$$

where k_r is the reaction kinetic coefficient expressed in $[m^3/kg \cdot h]$.

The final model then becomes

$$\frac{dC_S}{dt} = \frac{Q^{in}}{V} \cdot (C_S^{in} - C_S) + \frac{Q^{in,ss}}{V} \cdot C_S - k_r \cdot C_S^2 \quad (6)$$

$$\frac{dM_S^s}{dt} = -\dot{M}_S + \alpha \cdot M_S^{ss} + k_r \cdot C_S^2 \cdot V \quad (7)$$

$$\frac{dV}{dt} = Q^{in} - Q^{in,ss} \quad (8)$$

The nominal operating parameters used for this study are listed in Table 1. When switching to the process dynamic model solution, they are used as the ODE system initial conditions.

Table 1. Operating parameters for crystallization.

| Parameter | Value | Unit |
|------------|-------|------------------------|
| α | 0.1 | 1/h |
| C_S^{in} | 3 | kg/m ³ |
| k_r | 0.1 | m ³ /(kg·h) |
| Q^{in} | 1 | m ³ /h |
| Q^{vap} | 1 | m ³ /h |
| V^{ss} | 3 | m ³ |

The main effectiveness indicator for a crystallization process is the crystal growth ratio, defined as the ratio between the mass of the final crystal product and that of the initial crystal seeds. For this study, it is given by

$$G = \frac{M_{seeds} + dM_S}{M_{seeds}} = \frac{M_S}{\alpha \cdot M_S^{ss}} \quad (9)$$

If the system satisfies the McCabe ΔL law, the linear growth of the crystal can be considered independent of the initial crystal size [23]. In this case study, as crystals are distributed along the unit based on their mass due to floatation, and the recycled ones are removed from the same crystallizer height, an almost uniform distribution can be assumed.

Therefore, the overall linear growth can be calculated as the growth of a single crystal using the equation:

$$\Delta L = \sqrt[3]{\frac{V_S}{V_0}} = \sqrt[3]{\frac{M_S}{M_0}} \quad (10)$$

The purpose of this study is to establish a correlation between the ΔL indicator and the variation in species concentration in the feed stream by means of a derivative-based surrogate model, which consists of a single differential equation to directly describe the response of crystal growth with respect to the input step disturbance. The choice of a relatively simple phenomenological model helps to evaluate the potential of the metamodelling procedure in terms of computational effort on a single unit before tackling larger and more complex systems. The results, discussed in the corresponding section, demonstrate significant improvements in computational time despite the relatively low complexity of the continuous crystallizer. However, before delving into the results, a comprehensive explanation of the response function-based surrogate modeling approach is provided in the following section.

3. Methodology

This Methodology Section addresses the two main aspects of the surrogate modeling study, namely, the Design of Experiment and the modeling method. In particular, the details regarding the selected sampling strategy and its application to the specific system under analysis are presented in the first subsection, while the fundamentals of the response function-based dynamic surrogate modeling approach are discussed in the second one.

3.1. Design of Experiment

In the DoE step, two main phases can be distinguished, namely, the domain definition and the sample generation. The former aims to select the variation range of the independent variables over which the accuracy of the final model is sufficiently high and no region with considerable deviation from the experimental data can be detected. In some studies, for instance, a repartition of the overall domain into multiple subdomains is performed to improve the accuracy of the surrogate model at the expense of the modeling procedure, which should be repeated as many times as the number of subdomains [24]. The latter, as mentioned in the introduction, aims to provide a uniform distribution of points to ensure homogeneous representation of all parts of the domain.

For this study, the feed properties, i.e., flowrate (Q_{in}) and concentration (C_{in}), were selected as independent variables. Specifically, to model the continuous crystallization dynamics, a perturbation domain within the $\pm 50\%$ range with respect to the nominal operating conditions was considered over a 10-h simulation. This choice allowed us to cover a 100% operating window and have a sufficiently heterogeneous profile without falling into singular cases. Thus, this approach was used to obtain both the DoE and the validation set profiles. The importance of the difference between the validation set and the DoE is based on the fact that, since the model is built to be as compliant as possible with respect to the DoE points, it will likely comply with similar trends, but only a robust model will minimize error even for very different values of the independent variables.

