A Data-Driven Bayesian Approach for Optimal Dynamic Product Transitions

Antonio Flores-Tlacuahuac*, Luis Fabián Fuentes-Cortés†

* Institute of Advanced Materials for Sustainable Manufacturing

Ave. Eugenio Garza Sada 2501, Monterrey, N.L., 64849, Mexico

Tecnologico de Monterrey

† Departamento de Ingeniería Química

Tecnologico Nacional de México - Instituto Tecnológico de Celaya

Celaya, Guanajuato 38010, Mexico

November 25, 2023

Abstract

In the processing industry, dynamic product transitions are essential for achieving high product quality, minimizing the use of raw materials and energy and reducing production costs. However, optimizing dynamic product transitions is a challenging task due to the complex dynamics of the process and the uncertainty in the measurements. In this work, a data-driven Bayesian approach for optimal dynamic product transitions is proposed. The proposed approach is based on a dynamic optimization problem that is solved using a Bayesian optimization algorithm. One of the advantages of this approach for process optimization tasks is that it does not require a first-principles dynamic mathematical model for drawing optimal solutions. The approach is applied to three case studies, and the results are comparable in performance quality with those obtained using a traditional gradient-based optimization approach. The results show that the proposed approach is able to find optimal transition trajectories that meet the product composition requirements using smooth control actions. The approach is also able to cope with noisy measurements, which is an important feature in real-world applications. The proposed approach has several advantages over traditional optimization approaches, including being data-driven, able to cope with noisy measurements, computationally efficient, and it requires modest computational effort. Complex on-line optimal control problems can benefit from adopting a data-driven Bayesian optimization scheme.

Keywords:: Bayesian optimization, non-linear systems, noisy measurements, dynamic optimization, data-driven systems.

*Corresponding Author: antonio.flores.t@tec.mx

1 Introduction

Dynamic product transitions are an ubiquitous operation within the processing industry.[1–3] The primary aim is to transition a given processing system from its initial steady-state to a predetermined final target steady-state.[4–12] This transition encompasses the dynamic alteration of the manipulated variables. Furthermore, the dynamic operational transition must be executed with the optimization of specific performance metrics in mind. This type of operations are also known under the name of optimal open-loop dynamic operation policies, since they are off-line calculated and can be later implemented by a control system. Of course, dynamic optimal product transition operations can also be achieved by an on-line real time optimization system, although this can involve large computational demands, noisy measurements and lack of model fidelity.

Implicit in the preceding remarks is the necessity of employing a first-principles dynamic mathematical model for optimization calculations through traditional gradient-based optimization algorithms.[13–15] This is owing to the essential need for derivative information by these algorithms. However, meeting the prerequisite of having a dependable first-principles dynamic model can be a challenging endeavor. In reality, there are instances where these models can be arduous to develop, as the fundamental physical and chemical mechanisms that underlie them may not be well understood. The process of formulating and validating such models can be time-consuming. Additionally, the model parameters might be inadequately known or subject to time-dependent variations. The difficulty in finding an adequate optimal solution, from a strictly engineering point of view, to this type of problem lies in the fact that its formulation involves dealing with discontinuities, non-differentiable expressions, excessively bounded feasible regions, or the existence of a multiplicity of local solutions.[16]

Black-box optimization represents an approach to system optimization that finds its ideal application in cases where a first-order model is unavailable.[17–20] In this method, optimization relies upon an empirical model or an approximation of the target system. Subsequently, this approximated or surrogate model can be subject to optimization. Generic or heuristic algorithms have been extensively employed for black-box optimization. However, one of their primary shortcomings is the absence of a solid theoretical foundation and a guarantee of attaining a true optimal solution. These types of approaches have been widely used in the solution of nonlinear or mixed integer nonlinear problems, trying to obtain valid optimal solutions and, in some cases, aspiring to the global solutions to complex problems. However, even the most optimistic approaches still lack a mathematical guarantee of optimality based on classical criteria, and in some cases, the reproducibility of the computational experiments performed is still in question.[16,21–24] One way to enhance these approaches has been the use of machine learning algorithms. However, this coupling can result in additional computational costs that end up being counterproductive. In attempts to address this, there have been endeavors to construct black-box models through the utilization of quadratic model representations, followed by the application of classical unconstrained optimization techniques. More recently, there has been a burgeoning interest in adopting a data-driven approach for the creation of reliable surrogate models, subsequently employing formal Bayesian statistics for system optimization.[25–28] Accordingly, data science,[29] materials design,[30] hyper-parameters tuning in modern machine learning strategies[31] and the optimal design of experiments[32] have been among the main fields where Bayesian optimization (BO) has been applied. Specifically, applications of BO in the PET chemical recycling process,[33] computational fluid dynamics,[34] tuning of process controllers,[35] heating, ventilation, and air conditioning (HVAC) plants,[36] chemical reactor design using computational fluid dynamics,[37] extending BO tools to deploy parallel computer architectures,[38] among other applications, have been recently published.

The majority of documented applications of BO primarily focus on the optimization of steady-state systems. Nevertheless, when addressing dynamic optimization environments, distinct challenges may arise that are not encountered in the optimization of steady-state systems. These challenges encompass issues like managing transitions between unstable operational points and the necessity of achieving minimal time transitions without resorting to aggressive control actions. This work aims to illustrate how dynamic optimal product transitions can be achieved using solely measurements of the manipulated and controlled variables. This approach signifies a black-box data-driven strategy for addressing the optimization of uncertain dynamic systems. In this work, we demonstrate the implementation of Bayesian optimal dynamic transitions through the examination of three practical industrial instances. These instances comprise two highly nonlinear behavior continuous stirred tank reactors and an ideal binary distillation column. Each example showcases intricate steady-state and dynamic behaviours, and serve to illustrate potential issues when calculating optimal transition trajectories by Bayesian techniques. Moreover, they serve as a foundation for extending these methodologies to address other intricate processing systems, including reactive distillation, pressure-swing adsorption, and full-scale processing plants.

2 Bayesian optimization of dynamic systems

About the Bayesian optimization approach

Bayesian optimization is especially suited for the optimization of complex and expensive-to-evaluate objective functions. In fact, BO is particularly useful when dealing with objective functions which are costly to compute and noisy measuring environments. Because Bayesian optimization is a type of iterative probabilistic optimization technique that recently came into use,[28,39,40] especially when compared to well established unconstrained and constrained optimization techniques, we will highlight the following points regarding advantages/disadvantages of this technique as follows. In Figure 1 a flowsheet of the steps taken during BO is shown.

- Initially, a selection of input and output measured variables must be made. The output variables pertain to the target aspects, serving as control variables, while the input variables, considered manipulated variables, actively steer the system dynamic behavior. Subsequently, an objective function should be formulated, aligning with the intended goals, such as meeting purity standards or minimizing the consumption of raw materials and energy. The objective function hinges upon the set of controlled and manipulated variables. Hence, the primary goal of the objective function lies in furnishing a quantifiable metric to track advancements toward achieving optimal conditions. While this stage theoretically allows for the utilization of extensive datasets, such vast data repositories are not strictly mandatory. In fact, one of the advantages of Bayesian optimization techniques lies in their capability to start iterations with a minimal set of measurements. This attribute proves especially beneficial when confronted with extensive and costly experimental setups.

- In BO the target objective function is commonly modelled in terms of a probabilistic surrogate model, like probabilistic or Bayesian neural networks, or using a Gaussian Process (GP) which is an extension of traditional finite dimension multivariable normal distributions to non-parametric infinite dimension mean and kernel functions. This extension allows to model random processes with improved prediction capabilities and it is theoretically supported by the Kolmogorov extension theorem.[28] In any case the aim of the surrogate approximation is to provide estimates of the objective function and to draw uncertainty levels associated with such estimates.

- Using estimates of the objective function, provided by the surrogate model, BO makes use of an *activation function* to draw new values of the decision variables to guide the search toward optimal process conditions. In this phase during the search for optimal conditions, BO iteratively balances two aspects: (a) exploration (searching in regions with high uncertainty) and (b) exploitation (focusing on regions with high expected objective values). The aim of the activation function is to decide where and how much to sample the objective function (i.e. to draw the values of the decision variables) such to gather the most valuable information.

- During each iteration, BO extracts values for the decision variables and the objective function. With this data in hand, the process can either continue seeking enhancement in the objective function or conclude the iterative procedure. Should additional iterations be necessary, the current measurements augment the previous set. The underlying concept involves employing posterior Bayesian inference to generate improved values

