Journal Pre-proof



Global surrogate process model adaptation using self-adaptive characteristic maps for cost-optimal production scheduling on the example of a power-to-methanol process

Max Kollmer, Mücahit Terzi, Akasha Nonavinakere Venkatesh, Mohit Singh, Jörg Matthes

PII: S2772-5081(26)00066-9
DOI: https://doi.org/10.1016/j.dche.2026.100353
Reference: DCHE 100353

To appear in: *Digital Chemical Engineering*

Received date : 12 August 2026
Revised date : 28 September 2026
Accepted date : 21 October 2026

Please cite this article as: M. Kollmer, M. Terzi, A.N. Venkatesh et al., Global surrogate process model adaptation using self-adaptive characteristic maps for cost-optimal production scheduling on the example of a power-to-methanol process. *Digital Chemical Engineering* (2026), doi: https://doi.org/10.1016/j.dche.2026.100353.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Published by Elsevier Ltd on behalf of Institution of Chemical Engineers (IChemE). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Journal Pre-proof

Revised manuscript (unmarked, editable source file)

Global surrogate process model adaptation using self-adaptive characteristic maps for cost-optimal production scheduling on the example of a power-to-methanol process

Max Kollmer[a,*], Mücahit Terzi[b], Akasha Nonavinakere Venkatesh[b], Mohit Singh[b], Jörg Matthes[a]

[a] *Karlsruhe Institute of Technology, Institute for Automation and Applied Informatics, Hermman-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen, 76344, Baden-Württemberg, Germany*
[b] *Karlsruhe Institute of Technology, Institute for Micro Process Engineering, Hermman-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen, 76344, Baden-Württemberg, Germany*

Abstract

This paper describes a framework for the flexible scheduling of a novel power-to-methanol process, which uses self-adaptive characteristic maps as surrogate process models. A newly developed global adaptation algorithm adjusts the process model during operation using new measurements to account for plant-model mismatch due to sudden or incremental drifts in process behavior. Through parallel drift detection, structural errors in the characteristic map are identified at the operating points in use, and the characteristic map is then adjusted across the entire operating range using Gaussian process regression (GPR). This ensures an accurate characteristic map across the entire operating range. The adapted characteristic maps are then used to calculate a cost-optimal schedule, taking into account various production constraints and dynamic electricity prices, which is used in a plant emulation to calculate the outputs.

The results show a rapid adaptation of the characteristic maps to changing process behavior using the newly developed algorithm. On average, the reduction in plant-model mismatch using the root mean squared error (RMSE) between the ground truth and the model characteristic map, considering a sudden drift of the process behavior, ranges between 76.7 % and 97.6 % using the developed global algorithm, while the local algorithm used as a benchmark has a significantly lower reduction of the RMSE of about 5 % to 12.7 %. The smaller plant-model mismatch reduces production costs by an average of 11.8 % compared to local adaptation and 21.4 % compared to no adaptation, considering various test cases. At the same time, the required production target is met more accurately and production constraints are met more often when using the global algorithm. In the case of incremental drift of the process behavior, production costs and the plant-model mismatch are also significantly lower compared to local or no adaptation.

Keywords: PtX, PtM, process scheduling, characteristic maps, model adaptation, plant-model mismatch

1. Introduction

The steadily increasing share of renewable energy necessitates efficient storage solutions to utilize excess renewable energy during periods of oversupply. Beyond traditional battery storage, load shifting in industrial processes offers an innovative way to stabilize the grid and supply energy when neither solar nor wind power is available. Consequently, demand response and demand-side management are becoming increasingly important for the industry [1]. Energy-intensive chemical processes, for example, can ensure grid stability through load shifting or load shedding, while simultaneously reducing their own costs by shifting their main energy consumption to times of lower prices. This is particularly relevant for flexibly operable chemical processes, which can dynamically adjust their production capacity — and thus their load — based on electricity prices. One example are Power-to-X (PtX) processes, which produce alternative fuels or chemicals from renewable energies. Due to the dynamic nature of renewable energy, these processes must be operated flexibly under changing process and boundary conditions. In doing so, they simultaneously convert electrical energy into chemical products that are easy to store and can be utilized later. To do this, a production schedule is generated that takes into account the most important process decisions, like the production capacity per time step over extended periods of several days [2]. This is widely used in many processes in the chemical industry [3, 4, 5, 6].

To optimize the production of chemical processes, a suitable process model is necessary to formulate the optimization problem and calculate an optimal schedule. Since the chemical processes under consideration are often very complex, making optimization with detailed white-box process models very costly and time-consuming, surrogate models are frequently used [7]. These have the advantage of being less computationally intensive and are therefore better suited for scheduling [8, 9]. There are many approaches to obtain surrogate models, including, for example, linear regression [10], radial basis functions [11], kriging [12], artificial neural networks [13], or linear-piecewise approximations [14]. In the field of chemical processes, several papers have been published in recent years that use surrogate models for optimization. To accelerate computation, linear regression and neural networks were explored as surrogate

*Corresponding author, E-mail address: max.kollmer@kit.edu

Journal Pre-proof

thermodynamic models for a chemical process [15]. Integrating these models into the optimization workflow reduced the time required to determine optimal operating conditions while maintaining prediction accuracy. The main focus, however, has been on the development of dynamic surrogate models, since the chemical processes need to be scheduled over time. Another study developed dynamic surrogate models of different orders using polynomial regression and tested these with an ethylene oxide production [16]. The computation time for scheduling can be reduced, while the prediction accuracy varies depending on the model order. The development of dynamic surrogate nonlinear autoregressive models with exogenous inputs is shown on different industrial processes [17]. The prediction of the process outputs shows high accuracy over long time horizons while reducing the computational effort.

Depending on the method used, typical surrogate models often cannot capture the complex process characteristics, as they are a simplified version of the considered process. Therefore, this approach uses static characteristic maps as surrogate models, as they are easy to obtain and adjustable, but still reflect the complexity of the process.

Furthermore, chemical processes often change their behavior over time due to aging, wear, or component replacement. As a result, the model no longer matches the real plant. This plant-model mismatch means that optimal production schedules can no longer be calculated. In the literature, methods like schedule recalculation [18], stochastic [19] or robust optimization [20] are used to account for uncertainties in the model. One drawback to that is that the error in the underlying surrogate model is not updated and is therefore not suitable for classical plant-model mismatch. Another possible approach is known as real-time optimization (RTO), where different methods exist in the literature [21]: model-parameter adaptation, modifier adaptation, and direct input adaptation. The first method is a two-step iterative method involving an update of the model followed by the optimization [22]. The latter is an approach in which the optimization problem is transformed into a feedback control problem, where a set of controlled variables has to be found, which leads to optimal plant performance [23]. In modifier adaptation, the cost function and the constraints are adjusted so that the Karush-Kuhn-Tucker conditions [24] of the model and the plant are consistent, which are the necessary conditions for optimality. The correction terms, which are calculated during the modifier adaptation, are added to the cost function and constraints of the optimization problem [25]. This is a widely known approach in RTO and has been used and studied for different applications [26, 27, 28, 29].