Regarding point generation, the sampling strategy selected for this research was Latin Hypercube Sampling (LHS) [25]. This near-random sample generation methodology is particularly suitable for once-through surrogate modeling since it provides uniform domain filling and ensures that two points are never aligned in the orthogonal direction with respect to the axes of independent parameters. For a better understanding, examples of LHS for bidimensional and tridimensional domains are shown in Figure 2 (a and b, respectively); it can be seen that, given the number of points n , the space is partitioned into an $n \times n$ chessboard, and data points are distributed on random locations within squares that do not belong to the same row or column of any other point.

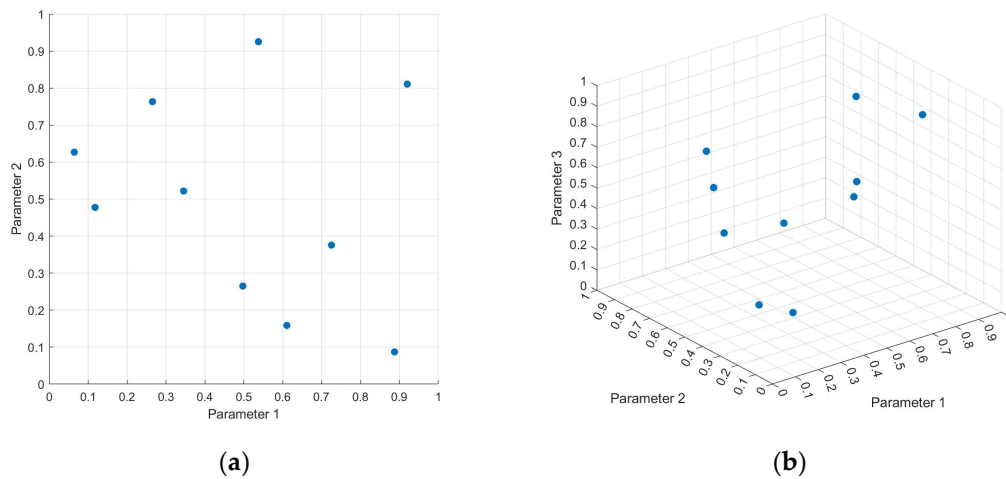


Figure 2. Latin Hypercube Sampling example: (a) 2D space; (b) 3D space.

The presented research employs LHS in two cases. The first one concerns the generation of $10 Q_{in} \times C_{in} \times t$ combinations for the system step perturbations, 5 for the DoE profile and 5 for the validation set. Then, LHS will be used to generate data samples for modeling from the growth vs. time trend whose size will range from 10 to 100 points. The independent variable profiles are thus obtained, and the outcomes of the related model simulation are reported in Section 4.1. From an operational point of view, the LHS was performed using the dedicated function *lhsdesign*, available in Matlab[®], so that the same software was used for the sampling and the modeling phases. Details concerning the latter are provided in the next section.

3.2. Response Function-Based Dynamic Surrogate Modeling

The surrogate modeling methodology proposed in this research is inspired by the derivative-based surrogate modeling approach proposed by Di Pretoro et al. (2022) [14] for an ethylene oxide process study and readapted to the continuous crystallization unit. The objective of this work is to use a single differential equation to correlate crystal growth and inlet parameters, expressed as deviation variables with respect to the nominal operating conditions. The model approach can be assumed to be analogous to that of the Response Surface Methodology [5]. However, in this case, the response surface is not modeled as an explicit function but as a differential response function, according to the well-established characteristic equations [26]. Although the system to be modeled is complex or could be composed of several sub-systems, in a series or parallel coupling layout, sometimes the correlation between the system outlet and the input variables can be simplified by low-order differential equations, without the need of a multiple sub-model combination. Examples of the parametrical differential form of a first- and second-order response function are reported in Equations (11) and (12), respectively:

$$\frac{dy}{dt} + \frac{1}{\tau} \cdot (y + A \cdot du + B) = 0 \quad (11)$$

$$\frac{1}{\omega^2} \cdot \frac{d^2y}{dt^2} + \frac{2 \cdot \zeta}{\omega} \cdot \frac{dy}{dt} + (A \cdot du + B) \cdot y + D \cdot du + E = 0 \quad (12)$$