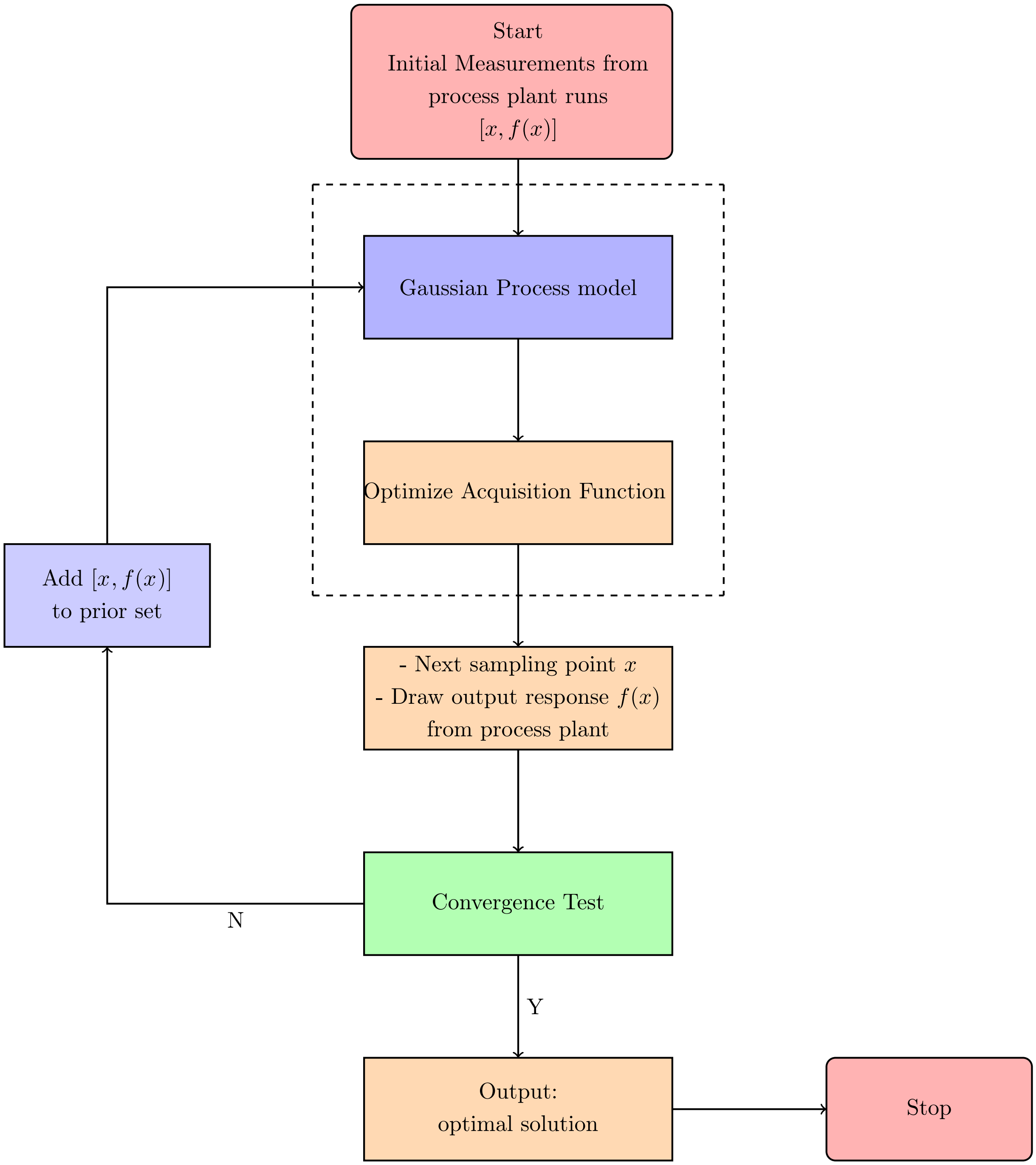


Figure 1: Flowsheet of the data-driven Bayesian Optimization approach. Starting from an initial set of process measurements a probabilistic surrogate model is built. Then, trough the optimization of an *acquisition function*, the values of the decision variables are updated. The aim of the acquisition function is to guide toward the searching of optimal conditions improving, at each iteration step, the probability of drawing the best optimal conditions. After some iterations an optimal solution can be found or the number of iterations can be exhausted. When the optimal solution has not been found, the information of the present iteration (i.e. objective function and decision variables) is accumulated and use it to draw a new surrogate model approximation and to infer new values of the decision variables. The dashed box comprises the two major steps of the Bayesian optimization strategy.

for the decision variables. As iterations progress and more information becomes accessible, the Gaussian Process approach typically yields improved predictive capabilities for the objective function.

BO is a well rooted theoretically sound optimization method which makes use of probabilistic models to guide search towards optimal and feasible optimality regions. BO has been widely used in Machine Learning tasks for hyper-parameters tuning[31] and to improve the performance of physical experiments.[32,39] In these scenarios it can be difficult or computationally demanding to deploy traditional or classical optimization techniques.

About the optimization of dynamic systems

When a first principles dynamic mathematical model of the system to be optimized is available, then gradient-based optimization techniques have been widely deployed for this purpose.[3,4] However, gradient-based dynamic optimization problems can be hard to use, especially when it comes to consider large scale, highly nonlinear behaviour, uncertainties and noisy measurements. Therefore, for this type of problems, it can be useful to deploy an efficient and reliable gradient-free optimization technique which does not require a first principles mathematical description to draw dynamic optimal transition trajectories. This type of free-gradient optimization techniques are known as black-box optimization techniques and have been deployed for tackling the problem addressed in the present work. Genetic and metaheuristic optimization techniques have been popular optimization techniques to deal with black-box optimization tasks.[41,42] However, these algorithms have not a solid theoretical foundation making difficult to assess the quality and validity of the solution found by these techniques. To partially cope with this issue some other black-box optimization works solve the dynamic optimization problem by building an approximate surrogate (i.e. quadractic function) around the intended dynamic trajectory, turning the problem into an unconstrained optimization problem that can be solved by traditional optimization techniques.[17]

Nowadays, particularly within the paradigms of Industry 4.0 and artificial intelligence, there exists a compelling imperative to address the efficient and reliable solution of optimization challenges through a predominantly data-driven methodology.[43] Within this framework, the fundamental concept revolves around the utilization of solely the measurements of process variables to both construct and address black-box optimization problems. This endeavor draws upon theoretically sound mathematical foundations, notably employing Bayesian statistical tools.

Relationship between Bayesian statistics and Bayesian optimization

Since BO techniques are rooted in Bayesian statistics inference concepts,[44–46] it is important to highlight the connections between these two disciplines.

- **Probabilistic Modeling**: Both Bayesian statistics and Bayesian optimization rely on probabilistic modeling. In Bayesian statistics, probability distributions are used to describe uncertainty in parameters or data. In Bayesian optimization, a Gaussian process as a probabilistic model is used to approximate the objective function and capture uncertainty in the optimization process.

- **Bayesian Inference**: Bayesian statistics is concerned with updating beliefs about parameters or data as new evidence is obtained. Bayesian optimization, on the other hand, involves iteratively updating the probabilistic surrogate model of the objective function with new evaluations. The Bayesian inference framework is used in both cases to make informed decisions.

- **Acquisition Functions**: In Bayesian optimization, an acquisition function is used to determine the next point to evaluate. These acquisition functions balance exploration (sampling in regions of high uncertainty) and exploitation (sampling in regions with high expected improvement). Bayesian statistics plays a role in estimating uncertainty, which is a crucial part of the acquisition function.

- **Sequential Decision-Making**: Both Bayesian statistics and Bayesian optimization involve making sequential decisions. In Bayesian statistics, beliefs are updated after each observation, and in Bayesian optimization, a decision is taken about where to evaluate the objective function next, continually refining the progress towards the optimal solution.

In summary, Bayesian optimization is a specific application of Bayesian statistics, where the focus is on finding the optimal solution for an expensive or unknown objective function through a sequence of data-driven decisions. Bayesian statistics provides the underlying framework and tools for modeling uncertainty, updating beliefs, and making informed decisions in the context of Bayesian optimization.

3 Problem definition

The issue to be elucidated within this study may be defined as follows. In the context of a data-driven processing system working in an initial steady-state operating point, the challenge at hand is to draw the control actions that facilitate the system transition from the aforementioned initial state to a designated target steady-state using only measurements of the controlled and manipulated variables. These control actions are to be drawn with the intent of minimizing or maximizing a predetermined objective function. Because these dynamic optimal transitions between the initial and final steady-states necessitate the implementation of time-dependent control actions, this turns the addressed optimization task into a black-box open-loop dynamic optimization problem.

4 Case studies

In this section, we demonstrate the methodology for delineating optimal dynamic transition trajectories through a data-driven Bayesian Optimization (BO) approach. The BO computations were executed utilizing an open-source Python implementation of the BO strategy.[47] Across all scenarios, a dynamic model representing the system under consideration was employed to emulate real-system behavior and generate samples utilized within the optimization methodology.

Hicks reactor

The reaction system, as proposed by Hicks and Ray,[48] has been extensively employed as a benchmark problem to assess algorithms designed for addressing non-linear multiplicity patterns and optimal product transition control policies.[6] In this study, we utilize this system to formulate Bayesian optimization data-driven optimal control policies. The dynamic mathematical model reads as follows.

$$\frac{dy_1}{dt} = \frac{1-y_1}{\theta} - k_{10}e^{-N/y_2}y_1 \tag{1}$$

$$\frac{dy_2}{dt} = \frac{y_f-y_2}{\theta} + k_{10}e^{-N/y_2}y_1 - \alpha u(y_2-y_c) \tag{2}$$

where y_1 stands for concentration, y_2 is the temperature, y_f is the feed stream temperature and u is the cooling flowrate. Additional parameter values are as follows: $\theta = 20$, $k_{10} = 300$, $N = 5$, $\alpha = 1.95\text{x}10^{-4}$, $t_f = 300$, $J = 100$, $c_f = 7.6$, $t_c = 290$. Moreover, $y_c = t_c/(J \cdot c_f)$ and $y_f = t_f/(J \cdot c_f)$.

For this case study the objective function (Ω) was defined as follows:

$$\Omega = \lambda_1(y_2 - y_2^t)^2 + \lambda_2\left(\frac{u-u^t}{u^t}\right)^2 \tag{3}$$

where the super-script t means target values, λ_1 and λ_2 are weighting function whose role is to stress the importance of the terms of the objective function such as to draw smooth control actions. As it can seen from the above equation,

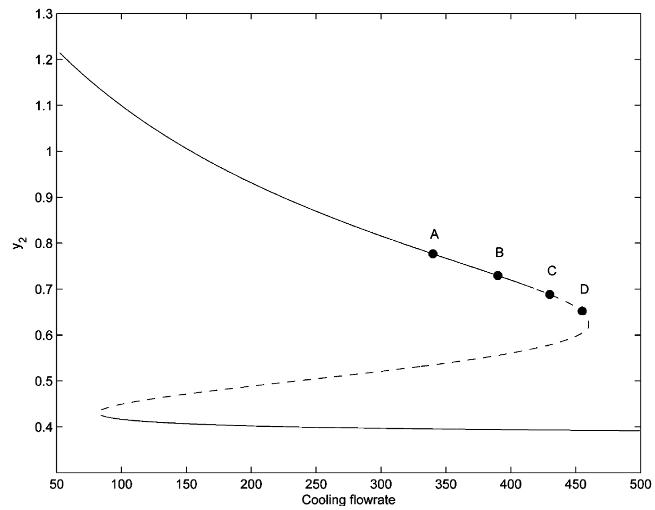
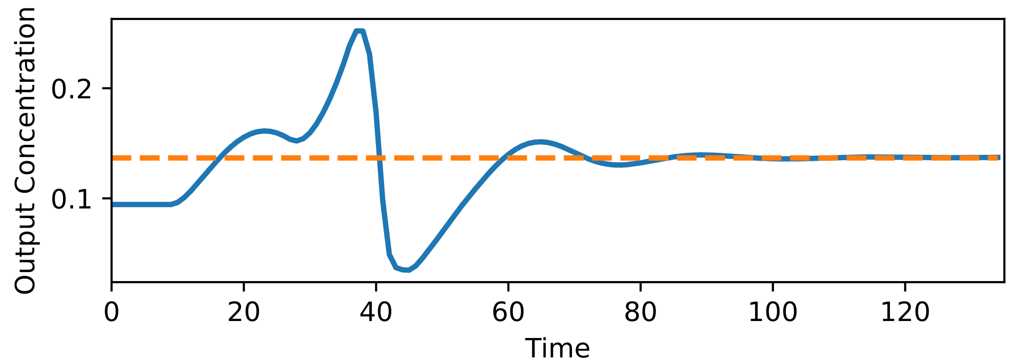


Figure 2: Nonlinear steady-state multiplicity map of the Hicks reactor. The continuous line stands for stable states, while the dashed line represent unstable steady-states.