However, the RTO methods are primarily designed for short time scales of seconds or minutes and adapt the model parameters in a way that the plant optimum in the steady state is achieved [30], whereas the scheduling algorithm used computes a production schedule spanning several days and optimizes the plant's production over a longer horizon including strategic decisions like plant shutdown and storage levels. Furthermore, RTO often requires continuous first-principles models, whereas the used optimization model includes both continuous and discrete variables. These are necessary for discrete decisions like different plant states, for example, shutdown or standby. Since gradients of discrete variables are harder to obtain, this can only be achieved with considerable effort [31, 32]. But since the gradient is necessary for modifier adaptation [33], this approach is not suitable for this problem.

To overcome these challenges, a two-step model-parameter adaptation approach using characteristic maps as a surrogate process model is introduced. By using characteristic maps, the nonlinear relationships between independent and dependent variables are captured. In this case, the adaptation approach does not adapt the parameters of the model; rather, the entire surrogate model (the characteristic map) is simply adapted. This has the added advantage that an updated characteristic map can also be used for other tasks, such as process monitoring for fault detection and isolation. The adaptation step can detect plant-model mismatches in the form of sudden drift (for example, due to initial model mismatch) or incremental drift over time and adjust the entire characteristic map. The contributions of this paper are the following:

- A new framework for model adaptation is presented that uses characteristic maps to capture the nonlinear behavior of chemical processes.
- We reduce the need for numerous simulations or expert knowledge to construct the characteristic maps, since only basic process inputs, outputs, and limits are necessary.
- We design a global adaptation algorithm that dynamically updates the entire characteristic map using only a few measurements to ensure physically accurate models within a short operational time.
- The framework allows for the specification of selected optimal operating points for model adaptation, which are taken into account in the scheduling of the process.
- By decoupling the surrogate model adaptation from the scheduling model, the characteristic maps can be used for secondary use-cases like process fault detection and isolation.

The rest of the paper is structured as follows: Section 2 describes the development of the new framework, including the surrogate process model, the scheduling, and the adaptation. Section 3 describes the PtM process as the case study and the data generation. This also includes the considered scenarios for the process scheduling and adaptation. The results are shown in Section 4 before the conclusion and outlook in Section 5.

2. Methodology

This section introduces the framework for the scheduling of chemical processes using self-adapting characteristic maps. First, the concept of characteristic maps as a process model is introduced, followed by a short explanation of the scheduling model. Finally, the self-adaptation algorithm is explained.

Journal Pre-proof

2.1. Surrogate process model

The process model of the plant describes the input-output behavior of the process and is used to compute the optimized schedule for the process operation. Since the considered PtM process is highly nonlinear and complex due to the process structure, including the physical absorption as well as the chemical reaction and reaction kinetics, modeling with a white-box model approach would take a lot of time and knowledge of the process. In addition, the transferability to another process would not necessarily be given. Finally, for scheduling purposes, it is often a few output variables — rather than the internal states — that are of interest. For this reason, a surrogate process model with a black-box model approach is adopted.

To keep the model as simple as possible, dynamics within the process model are not taken into account directly in the model. Consequently, in our approach, only static black-box models of the process are used. By neglecting the dynamics, the model is simplified. This is possible, among other things, because chemical processes such as the PtM process have dynamics, but are often operated at a stationary operating point, and the dynamics must therefore only be considered during start-up, shutdown, and operating point changes. In this case, the dynamics of the process are not reflected by the model, but are still to be taken into account for use in scheduling via different kinds of constraints (e.g. ramping constraints between different operating points or transition constraints which describe the startup and shutdown of the plant).

Thus, for the scheduling, only the static input-output behavior is of interest, which can be approximated by a characteristic map for each output

$$y_i = \Phi(u) \approx M(u) \tag{1}$$

These characteristic maps always have one output variable (y_i) and can have any number of input variables ($\mathbf{u}$), depending on which variables influence the output variable. Thus, there is a characteristic map for each output variable. Each characteristic map in turn consists of a certain number of operating points $\mathbf{x}$, with $\mathcal{X}$ denoting the set of all operating points of this characteristic map. Although each input variable can theoretically be selected continuously between two limit values, it is assumed that each input variable can be discretized with sufficient accuracy. This means that each operating point is defined by a specific combination of input variables. Fig. 1 shows an example of a two-dimensional characteristic map. The nonlinear relationship is difficult to capture using a parametric function, which is why characteristic maps are well suited for this purpose.

2.2. Scheduling

Scheduling is used to calculate an optimal production schedule. Common objectives include, for example, minimizing costs or maximizing the production of the final product. In the context of demand response, it can also refer to grid-friendly operation, in which electricity consumption is reduced at certain times to avoid peak loads and stabilize the power grid.

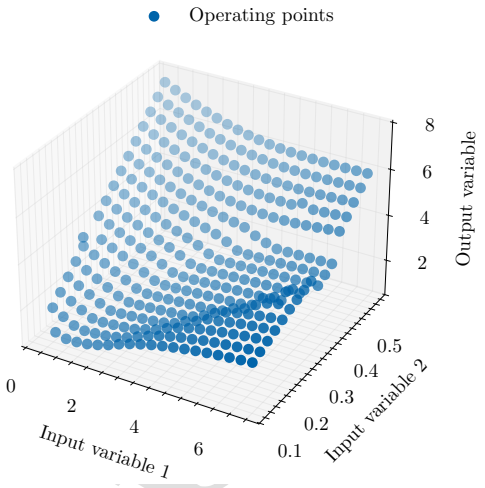


Figure 1: Example of a two-dimensional characteristic map. Each dot represents a single operating point for this specific output variable.

Generally, a schedule is not calculated just once, but repeatedly at regular intervals to ensure that the latest information on constraints — like storage tank levels, demand for the end product or electricity price forecasts — is always available. This recalculation usually takes place at fixed intervals. This makes it a model predictive controller (MPC), which iteratively calculates a schedule with new measurement data from the process.

The optimization model is formulated as a mixed-integer linear problem (MILP) in order to simultaneously consider both discrete parameters, such as operating states and operating points of the characteristic maps, and continuous variables, and is therefore a hybrid model. The general MILP formulation can be found in [34].

2.3. Global adaptation algorithm

The novel self-adapting algorithm presented in this paper has the goal to update the process model using online measurement data to maintain an accurate characteristic map over long-term process operation. The framework considers two types of process drifts introduced by [35], the **sudden drift** and the **incremental drift**. Other types like gradual or reoccurring drifts are omitted as they are either non-typical for chemical processes or can be modeled by multiple sudden drifts.