The known terms used in these equations refer to the characteristic parameters of the process dynamics (τ , ω , ζ , etc.) and to the coefficients related to the specific system that is modeled (A , B , D , E , etc.), while u and y represent the independent and dependent variables, respectively. To select the most suitable characteristic differential equation, the approach can be either knowledge-based or automated. The former requires the analysis of the phenomenological model trend and its association with an already known low-order model by the decision maker, while the latter consists of iterative automated tests with

equations of different orders and the selection of the equation corresponding to the highest compliance. In the first case, once the most compliant function has been detected, the function coefficients are obtained by solving the Mean Absolute Error (MAE) minimization problem using an optimization algorithm in the form of

$$MSE = \min_{coeff} \frac{1}{n} \cdot \sum_{i=1}^n \left(\frac{y^{exp} - y^{mod}}{y^{exp}} \right)^2 \quad (13)$$

where y^{exp} is the value of the dependent variable corresponding to the DoE sampling, while y^{mod} is the respective value obtained by the dynamic surrogate model. In the second case, this procedure is replicated for the entire set of differential response functions.

For this study, the most suitable model was detected by the authors, and, as already mentioned, the error minimization was performed by using Matlab[®] built-in optimization functions. The same software was also used for profile graphics and data post-treatment. With regards to the latter, the performance indicators used to assess the accuracy and computational efficiency of the surrogate model are the machine computational time and the Normalized Mean Square Error (NMSE) over the 10-h simulation, respectively. The NMSE can be defined as

$$NMSE = 1 - \frac{\|y^{exp} - y^{mod}\|^2}{\|y^{exp} - \tilde{y}^{exp}\|^2} \quad (14)$$

where \tilde{y}^{exp} is the average value of one experimental point in the case of multiple sets. In the case of an automated procedure, this indicator could be used for the a posteriori selection of the most suitable differential response function among those that have been tested.

The results concerning the sampling and the modeling step are presented and discussed in detail in the following section.

4. Results and Discussion

This section includes both the DoE and the modeling phase results. In the first subsection, the deviations of the independent parameters and the related response profiles are presented for both the DoE and the validation set. In the second subsection, the comparison between the surrogate model and the crystallizer growth indicator profile is discussed, and performance indicators for the proposed modeling methodology are calculated.

4.1. Dynamic Trend Profiles

As a first step, the external perturbation profiles were generated according to the LHS method to obtain the DoE dataset and the validation profile to be compared to the modeling phase outcome, as reported in Figures 3 and 4, respectively.

The step deviations for inlet flowrate (red dashed line) and concentration (blue dotted line) correspond to a linear growth deviation (blue line) with respect to the steady-state value under nominal operating conditions. Thus, based on the trend in Figure 3b, data samples ranging in size from 5 to 100 were generated to perform a sensitivity analysis of the modeling procedure accuracy.

However, the obtained results not only serve the purpose of providing points for the sample generation and the final model validation but, in this case, they are also exploited to decide which response function order can best model the continuous crystallizer unit. As can be observed in Figure 3b, the solute concentration in the unit exhibits a first-order dynamic behavior, consistent with the model related to Equation (6). Conversely, the dynamic trend of crystal linear growth is much more difficult to deduce using the model equations since it does not directly appear in any of them; it is the result of further operations based on the phenomenological model solution. A second-order response function was then tested as a possible surrogate model for the linear growth dynamics correlation with

respect to the process input parameters. The resulting outcome is reported and analyzed in the following section.

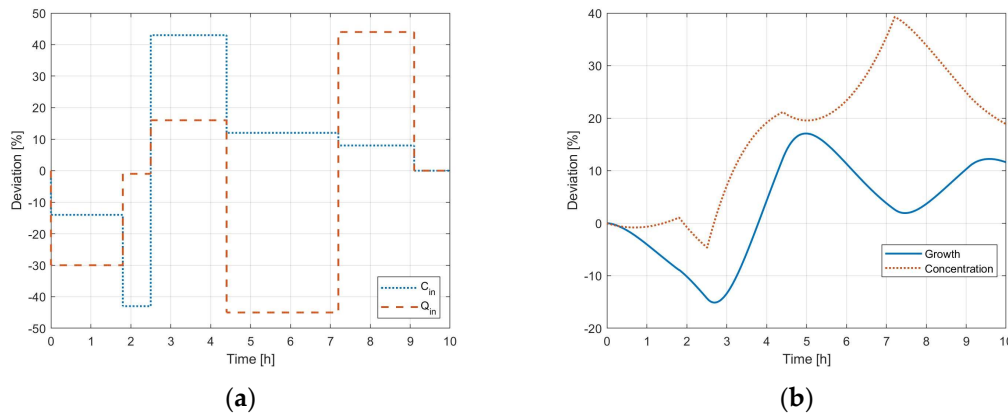


Figure 3. Dynamic profiles for the DoE: (a) independent input variables; (b) crystal growth.