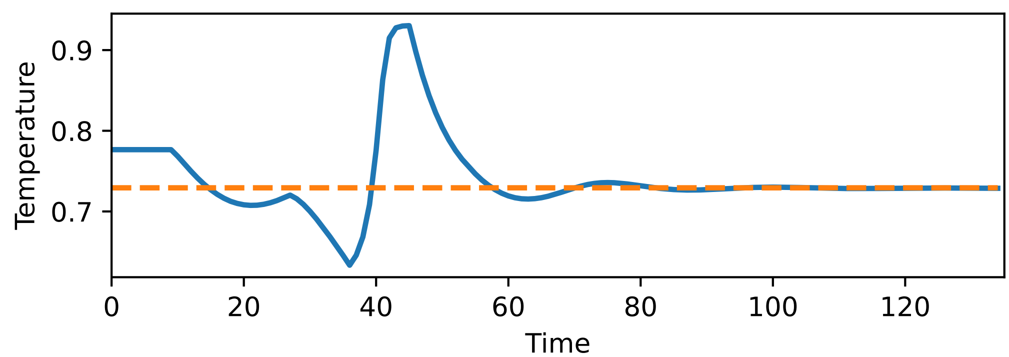
we used an scaled version of the control action such that all decision variables in the objective function features similar order of magnitude values. By trial and error, we came to the conclusion that setting $\lambda_1 = 1, \lambda_2 = 10$ leads to acceptable performance of the manipulated variables over the output variables. Notice, that the y_1 output variable does not appear in the above objective function. Since we only have available a single input variable (u) it would be hard to meet both output variables. Anyway, as shown later, good tracking behaviour of y_1 was also observed. For running the Bayesian optimization calculations the cooling flowrate was set within the bonds: [200,460], the Expected Improvement acquisition function was also used and a sampling rate of 15 minutes was enforced. Even when the Bayesian optimization approach can use a set of initial measurements of the input-output variables for building an initial Gaussian Process surrogate, we decided to use just a single initial measurement. It makes sense to use this single measurement as the initial steady-state conditions vector. In fact, one of the claimed advantages[28] of Bayesian optimization lies in its advantage of finding optimal operating conditions considering only a small set of measurements. As the optimization search progress more information would be available leading to better model representation. In this sense, Bayesian optimization offers a clear advantage over traditional machine learning models which commonly require large data sets for model representation. This advantage is also clear when contrasted with approaches that couple machine learning with evolutionary algorithms, where the size of the dataset ends up becoming a problem for the convergence and performance of evolutionary algorithms.[24,49]

Following, to get a clearer idea about the issues when drawing dynamical optimal product transitions, the nonlinear steady-state multiplicity map of this system is displayed in Figure 2. If the multiplicity map shows monotonic behaviour, then dynamical optimal product transitions should not be difficult to attain. However, some problems could emerge when multiple steady-states are present within the optimality region, since there is the chance of not being able to draw the target steady-states. Accordingly, Figure 2 displays the nonlinear multiplicity map of the Hicks reaction system. As shown in this Figure, the Hicks reaction system clearly features highly nonlinear behaviour around the intended operating region of the reactor.

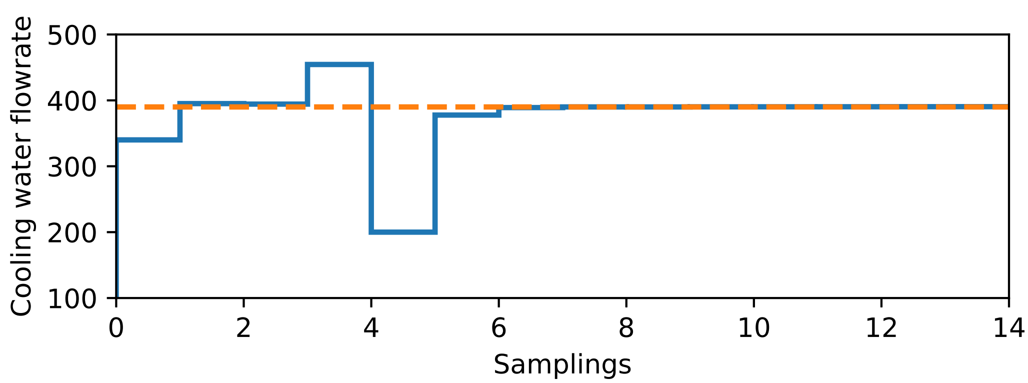
To test the advantages of the data-driven Bayesian optimization approach for optimally driving the Hicks reactor between steady-states some optimal dynamic state transitions are next detailed. First, we drawn the $A \rightarrow B$ open-loop optimal dynamic transition whose location is shown in Figure 2. This is a dynamic transition where A stands for the initial steady-state and B is the final or target steady-state. In addition, as clearly noted from Figure 2, both steady-states are stable. The states and manipulated variables dynamic optimal transitions are shown in Figure 3. As displayed in this Figure, the Bayesian optimization strategy finds the optimal dynamic transition in just a few steps. After hitting the bounds, the input variable stays close to the target operating region. Moreover, as previously mentioned, even when the y_1 variable was not part of the objective function, it also achieves is final steady-state value as a consequence of the reactor dynamic behaviour. Those results can be compared with similar dynamic optimal transitions results drawn using a first-principles dynamic mathematical model (see Figure 6 in Flores et. al.[6]). As noticed, the proposed BO data-driven approach provides acceptable dynamical optimal transition trajectories.



(a) Dimensionless concentration.



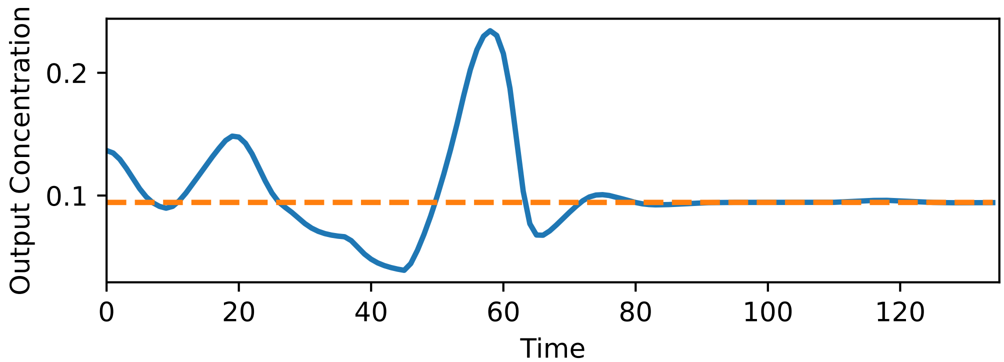
(b) Dimensionless temperature.



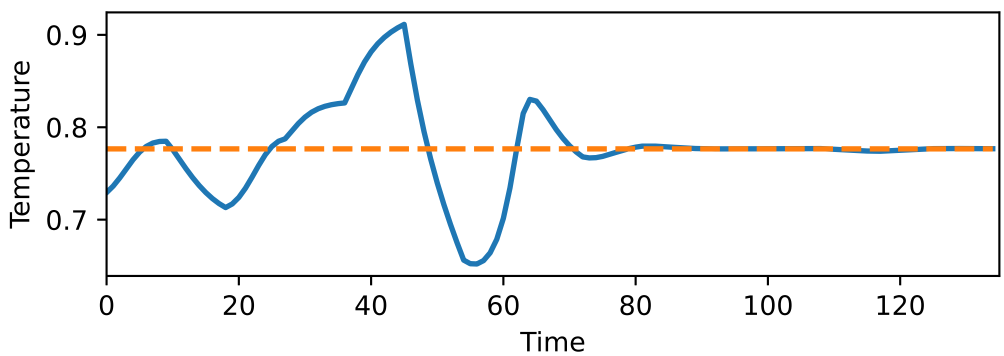
(c) Cooling flowrate.

Figure 3: Hicks reactor optimal dynamic transition using a pure data-driven Bayesian optimization approach: $A \rightarrow B$ product transition. The dashed line stands for the target values of the variables. Initial steady state (y_1, y_2, u):[0.0944, 0.7766,340], final steady-state (y_1, y_2, u): [0.1367,0.7293,390].

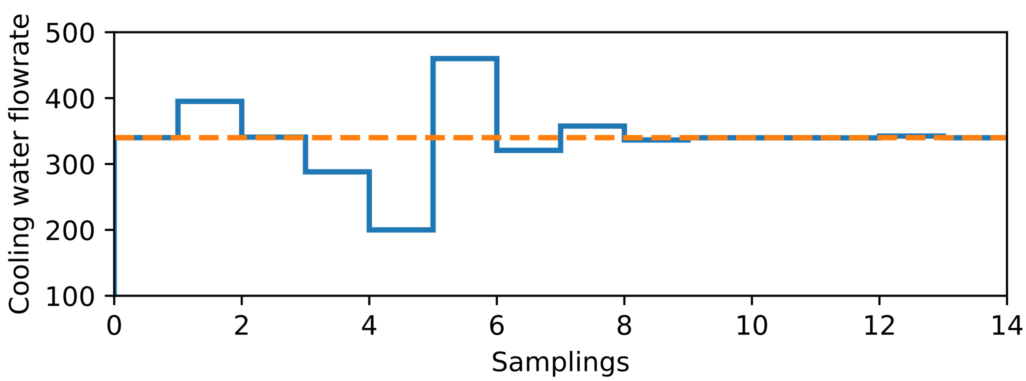
For testing the advantages of the optimization approach an optimal dynamic transition in the opposite direction ($B \rightarrow A$) was also attempted. The results are shown in Figure 4. As displayed in this Figure, the Bayesian optimization approach again quickly finds the dynamic transition trajectory. It should be stressed that this time, to test the ability of the Gaussian Process surrogate to build a reliable process representation and the acquisition function to provide good sampling points, the initial measured operating point was not taken as the initial steady-state. Instead, the single measurement point was set as the final target steady-staten operating vector. Even with this modification the results turn out to be satisfactory. Once again, the y_1 output variable attains its final steady-state value as well.