Sudden drift describes a fast change in process behavior where output variables change while inputs remain constant. Sudden drifts typically occur under two conditions, which are initial model inaccuracies or component modifications. Initial characteristic maps derived from simulation data or expert knowledge rarely match the real physical process due to modeling simplifications. Equally, component modifications like equipment replacement (e.g., replacing a reactor catalyst, an absorption medium, or moving to a pump with different efficiencies) instantly introduce new process characteristics. In this work, the initial characteristic maps are created by scaling

Journal Pre-proof

or modifying the ground truth characteristic map or by simulations with different physical boundary conditions.

Incremental drift captures slow, continuous variations in process over time due to, for example, aging, fouling, or catalyst deactivation. While a linear or constant change per time step is common, drift profiles over time can be highly nonlinear or accelerate. Because the physical degradation generally affects the complete process operating range, a drift observed at a single operating point implies that the entire characteristic map is affected. However, the exact drift behavior can vary across different regions of the operating range depending on input conditions. To account for that, the framework models multiple spatial drift distributions across operating points.

The adaptation is executed at discrete intervals during process operation. A single iteration k comprises a data acquisition phase followed by an adaptation phase (Fig. 2). In each iteration, the old characteristic map is updated with the new measurements. To do this, the residuals between the current characteristic map and the measured process values are calculated, and this residual characteristic map $\Delta_k^v(\mathbf{x})$ is then extended to all operating points, including those that were not measured. The new values of the characteristic map

$$M_{k+1}^v(\mathbf{x}) = M_k^v(\mathbf{x}) + \Delta_k^v(\mathbf{x}) + D_k^v(\mathbf{x}, k) \tag{2}$$

are then obtained by adding the residual characteristic map to the current characteristic map with an additional correction term $D_k^v(\mathbf{x}, k)$ to account for the incremental drift. The index v is the output variable, since a characteristic map exists for every considered output variable. The adaptation algorithm is structured in different steps. A schematic overview of the data flow of the global adaptation algorithm can be found in Fig. 3. After data acquisition and preprocessing, a sudden and incremental drift detection and handling step is performed. With this information, the data is prepared for the subsequent regression. The main step is the characteristic map update with the following active learning. The individual steps are further explained in the following on the example of a one-dimensional map.

2.3.1. *Data acquisition and preprocessing*

Data acquisition and preprocessing is the first step of the algorithm. The process is generally operated across different operating points $\mathbf{x}_i$, where all measurable data from the process sensors are recorded over time.

Using the raw measurements $y_i^v(\mathbf{x})$ from iteration k and process variable v, the operating points are detected via changes in the input variables, and the time series is partitioned accordingly. The number of operating points and the exact operating conditions depend on the schedule and the plant's operation. Second, the data within each segment is checked for stationarity, isolating the time steps t relevant for adapting the characteristic map. To reduce noise, the mean value of the $N_\mathbf{x}$ stationary measurements is calculated for each variable v and operating point $\mathbf{x}$ between the stationary boundaries $T_{\mathbf{x},k}^{\text{start}}$ and $T_{\mathbf{x},k}^{\text{end}}$:

$$\overline{y}_k^v(\mathbf{x}) = \frac{1}{N_\mathbf{x}} \sum_{t=T_{\mathbf{x},k}^{\text{start}}}^{T_{\mathbf{x},k}^{\text{end}}} y_{t,k}^v(\mathbf{x}) \tag{3}$$

To decouple the raw measurements from time, they are stored as a list for every operating point:

$$\mathbf{Y}_k^v(\mathbf{x}) = \{y_{1,k}^v(\mathbf{x}), \ldots, y_{N_\mathbf{x},k}^v(\mathbf{x})\} \tag{4}$$

Furthermore, the sample variance is computed using these raw measurements:

$$\sigma_k^2(\mathbf{x}) = \frac{1}{N_\mathbf{x} - 1} \sum_{t=T_{\mathbf{x},k}^{\text{start}}}^{T_{\mathbf{x},k}^{\text{end}}} \left(y_{t,k}^v(\mathbf{x}) - \overline{y}_k^v(\mathbf{x})\right)^2 \tag{5}$$

2.3.2. *Drift detection and handling*

The algorithm identifies and responds to two types of process drifts to maintain an accurate characteristic map. The following sections outline the operational rules, thresholds, and response mechanisms for both sudden and incremental drift detection. To prevent an incremental drift from being mistakenly detected as a sudden drift, both drift detectors are not active at the same

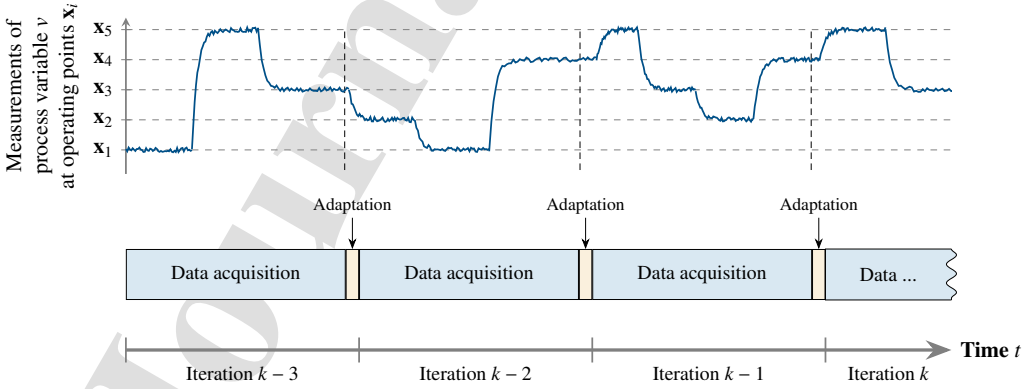


Figure 2: Schematic representation of the adaptation process with a fixed time between adaptations. The sizes of the blocks are not to scale. An adaptation typically takes significantly less time than the data acquisition period.