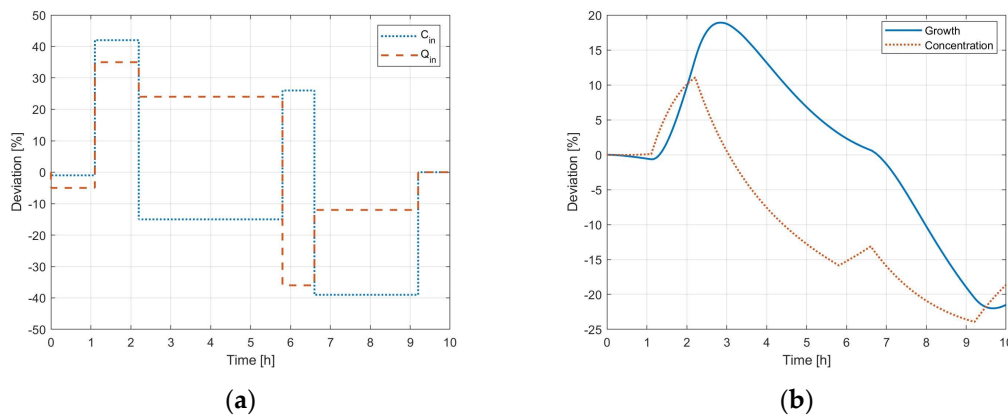


Figure 4. Dynamic profiles for the model validation: (a) independent input variables; (b) crystal growth.

4.2. Dynamic Surrogate Model

The modeling phase, based on the methodology previously proposed by the authors [14], was performed first. Before analyzing the results, it is worth noting that all variables presented in this section are expressed as deviation variables with respect to the nominal steady state, defined as

$$\hat{G} = \frac{G - G^{ss}}{G^{ss}} \tag{15}$$

Therefore, the initial conditions are always set to [0,0], and all dependent variables are expressed as a percentage of the steady-state ones.

Figure 5 shows the results for a first attempt at crystallization modeling based on a global second-order response over the entire domain. In particular, the graph refers to the result options that best fit the first step response, when accounting for the concentration perturbation only. As can be seen, despite the similar trend in terms of positive/negative slope and proportional gain, the error with respect to the other perturbations increases along with the simulation’s elapsed time. Several attempts to improve the quality of the overall profile have been carried out both for coupled and standalone disturbances, but whenever a more compliant dynamic trend is obtained over a given region, the mismatch increases for other scheduling windows. This inconvenience is due to the fact that the characteristic parameters of the model, such as the characteristic frequency $\frac{2\pi}{\omega}$ and damping factor ξ , are not constant during the operation because of the variation in the system characteristics. In particular, the variation in volume during the operation, due to the inlet flowrate perturbation, affects the unit residence time and the system inertia. It is important

to note that, in the previous study where this methodology worked for the entire window scheduling [14], the model correlation was carried out for the product flowrate as a function of the reactant one, and the system in between was a closed-loop system with fixed design and operational parameters.

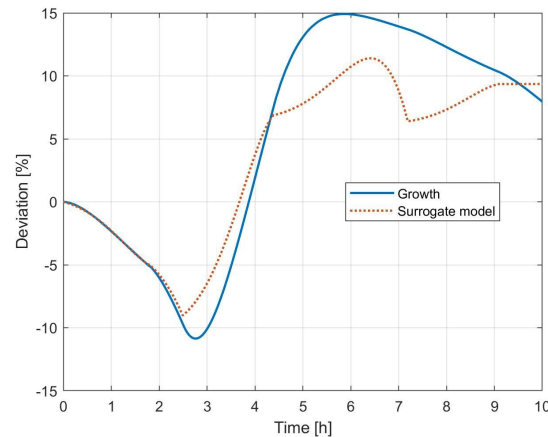


Figure 5. Dynamic surrogate model results with a single set of characteristic parameters.

Therefore, to deal with this issue, a domain partition was performed and the modeling procedure was repeated with a subdomain-wise approach, as suggested in the literature [27]. The obtained results are reported in the following section along with the analysis of the model performance indicators.