(a) Dimensionless concentration.



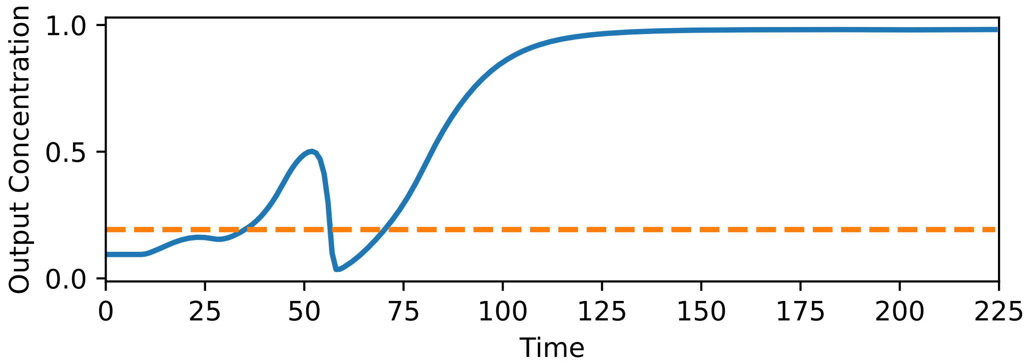
(b) Dimensionless temperature.



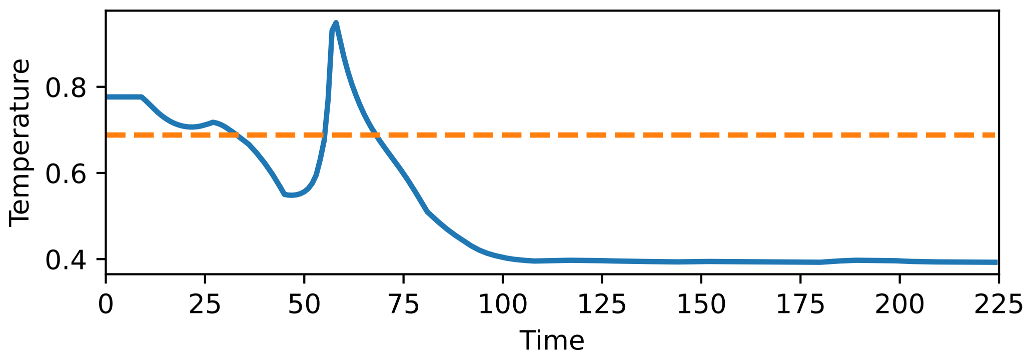
(c) Cooling flowrate.

Figure 4: Hicks reactor optimal dynamic transition using a pure data-driven Bayesian optimization approach: $B \rightarrow A$ product transition.

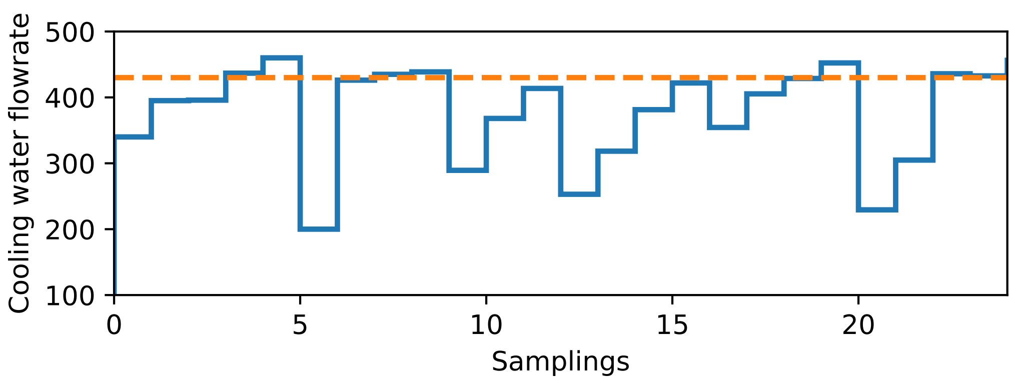
However, as displayed in Figure 5 for the $A \rightarrow C$ transition, when it comes to consider dynamic optimal transitions where the target steady-state (C) is an open-loop unstable operating point, the Bayesian optimization



(a) Dimensionless concentration.



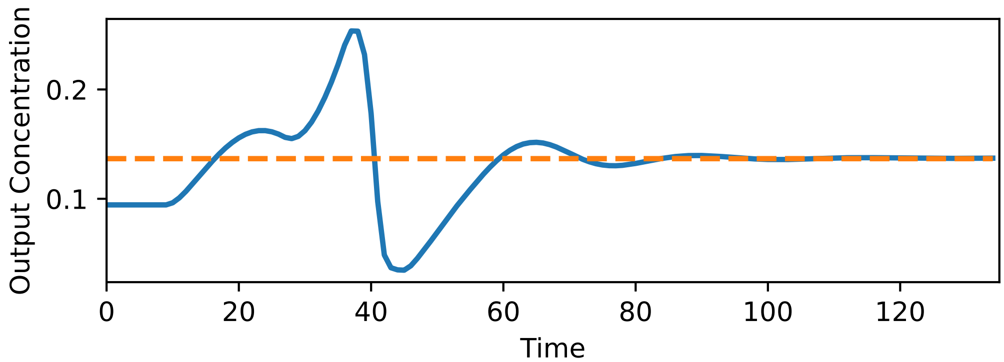
(b) Dimensionless temperature.



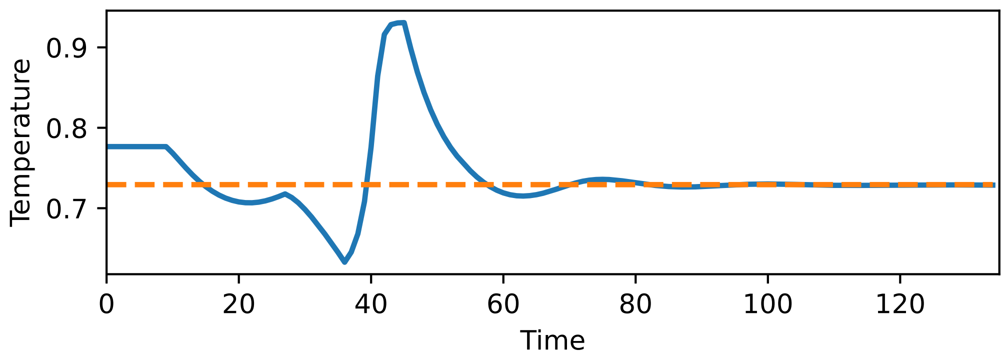
(c) Cooling flowrate.

Figure 5: Hicks reactor optimal dynamic transition using a pure data-driven Bayesian optimization approach: $A \rightarrow C$ product transition. Initial steady state (y_1, y_2, u):[0.0944, 0.7766,340], final steady-state (y_1, y_2, u): [0.1926,0.6881,430].

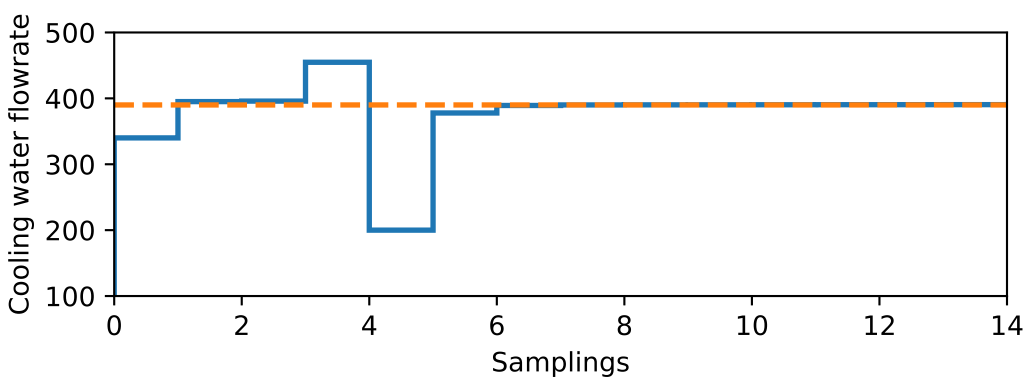
approach fails to draw the optimal trajectory. From Figure 2, it is clear that the optimization approach converges towards the corresponding lower branch stable steady-state. This is a well known drawback of dynamic optimization procedures which directly take into account the dynamic process behaviour for this purpose. As far as we know, the full discretization approach is the only known approach for carrying out this type of calculations [4] (optimal dynamic transitions involving open-loop unstable operating points). However, the full discretization approach requires an explicit first-principles dynamic mathematical model. As mentioned in the discussion about data-driven approaches to process optimization, we have made the assumption that this type of model is not available. There are at least two ways to cope with this problem. First, we could try to stabilise the reaction system by incorporating a control mechanism and then draw the transition trajectories and, second, the transition trajectories can be drawn by using an on-line control algorithm. The first approach leads to the calculation of open-loop optimal trajectories that ought to be later implemented, while the second approach enforces simultaneously drawing optimal trajectories and its implementation. This is a point for future consideration.



(a) Dimensionless concentration.



(b) Dimensionless temperature.



(c) Cooling flowrate.

Figure 6: Hicks reactor optimal dynamic transition using a pure data-driven Bayesian optimization approach: $A \to C$ product transition. Initial steady state (y_1, y_2, u):[0.0944, 0.7766,340], final steady-state (y_1, y_2, u): [0.1926,0.6881,430]. To assess the influence of noisy measurements on the Bayesian optimization performance 20% noise error was added to the objective function.