Journal Pre-proof

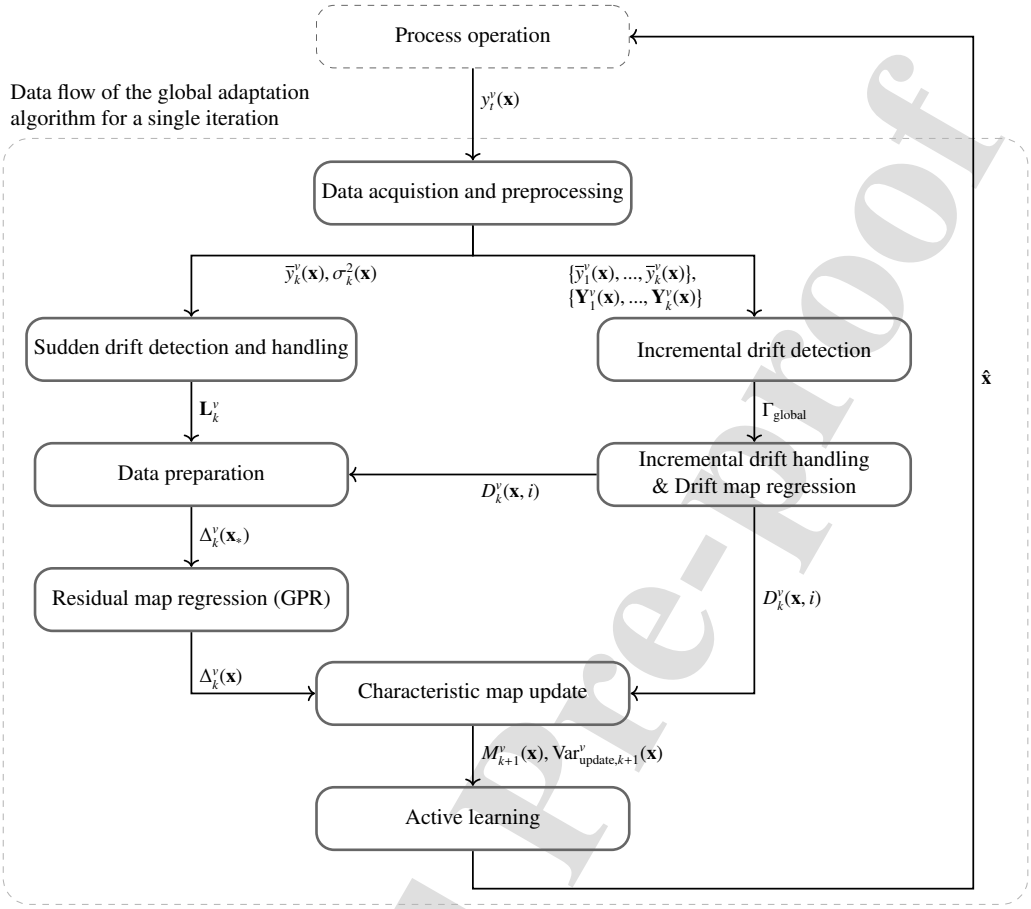


Figure 3: Schematic overview of the data flow of the developed global adaptation algorithm. The sudden drift detection and handling and the incremental drift detection and handling are processed in parallel.

time. The sudden drift detector is activated at the start of operation to adjust the initial characteristic map or after components have been replaced. The incremental drift detector is activated as soon as the system has been in operation and the sudden drift detector is deactivated.

Sudden Drift Detection and Handling

The algorithm performs three sequential checks to detect sudden drifts in process behavior and determine how to update the data set $\mathbf{L}_k^v$ used for the characteristic map update. An overview of this process is shown in Fig. 4.

Check 1: Map Consistency: The Mahalanobis distance quantifies the difference between measured values $\overline{y}_k^v(\mathbf{x})$ and the characteristic map $M_k^v(\mathbf{x})$ for all operating points $\mathbf{x} \in \mathcal{X}_k$:

$$D_{\text{M,map}} = \sqrt{\left(\overline{y}_k^v(\mathbf{x}) - M_k^v(\mathbf{x})\right)^\top \Sigma^{-1} \left(\overline{y}_k^v(\mathbf{x}) - M_k^v(\mathbf{x})\right)}, \quad (6)$$

where $\Sigma = \sigma_k^2(\mathbf{x}) + \text{Cov}(M^v(\mathbf{x}))$. For $k = 1$, $\text{Cov}(M^v(\mathbf{x}))$ is set manually; for subsequent iterations, it is derived from prior map updates. A χ^2-distribution determines the threshold $t_{\text{M,map}}$.

- If $D_{\text{M,map}} < t_{\text{M,map}}$ for all $\mathbf{x} \in \mathcal{X}_k$, measurements are consistent with the map, and $\mathbf{L}_k^v$ expands to include new data:

$$\mathbf{L}_k^v = \left\{\overline{y}_{i,x}^v \mid 1 \le i \le k, \mathbf{x} \in \mathcal{X}_i\right\}. \quad (7)$$

- If $D_{\text{M,map}} \ge t_{\text{M,map}}$ for any $\mathbf{x} \in \mathcal{X}_k$, the algorithm proceeds to Check 2.

Check 2: History: If no prior measurements exist for any $\mathbf{x} \in \mathcal{X}_k$, a sudden drift cannot be distinguished from map inaccuracy, resulting in the data set being unchanged:

$$\mathbf{L}_k^v = \mathbf{L}_{k-1}^v. \quad (8)$$

Check 3: Measurement Consistency This check compares current measurements with the most recent historical data for

Journal Pre-proof

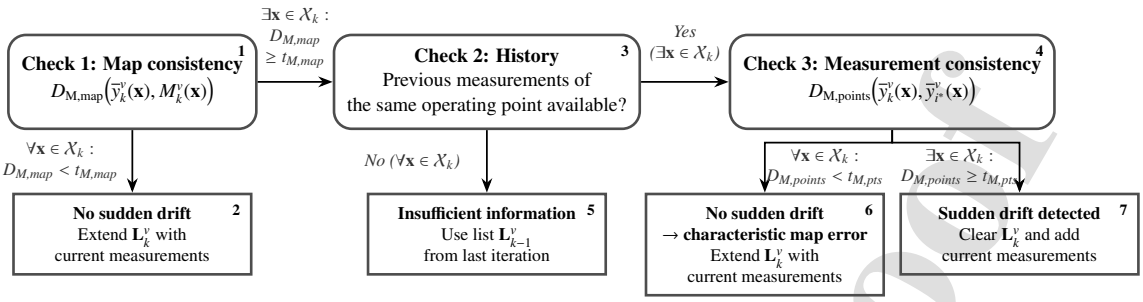


Figure 4: Schematic overview of the sudden drift detection and handling step. Three different checks have to be done to identify or rule out a sudden drift.

each $\mathbf{x} \in \mathcal{X}_k$:

$$D_{\mathrm{M,points}} = \sqrt{\left(\overline{y}_k^v(\mathbf{x}) - \overline{y}_{i^*}^v(\mathbf{x})\right)^\top \Sigma^{-1} \left(\overline{y}_k^v(\mathbf{x}) - \overline{y}_{i^*}^v(\mathbf{x})\right)}, \tag{9}$$

$$\tag{10}$$

$$i^*(\mathbf{x}) = \max\{i < k \mid \mathbf{x} \in \mathcal{X}_i\}. \tag{11}$$

- If $D_{\mathrm{M,points}} < t_{\mathrm{M,pts}}$ for all $\mathbf{x} \in \mathcal{X}_k$, no sudden drift is detected, and $\mathbf{L}_k^v$ is updated as in (7).
- If $D_{\mathrm{M,points}} \geq t_{\mathrm{M,pts}}$ for any $\mathbf{x} \in \mathcal{X}_k$, a sudden drift is confirmed; therefore, the data set is reset to include only current measurements:

$$\mathbf{L}_k^v = \{\overline{y}_k^v(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}_k\}. \tag{12}$$

The updated $\mathbf{L}_k^v$ is passed to the next step, where it is used for the map update.