4.3. The Domain Partition-Based Modeling Approach

In the light of the previous results, the first step of the corrected procedure was to partition the domain according to the discretization of step disturbances. Apart from this modification, the following modeling phase was carried out exactly as described in the Methodology Section for each subdomain independently, regardless of the time interval size.

Therefore, Figure 6 shows the results for the domain-wise second-order differential equation model with a 2000-point sample. In this case, the global model is characterized by as many $[\omega, \zeta]$ binaries as its subdomains, i.e., 6. Compared to the previous approach, the new model proves to be very accurate (NMSE equal to 99.9358% and 99.9931% for mass and linear growth, respectively) and meets the initial expectations in terms of compliance with the DoE trend. Predictably, the main criticalities can be observed at the subdomain boundaries where the error is slightly higher, and non-differentiable singular points can be noticed, though without causing any relevant issue to the overall trend. Furthermore, when applying the cubic root to shift from mass to linear growth, the errors are further mitigated, except for points close to the initial condition.

Regarding computational performance, the calculations were performed using a laptop running Matlab[®] version 2023b on the Windows 11 operating system, equipped with a 13th-Gen. i9 2.60 GHz processor and 64 GB of RAM. The resulting trends of computational time and model accuracy as a function of the sample size are reported in Figures 7a and 7b, respectively.

The value of the Normalized Mean Square Error, calculated as explained in Equation (14), is reported in Figure 7a for both the phenomenological and surrogate models. Generally, the NMSE trend increases with the increasing number of points, as expected due to the accuracy improvement. However, it can be observed that compliance with the benchmark results is achieved much faster by the rigorous model, with an almost negligible mismatch for samples with more than 125 points. Conversely, the surrogate model follows a hyperbolic trend with higher eccentricity, and the error increases faster when approaching small samples. Nevertheless, the overall mismatch with respect to the phenomenological model

becomes lower than 5% for 250 points and falls below 1.5% for samples with more than 500 points.

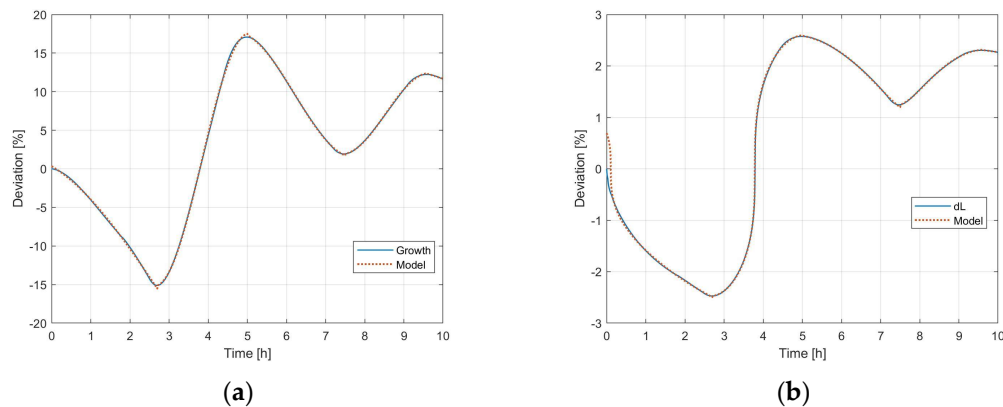


Figure 6. Comparison between phenomenological and domain partition-wise surrogate model with respect to the DoE dataset: (a) crystal growth ratio; (b) crystal linear growth.

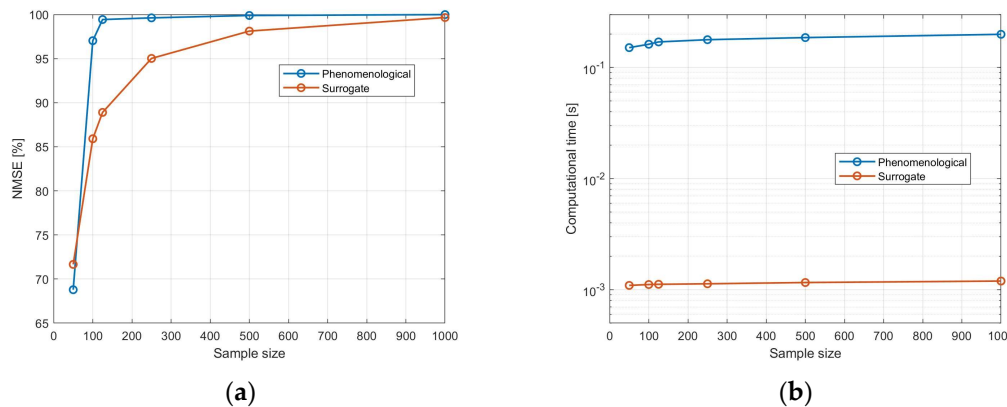


Figure 7. Computational performance indicators: (a) (1-NMSE) vs. sample size; (b) computational time vs. sample size.