In practical scientific and technical applications the presence of noise, when measurements are drawn, is inevitable.

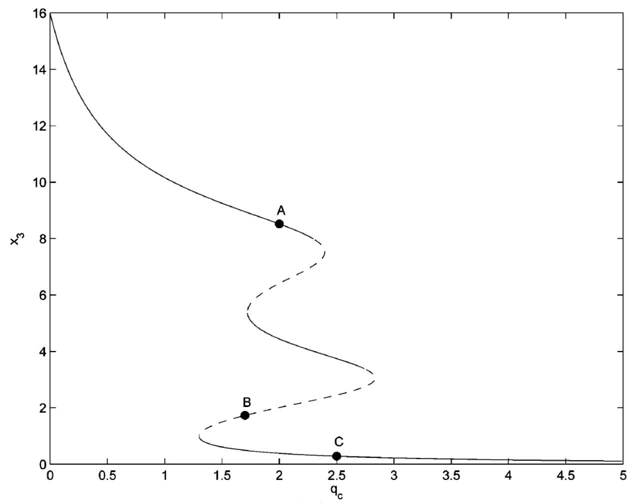


Figure 7: Nonlinear steady-state multiplicity map of the consecutive reaction system. The continuous line stands for stable states, while the dashed line represent unstable steady-states.

The presence of noisy measurements tends also to outperform the performance of numerical optimization algorithms making difficult to draw optimality operating regions. However, since noise effects are inevitable, its impact on the quality of the optimal dynamic transition trajectories should be assessed. Therefore, to take into account the presence of noisy measurements on the performance of the Bayesian optimization approach, we have added 20% noise to the objective function shown in Equation 3 by just drawing random numbers in the [-1,+1] interval. The dynamic optimal transition trajectories are displayed in Figure 6. As observed, even in presence of significant noise levels, the Bayesian optimization strategy successfully tracks the dynamic optimal transition trajectory meeting the product requirements. However, to tolerate this noise level and to draw the optimality region we have to set the weight value (λ_2) in Equation 3 up to 1x10^5. This change makes sense since its effect is to give priority to the effect of the manipulated variables on the output system response.

Consecutive reaction system

In this case study we will address the data-driven dynamic optimization of a consecutive reaction system.[6] As displayed in Figure 7 the system exhibits multiplicity steady-state nonlinear behaviour around the intended operating region. The dynamic mathematical model of this system reads as follows:

$$\frac{dx_1}{dt} = q(x_{1f} - x_1) - x_1 k_1 \phi \tag{4}$$

$$\frac{dx_2}{dt} = q(x_{2f} - x_2) - x_2 \phi S k_2 + x_1 \phi k_1 \tag{5}$$

$$\frac{dx_3}{dt} = q(x_{3f} - x_3) + \delta(x_4 - x_3) + \beta\phi(x_1 k_1 + \alpha x_2 k_2 S) \tag{6}$$

$$\frac{dx_4}{dt} = \delta_1(q_c(x_{4f} - x_4) + \delta\delta_2(x_3 - x_4)) \tag{7}$$

where,

$$k_1 = e^{x_3/(1+(x_3/\gamma))} \tag{8}$$

$$k_2 = e^{\psi x_3/(1+(x_3/\gamma))} \tag{9}$$

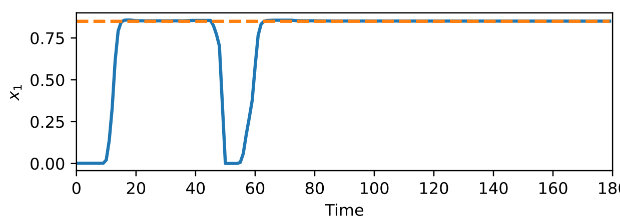
in the above equations x_1 is the dimensionless concentration of reactant X, x_2 is the dimensionless concentration of reactant Y, x_3 is the dimensionless reactor temperature, and x_4 is the dimensionless cooling jacket temperature.

The nominal parameters values are given as follows: $q = 1$, $x_{1f} = 1$, $x_{2f} = 0$, $x_{3f} = 0$, $x_{4f} =$ -1, $\beta = 8$, $\phi = 0.133$, $\delta = 1$, $\alpha = 1$, $S = 0.01$, $\psi = 1$, $\delta_1 = 10$, $\delta_2 = 1$, $\gamma = 1000$.

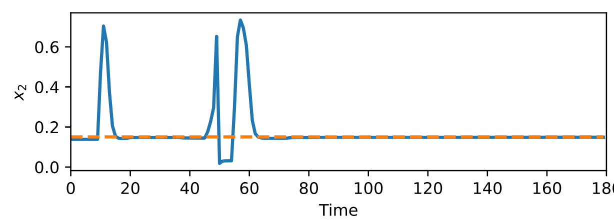
In this case study we draw the dynamic optimal transition trajectories between steady-state around the multiplicity region. To this end we will consider that the cooling flowrate (q_c) is the manipulated variable and that at the end of the transition all states should achieve its target values. Because there is only a single input available for driving the system, we will assume that good tracking of all states can be achieved trough the control of the dimensionless reactor temperature (x_3). Accordingly, the objective function (Ω) is defined as follows:

$$\Omega = \lambda_1(x_3 - x_3^t)^2 + \lambda_2 \left(\frac{q_c - q_c^t}{q_c^t} \right)^2 \tag{10}$$

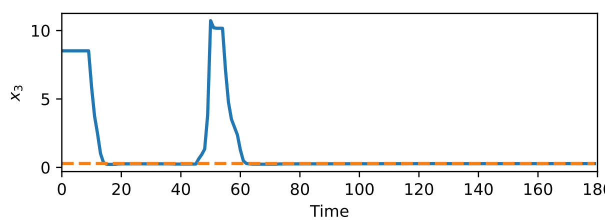
where the superscript t stands for target values and, as in the past case study, λ_1, λ_2 are weighting functions used to stress the importance of the input and output terms. For the noiseless measurements scenario the following values of the weights provided good dynamic open-loop performance: $\lambda_1 = 1, \lambda_2 = 10$. Moreover, a 15 minutes sampling rate was set and the manipulated variable (q_c) was constrained within the following closed interval: [1,3]. The optimal dynamic transition trajectories between the A and C operating points,[6] shown in Figure 7, are displayed in Figure 8. As shown in this Figure, good performance of the Bayesian optimization approach for drawing dynamic optimal transition trajectories is observed. As done in the previous case study, the results of the present case study can be compared with similar dynamic optimal transitions results (see Figure 2 in Flores et. al.[6]). Once again, the proposed BO data-driven approach displays comparable quality dynamical optimal transition trajectories.



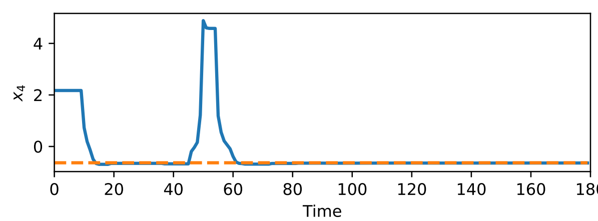
(a) x_1 dimensionless concentration.



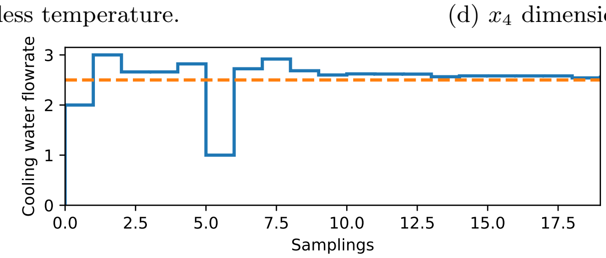
(b) x_2 dimensionless concentration.



(c) x_3 dimensionless temperature.



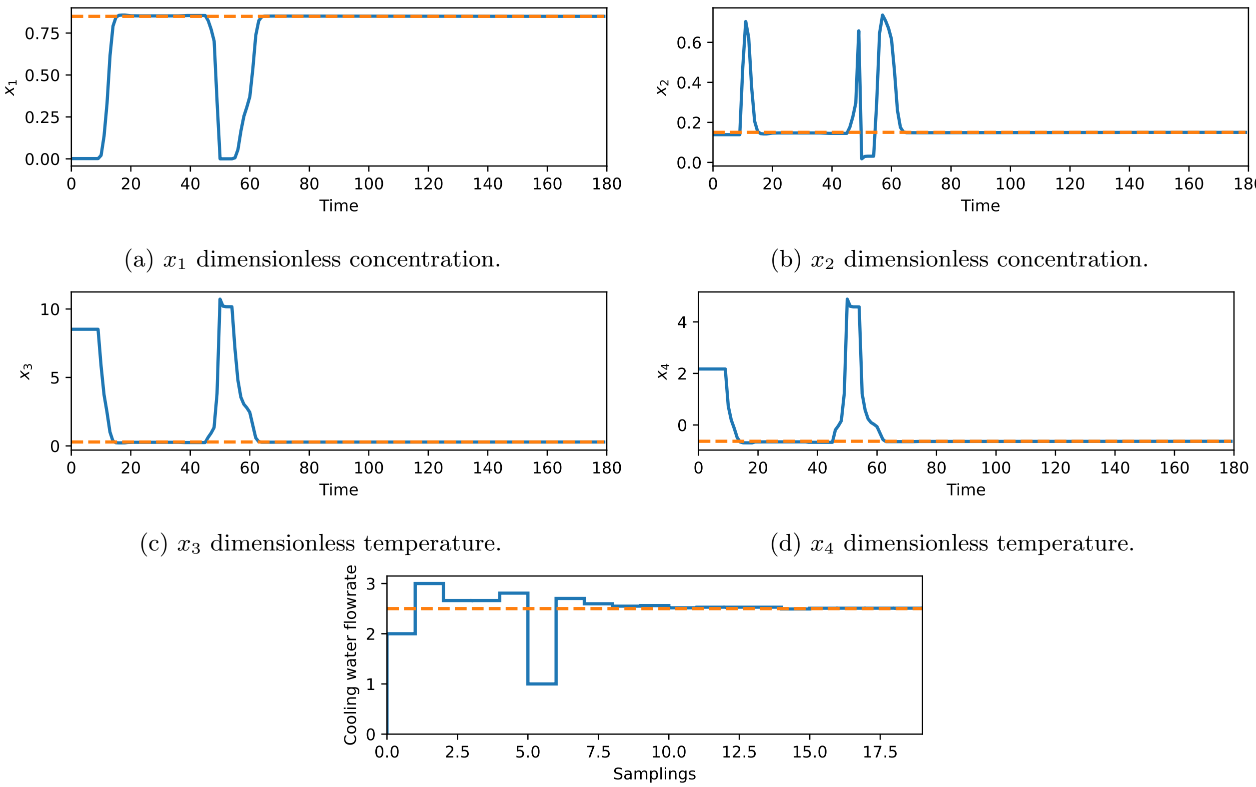
(d) x_4 dimensionless temperature.