Incremental drift detection and handling

Parallel to the detection of sudden drift, the algorithm also identifies incremental drifts. This is done using a Kolmogorov-Smirnov (KS) test with a significance level of α_{KS}. A schematic overview can be seen in Fig. 5.

For each operating point $\mathbf{x}$, the KS-test compares pairs of measurement distributions $\mathbf{Y}_i^v(\mathbf{x})$ and $\mathbf{Y}_j^v(\mathbf{x})$ $(i > j)$ to determine if they originate from the same underlying distribution. If at least ζ of these comparisons fail, the operating point is labeled as conspicuous, flagging a potential incremental drift.

The decision on whether an incremental drift is globally present is based on the number of conspicuous operating points:

$$\Gamma_{\mathrm{global}} = \begin{cases} \text{True} & \text{if } \sum_{\mathbf{x} \in \mathcal{X}} \gamma(\mathbf{x}) \geq \xi, \\ \text{False} & \text{otherwise,} \end{cases} \tag{13}$$

The larger the threshold ξ is set, the more operating points must be conspicuous for a drift to be detected. This makes the detection more robust; however, with a higher ξ, it also takes longer for a drift to be detected. The exact parameters can be found in Table A.5.

If an incremental drift is confirmed (Γ_{global} = True), a drift characteristic map $D_k^v(\mathbf{x}, i)$ is computed using Gaussian Process Regression. This map accounts for both spatial and temporal drift trends across the process operating domain.

For the calculation of the drift characteristic map $D_k^v(\mathbf{x}, i)$, the residuals between measurements from different iterations are normalized to reflect the drift per operating point. These residuals are used to calculate the drift at the unseen operating points using the GPR. The uncertainty of these residuals, α_{drift}, is calculated as:

$$\alpha_{\mathrm{drift}} = \sigma_i^2(\mathbf{x}) + \sigma_j^2(\mathbf{x}), \tag{14}$$

where i and j represent iterations being compared and $\sigma^2(\mathbf{x})$ the measurement noise. These uncertainties are combined into a noise matrix, $\Sigma_{k,\mathrm{drift}}^v(\mathbf{x}, i)$. The drift values are then calculated at unseen operating points, producing a complete drift characteristic map, $D_k^v(\mathbf{x}, i)$ at all operating points, with its associated variance.

Connection to map update

The detection results directly control how measurement data is incorporated into the map update procedure.

- For **sudden drifts**, $\mathbf{L}_k^v$ is either:
  - Expanded to include new measurements (Eq. 7), or
  - Reset to current measurements only (Eq. 12).
- For **incremental drifts**, if confirmed (Γ_{global} = True), the drift characteristic map $D_k^v(\mathbf{x}, i)$ is used to augment measurement data (see 2.3.3), ensuring all operating points reflect current process behavior.

This ensures the framework responds adaptively, balancing data retention with drift correction to maintain model accuracy.

2.3.3. Data preparation

The goal of this step is to calculate the residuals $\Delta_k^v(\mathbf{x}_*)$ between the measurements in the list $\mathbf{L}_k^v$ and the current characteristic map $M_k^v(\mathbf{x}_*)$ at the measured operating points $\mathbf{x}_*$. The residual map

$$\Delta_k^v(\mathbf{x}_*) = \overline{y}_{i^*}^v(\mathbf{x}_*) - M_k^v(\mathbf{x}_*) + \sum_{i=i^*}^{k} D_k^v(\mathbf{x}_*, i) \tag{15}$$

is calculated for all operating points $\mathbf{x}_*$ that have been measured and are in the list L_k^v. If an incremental drift is detected

Journal Pre-proof

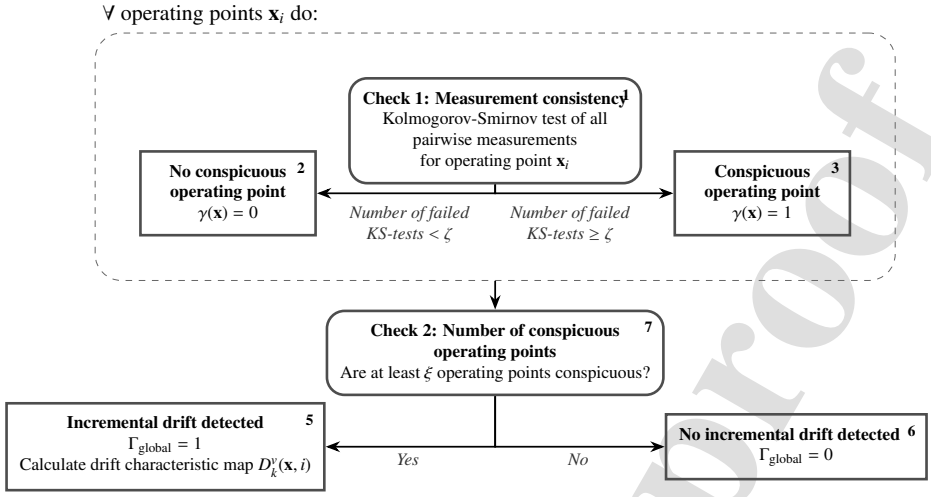


Figure 5: Schematic overview of the incremental drift detection and handling step. Two different checks have to be done to identify or rule out an incremental drift.

($\Gamma_{\text{global}} = 1$), the drift characteristic map $D_k^v(\mathbf{x}_*, i)$ is used to augment older measurements from the last observed iteration i^* to the current iteration k. Fig. 6 shows the calculation of the residual characteristic map for an exemplary characteristic map and some measurements. The characteristic map value of the respective operating point is subtracted from the measurements (left), which leads to the residual characteristic map $\Delta_k^v(\mathbf{x})$ (right).

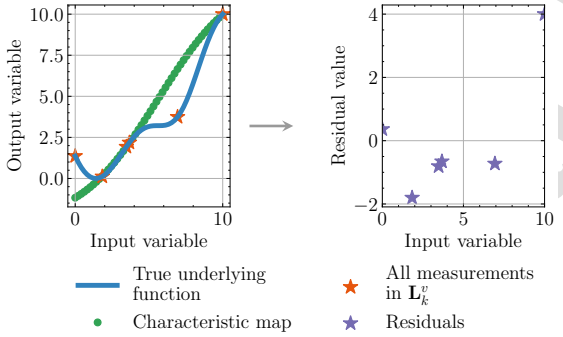


Figure 6: Left: Six measurements are used to calculate the residuals between the measurements $\bar{y}_k^v(\mathbf{x}_*)$ and the characteristic map $M_k^v(\mathbf{x}_*)$. Right: Residuals $\Delta_k^v(\mathbf{x}_*)$ at the operating points $\mathbf{x}_*$.