Computational time exhibits a linear trend for both the phenomenological and surrogate models as the number of points to be calculated increases. However, a remarkable difference can be clearly noticed when comparing the two: the solution of the entire model with respect to a single second-order differential equation requires a computational time that is more than two orders of magnitude higher, despite the low complexity of the crystallizer unit. Based on these results and considering the low slope of the computational time behavior with respect to the sample size, it can be concluded that a sample size of 500 points is a good compromise between model accuracy and computational effort; if a higher accuracy is required and a small increase in model solution time can be tolerated, the 1000-point sample is definitely the best option, corresponding to a totally feasible computational performance.

Finally, the same procedure was repeated on the validation dataset using the optimal number of 1000 points per subdomain sample. The results are presented in Figure 8. From a qualitative point of view, the same remark as for the DoE-based trends remains valid. In particular, the NMSE for the mass and linear growth functions is equal to 99.9682% and 99.9964%, respectively. Cuspid points can still be detected in correspondence with step perturbation changes, but even in this case they do not considerably affect the overall accuracy except in proximity of the linear growth trend initial conditions. Therefore, based on the obtained outcome, the surrogate modeling approach exploiting domain partition for the description of continuous crystallizer dynamics can be validated.

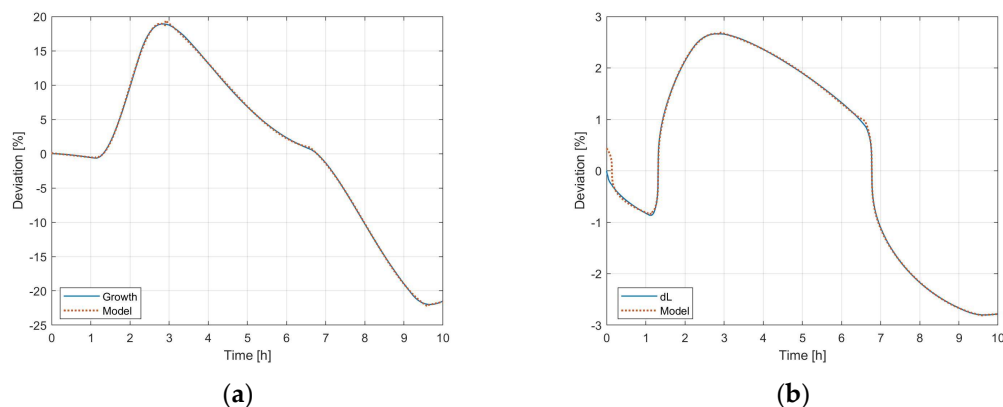


Figure 8. Comparison between phenomenological and domain partition-wise surrogate model with respect to the validation dataset: (a) crystal growth ratio; (b) crystal linear growth.

5. Conclusions

In the presented research, a surrogate model was developed to describe the dynamics of a continuous crystallizer unit by tackling some critical issues that were encountered. The analysis of these criticalities led to important methodological considerations, which can be summarized as follows:

- When the system properties undergo significant variations during operation scheduling, a single model cannot fully describe the process behavior throughout the entire simulation window. In this case, multiple modeling phases over a domain partition should be conducted;
- Conversely, for the simulation of a single transient, employing a response function-based dynamic surrogate model yielded highly accurate results in terms of independent variable trends and in the analysis of the indirect outcome parameters;
- Additionally, for a single modeling region, the sample size corresponding to the best compromise between computational effort and accuracy can be identified;
- The use of surrogate models for dynamic simulation substantially enhances computational efficiency while maintaining a high level of accuracy. However, full automation is currently unattainable, as a preliminary analysis of the system physical behavior is necessary to select the most suitable modeling methodology.