(e) Cooling flowrate.

Figure 8: Consecutive reaction system. Optimal dynamic transition trajectories between the A and C steady-states shown in Figure 7. Initial steady-state $(x_1, x_2, x_3, x_4, q_c) =$ [0.0016,0.1387,8.5188,2.1729,2]; Final steady-state $(x_1, x_2, x_3, x_4, q_c) =$ [0.8495,0.1503,0.2871,-0.6323,2.5]

As done in the past case study, an assessment of the effect of noisy measurements on the performance of the optimization strategy was done. To this end 20% noise was added to the objective function. Figure 9 displays the dynamic optimal trajectories under a noisy measurements scenario. To smooth the effect of noisy measurements on optimal system response the weighting functions were set as follows: $\lambda_1 = 10^2, \lambda_2 = 10^4$. It should be stressed that even in presence of considerable noise levels, the Bayesian optimization approach is clearly able to draw dynamic optimal transition trajectories. This fact becomes notoriously important especially when highly nonlinear processing systems are addressed.



(a) x_1 dimensionless concentration.

(b) x_2 dimensionless concentration.

(c) x_3 dimensionless temperature.

(d) x_4 dimensionless temperature.

(e) Cooling flowrate.

Figure 9: Consecutive reaction system for the $A \to C$ transition as shown in Figure 7 under a noisy measurements scenario.

Binary distillation column

This case study deals with the dynamic optimization of an ideal equilibrium binary distillation column as introduced by Morari and Zafiriou.[50] The model detailed in[50] is a basic model of the distillation operation that only takes into account composition dynamics, with no consideration for holdup dynamics as discussed in[51] or tray thermal effects. However, we decided to consider the relatively simple model discussed in[50] because it suits our purpose of demonstrating how to draw optimal dynamic transition trajectories by data-driven optimization techniques in a complex large scale nonlinear system. Accordingly the composition dynamic model reads as follows.

- Reboiler ($j = 1$)

$$\frac{dx_j}{dt} = \frac{Lx_{j+1} - Vy_j - Bx_j}{M_b} \tag{11}$$

- Feed tray ($j = N_f$) (saturated liquid feed stream)

$$\frac{dx_j}{dt} = \frac{Lx_{j+1} + Vy_{j-1} - Lx_j - Vy_j + Fz_f}{M_t} \tag{12}$$

- Condenser ($j = N_t$)

$$\frac{dx_j}{dt} = \frac{Vy_{j-1} - Lx_j - Dx_j}{M_c} \tag{13}$$

- Inner trays other than feed, condenser and reboiler trays $j \neq 1, j \neq N_f, j \neq N_t$

$$\frac{dx_j}{dt} = \frac{Lx_{j+1} + Vy_{j-1} - Lx_j - Vy_j}{M_t}, \quad \forall j > N_f \tag{14}$$

$$\frac{dx_j}{dt} = \frac{(L+F)x_{j+1} + Vy_{j-1} - (L+F)x_j - Vy_j}{M_t}, \quad \forall j < N_f \tag{15}$$

- Ideal vapor and liquid equilibrium relationship

$$y_j = \frac{\alpha x_j}{1 + (\alpha - 1)x_j} \tag{16}$$

For dynamic optimization purposes we consider the operation of the column from a nominal steady-state and take it to a final target steady state. For achieving the transition trajectory we assume that the reflux flowrate (L) and the boilup flowrate (V) are the manipulated variables, while the output target variables are the distillate and bottoms compositions. This corresponds to a classical LV two-points control structure. Additional data are as follows: $\alpha = 1.5$, $z_f = 0.5$, $N_t = 40$, $N_f = 21$, $F = 1$, $M_c = 2$, $M_b = 5$ and $M_t = 0.5$. Moreover, for Bayesian optimization purposes box constraints on the manipulated variables were enforced as follows: $L \in [2, 3], V \in [3, 4]$, a sampling rate of 30 minutes was set, and decisions about where to take optimal samples were drawn using the Expected Improvement acquisition function.

For this case study the objective function (Ω) was defined as follows:

$$\Omega = (x_b - x_b^t)^2 + (x_d - x_d^t)^2 + \lambda_L(L - L^t)^2 + \lambda_V(V - V^t)^2 \tag{17}$$

where x stands for mol fraction, the sub-index d, b denote distillate and bottoms streams, respectively, and the superscript t is the target value. λ is a weighting function to stress control actions. After trial and error, we came to the conclusion that by setting $\lambda_L = \lambda_V = 10^2$ good performance of the dynamic optimization behaviour was drawn. No scaling of the objective function variables was needed since all these variables feature values in the same order of magnitude.

Accordingly, Figure 10 displays the dynamic optimal transition trajectory when changing from the A to the C steady-state operating points as defined in the appendix section of Morari and Zafiriou.[50] As observed, good tracking behaviour is drawn successfully meeting product compositions targets within the manipulated variables closed interval bounds. It should be stressed that the optimal control of distillation columns is a complex task especially when dual composition control is requested. Even, so the Bayesian optimization strategy quickly draws the dynamic trajectory. As in the past example, a single initial measurement was deployed. The single measurement refers to the initial steady-state conditions of the A operating point.[50] Since our aim is to draw optimal transition trajectories, it makes sense to use this point as the initial measurement vector instead of using random initial measurements.

Finally, as done in the previous case studies, the influence of noisy measurements on the performance of the Bayesian optimization approach was assessed by adding 20% noise level to the objective function shown in Equation 17. The performance of the optimizationn scheme taking into account noisy measurements is displayed in Figure 11. To cope with noisy measurements and to draw the optimal transitions trajectories the weighting function values in Equation 17 were set as follows: $\lambda_L = \lambda_V = 10^3$. As shown in 11, even in presence of noisy measurements, the Bayesian approach successfully draws optimal transition trajectories meeting product composition requirements. In fact, as shown from Figures 10 and 11, the dynamic shape of the manipulated variables is similar. Once again, increasing the weighting functions values have the effect of stressing the importance of the manipulated variables on the column distillate and bottoms compositions.

5 Conclusions

In this work, a data-driven Bayesian approach for optimal dynamic product transitions has been proposed. The approach is based on a dynamic optimization problem that is solved using a Bayesian optimization algorithm. The approach has been applied to three case complex nonlinear behaviour case studies. The results show that the proposed approach is able to find dynamical optimal transition trajectories that meet the product composition requirements. The approach is also able to cope with noisy measurements, which is an important feature in real-world applications. The results also show that the proposed approach is able to find optimal transition trajectories that are similar to those obtained using a traditional optimization approach, but with a much lower computational cost. The proposed approach has several advantages over traditional optimization approaches. First, it is data-driven, which means that it does not require a priori knowledge of the system dynamics. Second, it is able to cope with noisy measurements, which is an important feature in real-world applications. Third, it is computationally efficient, which makes it suitable for on-line real-time optimization. In conclusion, the proposed data-driven Bayesian approach for optimal dynamic product transitions is a promising approach that has the potential to improve the efficiency and effectiveness of dynamic product transitions in the processing industry.

AUTHOR CONTRIBUTIONS

Antonio Flores-Tlacuahuac: Conceptualization (equal); funding acquisition(equal); investigation (equal); methodology (equal); supervision (equal);writing–review and editing (equal). **Luis-Fabian Fuentes-Cortés**: Conceptualization (equal); investigation (equal); methodology (equal); supervision (equal);writing–review and editing (equal).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon request.

ORCID

Antonio Flores-Tlacuahuac: https://orcid.org/0009-0004-1291-0973.
Luis-Fabian Fuentes-Cortés: https://orcid.org/0000-0003-1191-6957.

SUPPLEMENTARY MATERIAL

- computer codes of the BO dynamic product transition algorithm as implemented for the three case studies of this work can be found in the following zipped file: **BO-codes.zip** .
- All the results figures in the paper are generated within the above computer codes.

References

[1] Antonio Flores-Tlacuahuac, E. S.-G., Lorenz T. Biegler Optimal Grade Transitions in the High-Impact Polystirene Polymerization Process. *Industrial and Engineering Chemistry Research* **2006**, *45*, 6175–6189, DOI: http://dx.doi.org/10.1021/ie051140q.

[2] Shi, J.; Biegler, L. T.; Hamdan, I.; Wassick, J. Optimization of grade transitions in polyethylene solution polymerization process under uncertainty. *Computers & Chemical Engineering* **2016**, *95*, 260–279, DOI: https://doi.org/10.1016/j.compchemeng.2016.08.002.

[3] Biegler, L. T. Advanced optimization strategies for integrated dynamic process operations. *Computers & Chemical Engineering* **2018**, *114*, 3–13, DOI: https://doi.org/10.1016/j.compchemeng.2017.10.016, FOCAPO/CPC 2017.

[4] Antonio Flores-Tlacuahuac, E. S.-G., Lorenz T. Biegler Dynamic Optimization of HIPS Open-Loop Unstable Polymerization Reactors. *Industrial and Engineering Chemistry Research* **2005**, *44*, 2659–2674, DOI: http://dx.doi.org/10.1021/ie049534p.

[5] Antonio Flores Tlacuahuac, G. R.-M., Enrique Saldivar-Guerra Grade Transition Dynamic Simulation of HIPS Polymerization Reactors. *Computers & Chemical Engineering* **2005**, *30*, 357–375, DOI: http://dx.doi.org/10.1016/j.compchemeng.2005.10.002.

[6] Antonio Flores-Tlacuahuac, L. T. B., Sebastian Terrazas-Moreno Global Optimization of Highly Nonlinear Dynamic Systems. *Industrial and Engineering Chemistry Research* **2008**, *47*, 2643–2655, DOI: http://dx.doi.org/10.1021/ie070379z.