The measurement noise is used as the uncertainty of the residuals in the regression, whereas the uncertainty of the initial model characteristic map is used as a prefactor of the covariance function [36]. It is the uncertainty of the characteristic map without a new measurement. The initial uncertainty $\sigma_{\text{map},0}$ must be defined in advance and describes how certain the initial characteristic map $M_0^v(\mathbf{x})$ is.

The combined uncertainty α_{residual} for the residuals consists of the measurement noise $\sigma_k^2(\mathbf{x}_*)$ and, if available, the uncertainty of the drift prediction and an age factor $\text{Var}_{\text{age},k}(\mathbf{x}_*)$ which down-weighs older measurements. The different uncertainties per operating point are taken into account by defining the heteroscedastic noise matrix $\Sigma_{k,\text{residual}}^v(\mathbf{x}_*)$ for iteration k, with the diagonal elements

$$\alpha_{\text{residual}} = \sigma_k^2(\mathbf{x}_*) + \text{Var}_{\text{drift},k}(\mathbf{x}_*, k) + \text{Var}_{\text{age},k}(\mathbf{x})_* \tag{16}$$

for every measured operating point $\mathbf{x}_*$ in the list $\mathbf{L}_k^v$.

2.3.4. Data regression

The goal of the data regression step is the calculation of the residual characteristic map $\Delta_k^v(\mathbf{x})$ for all operating points from the residual characteristic map $\Delta_k^v(\mathbf{x}_*)$ at the measured operating points.

Because the functional relationship between the input and the output variables can be arbitrary, nonparametric regression models are used. A widely used method for this is Gaussian process regression [37]. It defines a distribution of possible functions and allows prediction of new points as well as an uncertainty measure for the output. In a first step, a prior has to be defined, which often assumes a mean of zero. Afterward, the model is updated with new measurements to obtain the posterior distribution. Since the Gaussian process regression assumes that the training data has generally a zero mean value [38] and the structure of the residuals is unknown, they are normalized beforehand to satisfy this assumption by subtracting the mean value of the residual map and dividing by the variance to obtain the normalized residual characteristic map $\Delta_{k,\text{norm}}^v(\mathbf{x}_*)$.

The training data set $\mathcal{D}_k = \left\{\left(\mathbf{x}_*, \Delta_{k,\text{norm}}^v(\mathbf{x}_*)\right)\right\}$ for the Gaussian process regression is constructed using the measured operating points $\mathbf{x}_*$ and the normalized residuals at these operating points. Applying the calculation of the mean and variance using

Journal Pre-proof

Gaussian process regression to our problem yields the following [37]:

$$\Delta_k^v(\mathbf{x}) = K(\mathbf{x}_*, \mathbf{x})[K(\mathbf{x}, \mathbf{x}) + \Sigma_k^v(\mathbf{x})]^{-1}\Delta_{k,\text{norm}}^v(\mathbf{x}_*) \tag{17}$$

$$\text{Var}_{\text{res},k}^v(\Delta_k^v(\mathbf{x})) = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, \mathbf{x})[K(\mathbf{x}, \mathbf{x}) + \Sigma_k^v(\mathbf{x})]^{-1}K(\mathbf{x}, \mathbf{x}_*) \tag{18}$$

$$\text{with } \mathbf{x}_* \in \mathcal{X},\ \mathbf{x} \in \bigcup_{i=1}^{k} P_i \mid \exists y \in \mathbf{L}_k^v : (y, i, \mathbf{x})$$

These equations describe how residuals are fitted at unmeasured operating points. The selection of a suitable kernel as a covariance function $K(\mathbf{x}, \mathbf{x})$ is often a difficult process that requires expert knowledge of how the data is related. To automate this process, an adaptive kernel selection strategy is developed, where different kernels are tested and validated. RBF, Matérn, constant, and linear covariance functions are tested as kernels, as well as combinations of all of them with different hyperparameters to account for superimposing local and global effects. The hyperparameters of the kernels are optimized using the log marginal likelihood. The exact configuration for the GPR can be found in Table A.5.

Various kernels are tested and validated using k-fold cross-validation. The best model is then determined based on several metrics:

- **Log marginal likelihood:** The log marginal likelihood (LML)

$$\log p(\mathbf{y}|X, \theta) = -\frac{1}{2}\mathbf{y}^\top K_y^{-1}\mathbf{y} - \frac{1}{2}\log|K_y| - \frac{n}{2}\log 2\pi \tag{19}$$

  evaluates how probable the entire distribution of data is under the calculated GPR model.

- **MRMSE of the cross validation:** The mean RMSE (MRMSE) through cross-validation is calculated as follows:

$$\text{MRMSE} = \frac{1}{k_{\text{cross-val}}}\sum_{i=1}^{k_{\text{cross-val}}}\sqrt{\left(\Delta_k^v(\mathbf{x}_*) - \left(\bar{y}_k^v(\mathbf{x}_*) - M_k^v(\mathbf{x}_*)\right)\right)^2} \tag{20}$$

  with $k_{\text{cross-val}}$ being the number of cross-validations.

- **Bayesian information criterion (BIC):** The BIC is a criterion widely used for model selection ([39]). It is defined as follows:

$$\text{BIC} = k_{\text{param}}\ln(n) - 2\ln(L) \tag{21}$$

  with L being the marignal likelihood, n the number of samples ($|\mathbf{L}_k^v(\mathbf{x})|$) and k_{param} the number of parameters θ_k. A small value for the BIC is preferred.

After calculating the metrics for all kernels, a vote is held. The kernel with the majority of the metrics having the best result is selected. If this is not the case, the kernel with the best BIC is used, as this metric is a compromise between model complexity and thus overfitting and prediction quality.

2.3.5. Characteristic map update

The map update computes the model characteristic map $M_{k+1}^v(\mathbf{x})$ for the next iteration with the residual map $\Delta_k^v(\mathbf{x})$. If a drift is present in the form of the drift characteristic map $D_k^v(\mathbf{x}, i)$, the update is calculated as follows:

$$M_{k+1}^v(\mathbf{x}) = M_k^v(\mathbf{x}) + \Delta_k^v(\mathbf{x}) + D_k^v(\mathbf{x}, k) \tag{22}$$

This updates all operating points — both measured and unmeasured — since the residual map is defined for all operating points. Simultaneously, the variance of the current map is also updated. Because the variance of the previous model characteristic map and the measurement uncertainties are directly incorporated into the GPR, this variance of the map $\text{Var}_{\text{map},k}^v$ must not be taken into account for iteration $k + 1$.

$$\text{Var}_{\text{map},k+1}^v(\mathbf{x}) = \text{Var}_{\text{residual},k}^v(\mathbf{x}) \tag{23}$$

Fig. 7 shows an exemplary update of the characteristic map $M_k^v(\mathbf{x})$ with the residuals $\Delta_k^v(\mathbf{x})$. After the characteristic map is updated, it is also checked for physical plausibility. For example, in the case of a mass flow or volume flow, all operating points with a value less than zero are set to zero. The same applies to the power output of individual components. For molar fractions, the values of the operating points are limited between zero and one.