Therefore, considering the significant outcomes discussed above, further testing and development of the presented procedure are worthwhile. The direct correlation between the input and output variables of interest through a single differential equation, capable of replacing the entire model with good accuracy, exhibits considerable potential in terms of computational time reduction. Therefore, this approach could enable applications such as Real-Time Dynamic Optimization and model predictive control to be used for a broader range of process systems, which currently demand extensive calculation times. Nevertheless, the proposed procedure requires substantial improvements before scaling up to more complex systems. The unconditional use of domain partition might indeed present a notable drawback in terms of computational performance for process systems with highly irregular behavior or wide operating windows. In such cases, the number of required partitions may significantly increase, necessitating adjustments to characteristic parameters for each iteration. Moreover, when dealing with more complex systems, the correlation between input and output variables overlooks the local feasibility limitations of individual units, potentially yielding operating profiles that cannot be attained in practice.

In conclusion, building upon the initial step carried out in this study, further research will focus on developing a comprehensive framework for applying the response function-based dynamic surrogate modeling methodology to process systems. This endeavor will address the aforementioned challenges and explore potential solutions to feasibility issues through a rigorous approach.

Author Contributions: Conceptualization, A.D.P.; methodology, A.D.P. and S.N.; validation, L.M.; data curation, L.M.; writing—original draft preparation, A.D.P.; writing—review and editing, L.M. and S.N.; supervision, L.M. and S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Nomenclature

| Symbol | Definition | Unit |
|-----------------|-------------------------------------|--------------------------------|
| C_S | Solute concentration | kg/m ³ |
| DoE | Design of Experiment | acronym |
| F | Mass rate | kg/h |
| G | Growth ratio | kg/kg |
| ΔH^{ev} | Latent heat of evaporation | J/kg |
| in | Inlet stream, superscript | superscript |
| k_r | Kinetic coefficient | m ³ /(kg·h) |
| l | Liquid phase, superscript | superscript |
| LHS | Latin Hypercube Sampling | acronym |
| ΔL | Linear growth | mm |
| M | Mass | kg |
| MINLP | Mixed-Integer Nonlinear Programming | acronym |
| n | Sample size | / |
| NMSE | Normalized Mean Square Error | acronym |
| out | Outlet stream, superscript | superscript |
| Q | Volume rate | m ³ /s |
| r_c | Crystallization reaction rate | kg/(m ³ ·h) |
| RSM | Response Surface Methodology | acronym |
| RTDO | Real-Time Dynamic Optimization | acronym |
| s | Solid phase, superscript | superscript |
| ss | Steady-state, superscript | superscript |
| t | Time | h |
| V | Volume | m ³ |
| W | Heating rate | W |
| α | Crystal product withdrawal fraction | kg/kg |
| β | Solvent vaporization fraction | m ³ /m ³ |
| ρ | Density | kg/m ³ |

References

- Kim, S.H.; Boukouvala, F. Surrogate-based optimization for mixed-integer nonlinear problems. *Comput. Chem. Eng.* **2020**, *140*, 106847. [\[CrossRef\]](#)
- Bhosekar, A.; Ierapetritou, M. Advances in surrogate-based modeling, feasibility analysis, and optimization: A review. *Comput. Chem. Eng.* **2018**, *108*, 250–267. [\[CrossRef\]](#)
- McBride, K.; Sundmacher, K. Overview of Surrogate Modeling in Chemical Process Engineering. *Chem. Ing. Tech.* **2019**, *91*, 228–239. [\[CrossRef\]](#)
- Davis, S.E.; Cremaschi, S.; Eden, M.R. Efficient surrogate model development: Impact of sample size and underlying model dimensions. *Comput. Aided Chem. Eng.* **2018**, *44*, 979–984. [\[CrossRef\]](#)
- Box, G.E.P.; Wilson, K.B. On the experimental attainment of optimum conditions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1951**, *13*, 1–45. [\[CrossRef\]](#)
- Fang, H.; Horstemeyer, M.F. Global response approximation with radial basis functions. *Eng. Optim.* **2006**, *38*, 407–424. [\[CrossRef\]](#)
- Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [\[CrossRef\]](#)
- Krige, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. S. Afr. Inst. Min. Metal.* **1951**, *52*, 119–139.
- Konakli, K.; Sudret, B. Polynomial meta-models with canonical low-rank approximations: Numerical insights and comparison to sparse polynomial chaos expansions. *J. Comput. Phys.* **2016**, *321*, 1144–1169. [\[CrossRef\]](#)