[7] Chu, Y.; You, F. Integration of production scheduling and dynamic optimization for multi-product CSTRs: Generalized Benders decomposition coupled with global mixed-integer fractional programming. *Computers & Chemical Engineering* **2013**, *58*, 315–333, DOI: https://doi.org/10.1016/j.compchemeng.2013.08.003.

[8] Wang, Y.; Biegler, L. T.; Ostace, G. S.; Majewski, R. A. Optimal polymer grade transitions for fluidized bed reactors. *Journal of Process Control* **2020**, *88*, 86–100, DOI: https://doi.org/10.1016/j.jprocont.2020.02.001.

[9] Shirude, S.; Padhiyar, N. Optimal grade transition of a non-isothermal continuous reactor with multi-objective dynamic optimization approach. *Chemical Engineering Research and Design* **2019**, *147*, 63–72, DOI: https://doi.org/10.1016/j.cherd.2019.04.040.

[10] Medina-Herrera, N.; Jiménez-Gutiérrez, A.; Ricardez-Sandoval, L. A.; Tututi-Avila, S. An approach for dynamic transitions in multiproduct reactive distillation columns. *Chemical Engineering and Processing - Process Intensification* **2020**, *153*, 107967, DOI: https://doi.org/10.1016/j.cep.2020.107967.

[11] Remigio, J. E.; Swartz, C. L. Production scheduling in dynamic real-time optimization with closed-loop prediction. *Journal of Process Control* **2020**, *89*, 95–107, DOI: https://doi.org/10.1016/j.jprocont.2020.03.009.

[12] Liñán, D. A.; Ricardez-Sandoval, L. A. Optimal design and dynamic transitions of multitask catalytic distillation columns: A Discrete-Steepest Descend Framework. *Chemical Engineering and Processing - Process Intensification* **2022**, *180*, 108655, DOI: https://doi.org/10.1016/j.cep.2021.108655.

[13] Jorge Nocedal, S. J. W. *Numerical optimization*; Springer, 2006.

[14] Biegler, L. T. *Nonlinear programming*; SIAM, 2010.

[15] Grossmann, I. E. *Advanced Optimization for Process Systems Engineering*; Cambridge University Press, 2021.

[16] da Costa Oliveira, A. L.; Britto, A.; Gusmão, R. A framework for inverse surrogate modeling for fitness estimation applied to Multi-Objective Evolutionary Algorithms. *Applied Soft Computing* **2023**, *146*, 110672, DOI: `https://doi.org/10.1016/j.asoc.2023.110672`.

[17] Negrellos-Ortiz, I.; Flores-Tlacuahuac, A.; Gutiérrez-Limón, M. Product dynamic transitions using a derivative-free optimization trust-region approach. *Industrial and Engineering Chemistry Research* **2016**, *55*, 8586–8601, DOI: `https://doi.org/10.1021/acs.iecr.6b00268`.

[18] Liu, Y.; Tang, L.; Liu, C.; Su, L.; Wu, J. Black box operation optimization of basic oxygen furnace steelmaking process with derivative free optimization algorithm. *Computers & Chemical Engineering* **2021**, *150*, 107311, DOI: `https://doi.org/10.1016/j.compchemeng.2021.107311`.

[19] Ma, K.; Sahinidis, N. V.; Bindlish, R.; Bury, S. J.; Haghpanah, R.; Rajagopalan, S. Data-driven strategies for extractive distillation unit optimization. *Computers & Chemical Engineering* **2022**, *167*, 107970, DOI: `https://doi.org/10.1016/j.compchemeng.2022.107970`.

[20] van de Berg, D.; Savage, T.; Petsagkourakis, P.; Zhang, D.; Shah, N.; del Rio-Chanona, E. A. Data-driven optimization for process systems engineering applications. *Chemical Engineering Science* **2022**, *248*, 117135, DOI: `https://doi.org/10.1016/j.ces.2021.117135`.

[21] de Oliveira, G. C.; Bertone, E.; Stewart, R. A. Optimisation modelling tools and solving techniques for integrated precinct-scale energy–water system planning. *Applied Energy* **2022**, *318*, 119190, DOI: `https://doi.org/10.1016/j.apenergy.2022.119190`.

[22] Abouhawwash, M.; Seada, H.; Deb, K. Towards faster convergence of evolutionary multi-criterion optimization algorithms using Karush Kuhn Tucker optimality based local search. *Computers & Operations Research* **2017**, *79*, 331–346, DOI: `https://doi.org/10.1016/j.cor.2016.04.026`.

[23] Del Ser, J.; Osaba, E.; Molina, D.; Yang, X.-S.; Salcedo-Sanz, S.; Camacho, D.; Das, S.; Suganthan, P. N.; Coello Coello, C. A.; Herrera, F. Bio-inspired computation: Where we stand and what's next. *Swarm and Evolutionary Computation* **2019**, *48*, 220–250, DOI: `https://doi.org/10.1016/j.swevo.2019.04.008`.

[24] Hellwig, M.; Beyer, H.-G. Benchmarking evolutionary algorithms for single objective real-valued constrained optimization – A critical review. *Swarm and Evolutionary Computation* **2019**, *44*, 927–944, DOI: `https://doi.org/10.1016/j.swevo.2018.10.002`.

[25] Mockus, J.; Eddy, W.; Mockus, A.; Mockus, L.; Reklaitis, G. *Bayesian Heuristic Approach to Discrete and Global Optimization: Algorithms, Visualization, Software, and Applications*; Springer, 1997.

[26] Carl Edward Rasmussen, C. K. I. W. *Gaussian Processes for Machine Learning*; The MIT Press, 2006.

[27] Hennign, P.; Osborne, M. A.; Kersting, H. *Probabilistic Numerics*; Cambridge University Press, 2022.

[28] Garnett, R. *Bayesian optimization*; Cambridge University Press, 2023.

[29] Francesco Archetti, A. C. *Bayesian Optimization and Data Science*; Springer, 2019.

[30] Packwood, D. *Bayesian Optimization for Material Science*; Springer, 2017.

[31] Nguyen, Q. *Bayesian Optimization in action*; Manning Publications, 2023.

[32] Nguyen, Q. *Experimentation for Engineers*; Manning Publications, 2023.

[33] Urm, J. J.; Choi, J. H.; Kim, C.; Lee, J. M. Techno-economic analysis and process optimization of a PET chemical recycling process based on Bayesian optimization. *Computers & Chemical Engineering* **2023**, *179*, 108451, DOI: https://doi.org/10.1016/j.compchemeng.2023.108451.

[34] Begall, M. J.; Schweidtmann, A. M.; Mhamdi, A.; Mitsos, A. Geometry optimization of a continuous millireactor via CFD and Bayesian optimization. *Computers & Chemical Engineering* **2023**, *171*, 108140, DOI: https://doi.org/10.1016/j.compchemeng.2023.108140.

[35] Coutinho, J. P.; Santos, L. O.; Reis, M. S. Bayesian Optimization for automatic tuning of digital multi-loop PID controllers. *Computers & Chemical Engineering* **2023**, *173*, 108211, DOI: https://doi.org/10.1016/j.compchemeng.2023.108211.

[36] Lu, Q.; González, L. D.; Kumar, R.; Zavala, V. M. Bayesian optimization with reference models: A case study in MPC for HVAC central plants. *Computers & Chemical Engineering* **2021**, *154*, 107491, DOI: https://doi.org/10.1016/j.compchemeng.2021.107491.

[37] Park, S.; Na, J.; Kim, M.; Lee, J. M. Multi-objective Bayesian optimization of chemical reactor design using computational fluid dynamics. *Computers & Chemical Engineering* **2018**, *119*, 25–37, DOI: https://doi.org/10.1016/j.compchemeng.2018.08.005.

[38] González, L. D.; Zavala, V. M. New paradigms for exploiting parallel experiments in Bayesian optimization. *Computers & Chemical Engineering* **2023**, *170*, 108110, DOI: https://doi.org/10.1016/j.compchemeng.2022.108110.

[39] Pourmohamad, T.; Lee, H. *Bayesian optimization with application to computer experiments*; SpringerBriefs in Statistics; Springer International Publishing, 2021.

[40] Liu, P. *Bayesian optimization: Theory and practice using Python*; Apress, 2023.

[41] Valencia-Rivera, G. H.; Benavides-Robles, M. T.; Vela Morales, A.; Amaya, I.; Cruz-Duarte, J. M.; Ortiz-Bayliss, J. C.; Avina-Cervantes, J. G. A systematic review of metaheuristic algorithms in electric power systems optimization. *Applied Soft Computing* **2024**, *150*, 111047, DOI: https://doi.org/10.1016/j.asoc.2023.111047.

[42] Camacho-Vallejo, J.-F.; Corpus, C.; Villegas, J. G. Metaheuristics for bilevel optimization: A comprehensive review. *Computers & Operations Research* **2024**, *161*, 106410, DOI: https://doi.org/10.1016/j.cor.2023.106410.

[43] Sansana, J.; Joswiak, M. N.; Castillo, I.; Wang, Z.; Rendall, R.; Chiang, L. H.; Reis, M. S. Recent trends on hybrid modeling for Industry 4.0. *Computers & Chemical Engineering* **2021**, *151*, 107365, DOI: https://doi.org/10.1016/j.compchemeng.2021.107365.

[44] Jaynes, E. *Probability Theory*; Cambridge University Press, 2003.

[45] Gelman, A.; Vehtari, A.; Carlin, J. B.; Stern, H.; Rubin, D.; Dunson, D. B. *Bayesian Data Analysis*; CRC Press, 2014.

[46] Calvetti, D.; Somersalo, E. *Introduction to Bayesian Scientific Computing*; Springer, 2007.