2.3.6. Active learning

Active learning is the selection of new operating points $\hat{\mathbf{x}}$ for the next iteration $k + 1$. The aim is to ensure that the selected operating points in the future yield maximum information gain for the model adaptation. It can consist of two conflicting parts, namely local exploitation and global exploration [40, 41]. While local exploitation aims to find regions with large prediction errors in order to sample additional operating points, global exploration aims to investigate regions that have not yet been sampled. The following methods are considered:

- Coverage score (S_{dist}) [42]: This metric ensures that the measured operating points to calculate the residual characteristic map are evenly distributed. It prevents large unseen spaces of operating points and promotes the exploration of these spaces. For uniform coverage of the operating point set $\mathcal{X}$, the normalized minimum distance of each operating point outside the data set $\mathbf{L}_k^v$ is calculated, resulting in the score

$$S_{\text{dist}}(\mathbf{x}) = \frac{d_{\min}(\mathbf{x}) - \min(d_{\min}(\mathbf{x}))}{\max(d_{\min}(\mathbf{x})) - \min(d_{\min}(\mathbf{x}))} \tag{24}$$

  with $d_{\min}(\mathbf{x})$ being the minimum distance of an unseen operating point to the nearest already measured operating point. The minimum and maximum distances are used to normalize the score.

- Uncertainty sampling (S_{var}): The operating points that show the greatest variance in the current model are selected. This is equivalent to an entropy search strategy,

Journal Pre-proof

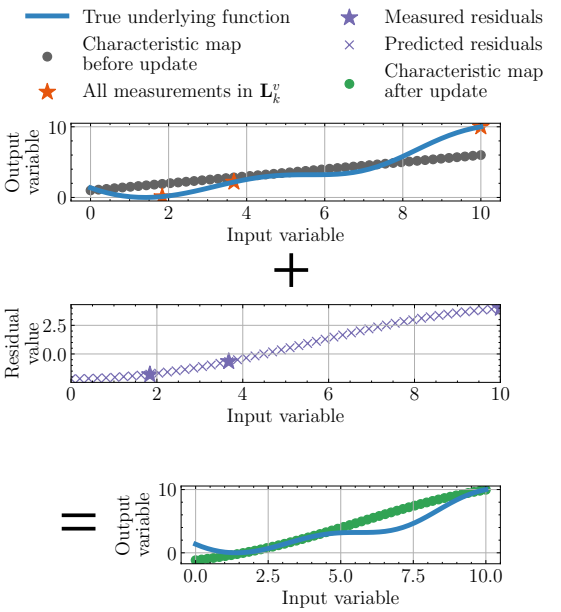


Figure 7: The current characteristic map before the update (top left) and the residual characteristic map (top right) are added together, resulting in the updated characteristic map (bottom). The true underlying function is shown as a reference.

as it is proportional to the logarithm of the variance [43]. Maximizing the variance scores results in the highest information gain and exploits the local prediction error. The score for the variance is calculated as

$$S_{\text{var}}(\mathbf{x}) = \frac{\text{Var}^v_{\text{map},k}(\mathbf{x}) - \min(\text{Var}^v_{\text{map},k}(\mathbf{x}))}{\max(\text{Var}^v_{\text{map},k}(\mathbf{x})) - \min(\text{Var}^v_{\text{map},k}(\mathbf{x}))} \tag{25}$$

using the variance per operating point of the characteristic map and its minimum and maximum value for normalization. This is used as uncertainty sampling and is equivalent to the maximum entropy design, since the natural logarithm is a strictly monotonically increasing function [44].

The final score

$$S_{\text{total}}(\mathbf{x}) = w_1 \cdot S_{\text{var}}(\mathbf{x}) + w_2 \cdot S_{\text{dist}}(\mathbf{x}) \tag{26}$$

is then calculated as a weighted sum, which places more emphasis on exploitation or exploration depending on the weighting factors w_1 and w_2. The actual values of the parameters result from manual tuning and can be found in Table A.5 in the appendix.

If the N best points get selected for the next iteration, this can lead to spatial redundancy, as points with a high variance are often close to each other. For this reason, the combined score S_{combined} is introduced, which is calculated dynamically every time after a point has been selected. Each point $\hat{\mathbf{x}}$ is determined iteratively for $i \in \{i, ..., N\}$ by maximizing the dynamic objective function:

$$\hat{\mathbf{x}}_i = \arg\max_{\mathbf{x}\in\mathcal{X}_{\text{rem}}} [\gamma S_{\text{total}}(\mathbf{x}) + (1-\gamma) S_{\text{diversity}}(\mathbf{x}, \mathcal{X}_{\text{sel}})] \tag{27}$$

Here, $S_{\text{combined}} = \gamma S_{\text{total}}(\mathbf{x}) + (1-\gamma) S_{\text{diversity}}(\mathbf{x}, \mathcal{X}_{\text{sel}})$ is the combined score, constructed with the score S_{total} and $S_{\text{diversity}}$. $\mathcal{X}_{\text{rem}}$ are the remaining operating points not in the data set $\mathbf{L}_k^v$ ($\mathcal{X} \setminus \sum_k \mathcal{X}_k$) and is therefore the candiate set. $\mathcal{X}_{\text{sel}}$ are the selected operating points during the active learning. $S_{\text{diversity}}(\mathbf{x}, \mathcal{X}_{sel}) = \min_{\mathbf{x}'\in\mathcal{X}_{sel}} \|\mathbf{x} - \mathbf{x}'\|_2$ is the distance of the operating point $\mathbf{x}$ to the nearest point in the already selected points for active learning. $\gamma \in [0, 1]$ is the parameter to control the trade-off between the usability for improving the model performance and the batch diversity. A two-step procedure is used to determine whether the proposed operating points will be used in the scheduling. First, the system checks whether an operating point would violate a production constraint; if so, that operating point is removed. In a second step, the scheduler has the option to skip individual proposed operating points if using all proposed points does not result in a valid solution, for example due to ramping constraints. However, this leads to an increase in costs, so it is only done if the schedule would otherwise be infeasible.

3. Case Study: PtM process

This section introduces the considered PtM process. Afterward, the generation of the characteristic maps is described, as well as the considered case studies for the scheduling and adaptation of the characteristic maps. The considered Power-to-methanol process is depicted in Fig. 8.