10. Breiman, L. Random Forests. *Mach. Learn.* **2021**, *45*, 5–32. [[CrossRef](#)]
11. Mamo, T.Z.; Di Pretoro, A.; Chiari, V.; Montastruc, L.; Negny, S. Benefits of feasibility constrained sampling on unit operations surrogate model accuracy. *Comput. Chem. Eng.* **2023**, *173*, 108210. [[CrossRef](#)]
12. Shokry, A.; Baraldi, P.; Zio, E.; Espuña, A. Dynamic Surrogate Modeling for Multistep-ahead Prediction of Multivariate Nonlinear Chemical Processes. *Ind. Eng. Chem. Res.* **2020**, *59*, 15634–15655. [[CrossRef](#)]
13. Di Pretoro, A.; Tomaselli, A.; Manenti, F.; Montastruc, L. Dynamic Surrogate Modeling for Continuous Processes Control Applications. *Comput. Aided Chem. Eng.* **2022**, *51*, 91–96. [[CrossRef](#)]
14. Zhang, X.; Yang, B.; Yu, T.; Jiang, L. Dynamic Surrogate Model Based Optimization for MPPT of Centralized Thermoelectric Generation Systems Under Heterogeneous Temperature Difference. *IEEE Trans. Energy Convers.* **2020**, *35*, 966–976. [[CrossRef](#)]
15. Di Pretoro, A.; Bruns, B.; Negny, S.; Grünewald, M.; Riese, J. Demand Response Scheduling Using Derivative-Based Dynamic Surrogate Models. *Comput. Chem. Eng.* **2022**, *160*, 107711. [[CrossRef](#)]
16. Chen, J.; Sarma, B.; Evans, J.M.B.; Myerson, A.S. Pharmaceutical Crystallization. *Cryst. Growth Des.* **2011**, *11*, 887–895. [[CrossRef](#)]
17. Hartel, R.W. Crystallization in Foods. In *Handbook of Industrial Crystallization*; Myerson, A.S., Erdemir, D., Lee, A.Y., Eds.; Cambridge University Press: Cambridge, UK, 2019; pp. 460–478.
18. Orehek, J.; Teslić, D.; Likožar, B. Continuous Crystallization Processes in Pharmaceutical Manufacturing: A Review. *Org. Process Res. Dev.* **2021**, *25*, 16–42. [[CrossRef](#)]
19. Ali, M.I. Struvite Crystallization in Fed-Batch Pilot Scale and Description of Solution Chemistry of Struvite. *Chem. Eng. Res. Des.* **2007**, *85*, 344–356. [[CrossRef](#)]
20. Zauner, R.; Jones, A.G. On the influence of mixing on crystal precipitation processes—Application of the segregated feed model. *Chem. Eng. Sci.* **2002**, *57*, 821–831. [[CrossRef](#)]
21. Qu, Y.; Cheng, J.; Mao, Z.S.; Yang, C. A perspective review on mixing effect for modeling and simulation of reactive and antisolvent crystallization processes. *React. Chem. Eng.* **2021**, *6*, 183–196. [[CrossRef](#)]
22. Aghajanian, S.; Koiranen, T. Dynamic modeling and semibatch reactive crystallization of calcium carbonate through CO₂ capture in highly alkaline water. *J. CO₂ Util.* **2020**, *38*, 366–374. [[CrossRef](#)]
23. McCabe, W.L. Crystal Growth in Aqueous Solutions¹: II—Experimental. *Ind. Eng. Chem.* **1929**, *21*, 112–119. [[CrossRef](#)]
24. Srinivas, S.V.; Karimi, I.A. Zone-wise surrogate modelling (ZSM) of univariate systems. *Comput. Chem. Eng.* **2023**, *174*, 108249. [[CrossRef](#)]
25. McKay, M.D.; Beckman, R.J.; Conover, W.J. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* **1979**, *21*, 239–245. [[CrossRef](#)]
26. Golnaraghi, F.; Kuo, B.C. *Automatic Control Systems*, 10th ed.; McGraw-Hill Education: New York, NY, USA, 2017.
27. Tang, A.C.; Fang, Y.; Pfister, P.-D. A surrogate model assisted with a subdomain model for surface-mounted permanent-magnet machine. In Proceedings of the 2021 IEEE International Magnetic Conference (INTERMAG), Lyon, France, 26–30 April 2021; pp. 1–5. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.