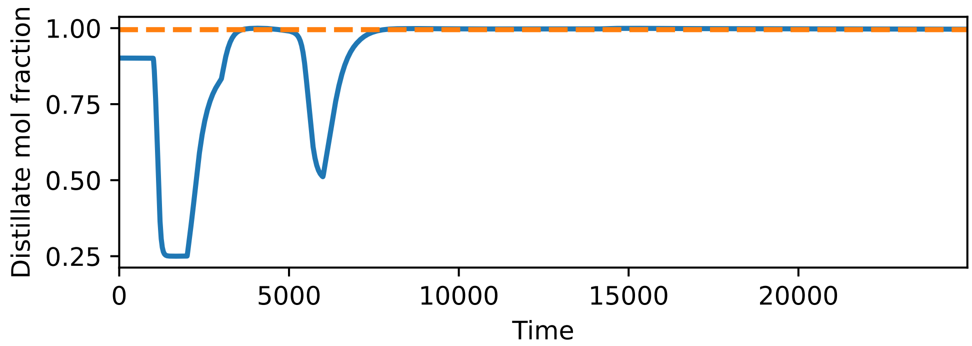
[47] Nogueira, F. Bayesian Optimization: Open source constrained global optimization tool for Python. 2014–; https://github.com/fmfn/BayesianOptimization.

[48] Hicks, W. H., G. A.; Ray Approximation Methods for Optimal Control Synthesis. *Canadian Journal of Chemical Engineering* **1971**, *40*, 522–529.

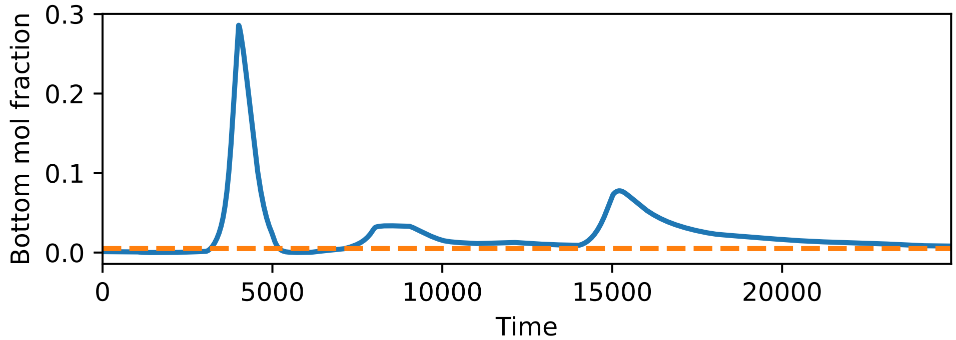
[49] Campelo, F.; Takahashi, F. Sample size estimation for power and accuracy in the experimental comparison of algorithms. *Journal of Heuristics* **2019**, *25*, 305–338, DOI: 10.1007/s10732-018-9396-7.

[50] Morari, M.; Zafiriou, E. *Robust Process Control*; Prentice-Hall, 1989.

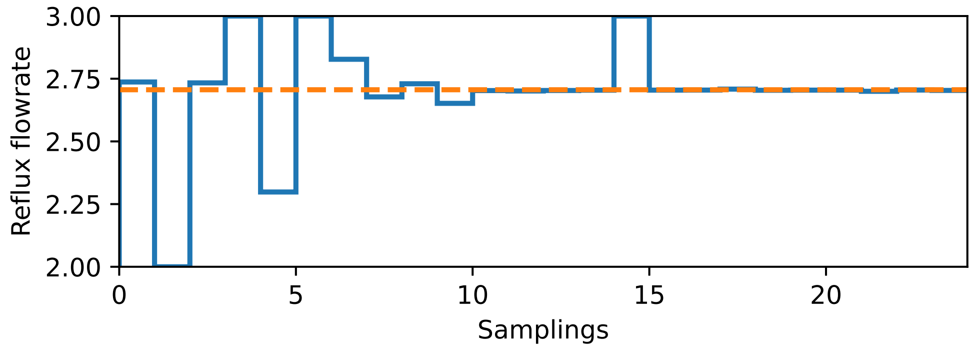
[51] Skogestad, S.; Postlethwaite, I. *Multivariable Feedback Control: Analysis and Design*; Wiley, 2005.



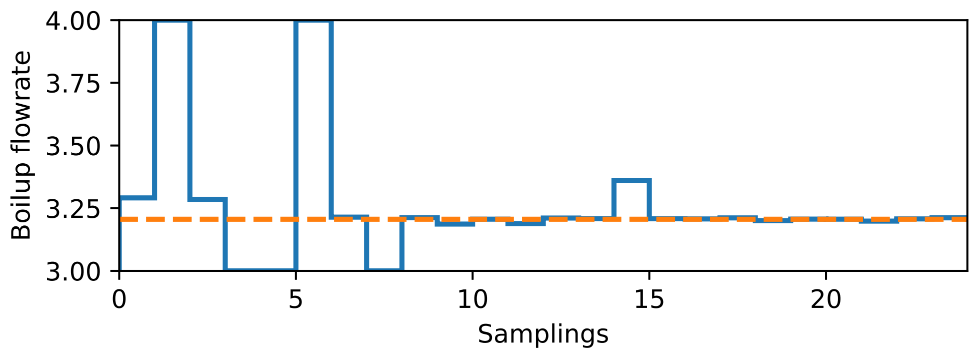
(a) Distillate concentration.



(b) bottoms concentration.

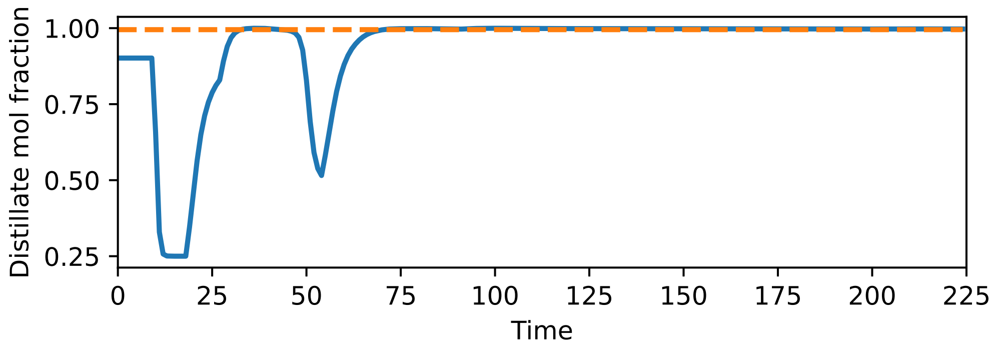


(c) Reflux flowrate.

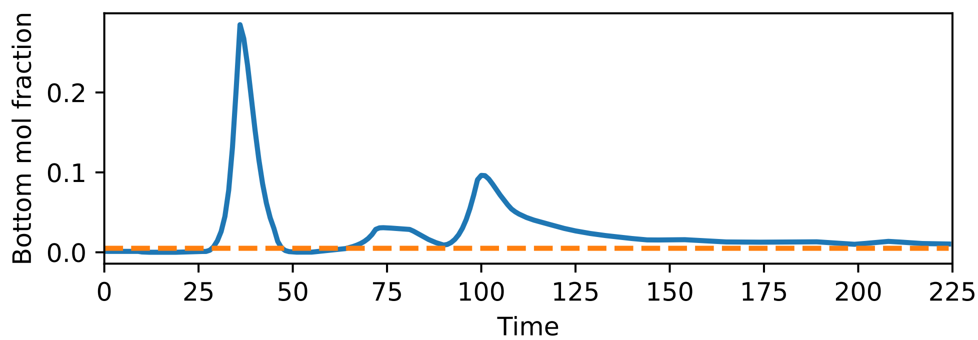


(d) Boilup flowrate.

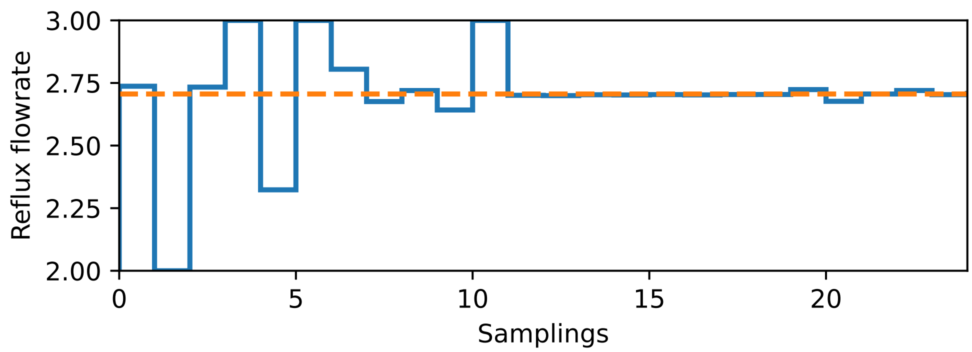
Figure 10: Ideal binary distillation column. Initial conditions vector (x_d, x_b, L, V): [0.9016,0.0015,2.737,3.291], Target conditions vector (x_d, x_b, L, V): [0.995,0.005,2.706,3.206]. The dashed line stands for the corresponding target values.



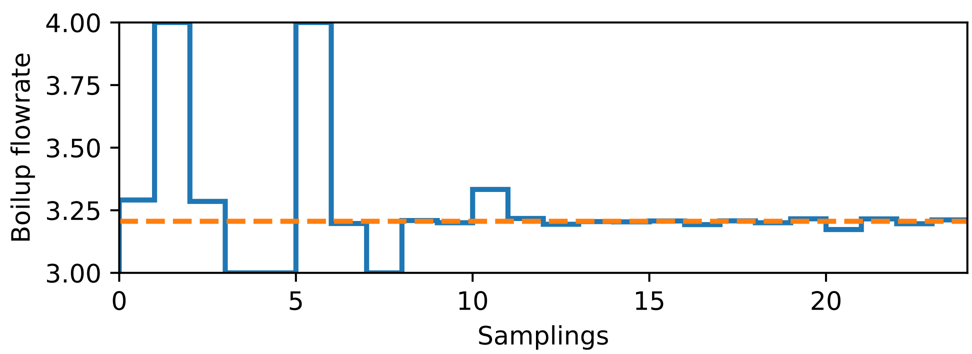
(a) Distillate concentration.



(b) bottoms concentration.



(c) Reflux flowrate.



(d) Boilup flowrate.

Figure 11: Ideal binary distillation column. Initial conditions vector (x_d, x_b, L, V): [0.9016,0.0015,2.737,3.291], Target conditions vector (x_d, x_b, L, V): [0.995,0.005,2.706,3.206]. The dashed line stands for the corresponding target values. To assess the influence of noisy measurements on the Bayesian optimization performane 20% noise error was added to the objective function.