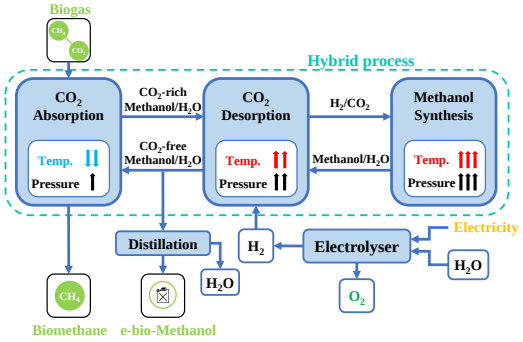


Figure 8: Block flow diagram of the novel Power-to-methanol process [45].

The process itself consists of three main units: The carbon dioxide capture unit (consisting of CO_2 absorption and desorption), the methanol synthesis unit and the distillation unit [46]. In addition, an electrolyser is also included. Biogas containing CO_2 enters the absorption column, where the carbon dioxide in the biogas is absorbed using methanol-water as the absorption

Journal Pre-proof

medium at a low temperature and pressure. The cleaned biogas (mostly CH_4) flows out of the top of the absorption column and can be used for other purposes, such as a substitute for natural gas. The CO_2-rich methanol-water mixture is heated and flows to the desorption column, where the CO_2 is stripped using hydrogen gas at higher temperature and pressure. The hydrogen is produced in an anion exchange membrane electrolyser (AEM). After the desorption, the lean methanol-water mixture flows back into the absorption column and thus forms a cycle, while the stripped CO_2 mixes with the hydrogen to yield synthesis gas. Over 99 % of the CO_2 is absorbed and then desorbed under optimum operating conditions, resulting in a high efficiency of the CO_2 capture. The average energy consumption is around 10 kW, primarily for heating of the desorption and cooling of the absorption. The CO_2 yield is approximately 72 kg d^{-1}.

The synthesis gas is heated, compressed and passes through four consecutive plug flow reactors, where H_2 and CO_2 react to form a methanol-water mixture. The overall CO_2 conversion rate is about 80 %, yielding roughly 77 L d^{-1}.

Some of the resulting methanol-water mixture is returned to the CO_2 capture to maintain a consistently fresh absorption medium. Most of it is stored and later used in the distillation, where the methanol is separated from the water to yield highly pure green methanol. The purity is roughly 97 % with a yield of about 52 L d^{-1}. The total power demand of the process is about 40 kW, whereby most of it is accounted for the electrolyser with nearly 20 kW, followed by the CO_2 capture and the distillation unit, while the methanol synthesis unit requires the least power. The actual pilot plant can be seen in Fig. 9.

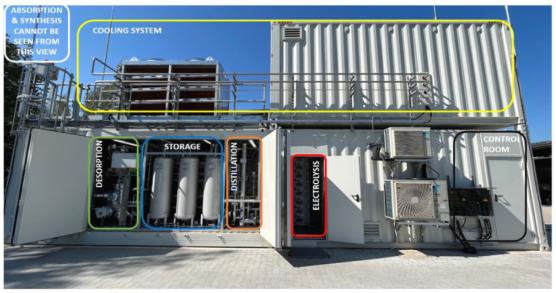


Figure 9: Container-sized pilot plant of the novel Power-to-methanol process.

The characteristic maps for the subprocesses are computed with the Aspen Plus process engineering software [47] for chemical processes. The exact structure of the simulation model can be found in [48]. Since the process design couples CO_2 capture and synthesis, these subprocesses are always considered together in the following. Distillation is not taken into account, as it is only operated at a steady state at a single operating point. The characteristic maps are determined using steady-state simulations at various operating points. The operating points are determined by various values of the input variables, which, for the process under consideration, are the mass flow rates of biogas and hydrogen. The range of values and the step size of the input variables are shown in Tab. 1. The total number of operating points considered is therefore 340.

Table 1: Step size and ranges of the considered input variables.

Input variable	Step size	Range
Biogas	0.35 kg h^{-1}	0.35 - 6.91 kg h^{-1}
Hydrogen	0.027 kg h^{-1}	0.11 - 0.55 kg h^{-1}

The output variables that play a role in scheduling are the power consumption of the subprocesses, the amount of methanol-water produced after synthesis, the ratio of hydrogen to carbon dioxide before synthesis, and the proportion of carbon dioxide in the purified biogas. A two-dimensional characteristic map is thus generated for each output variable. Fig. 10 shows an example of the characteristic map for the amount of methanol-water produced as a function of the inflowing biogas and hydrogen.

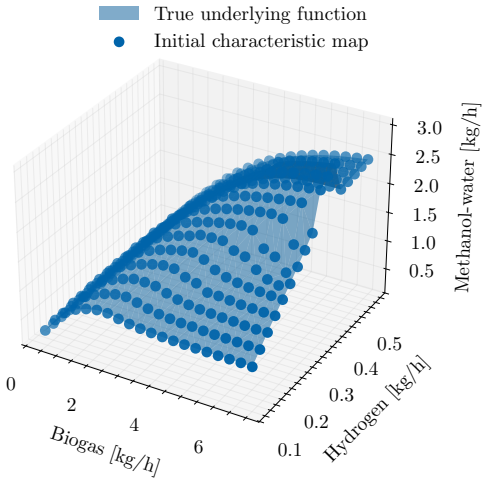


Figure 10: Example of the true underlying function and the characteristic map of the methanol-water production, depending on the inflowing biogas and hydrogen, generated with steady-state simulations.

As described in the algorithm above, there are two types of drift that must be taken into account: sudden and incremental. To test for sudden drift, various initial characteristic maps are created that differ from the true process characteristic map. Tab. 2 shows the types of initial characteristic maps considered. Cases S1 and S2 assume wrong model assumptions in form of a constant offset or little to no knowledge of the process. These have no direct physical source and are used as a benchmark to test how well the adaptation works for characteristic maps with little or no process knowledge or wrong model assumptions. Cases S3 and S4 are created using a modified simulation model, assuming a different catalyst activity. These might

Journal Pre-proof

Table 2: Different sudden and incremental drifts considered to test the adaptation of the process behavior.

Case #	Physical source	Affected outputs	Range of deviation $d(\mathbf{x}) = M_k^v(\mathbf{x}) - P_k^v(\mathbf{x})$	Description
Cases for sudden drift				
S1	-	All output variables	$d(\mathbf{x}) = a \cdot \overline{P}_k^v$, $a = -0.8..0.8$	Wrong model assumptions
S2	-	All output variables	$d(\mathbf{x}) = P_k^v(\mathbf{x}) - \overline{P}_k^v$	No knowledge of the process
S3	Catalyst activity	Mass flow of methanol-water Power of the synthesis	$d(\mathbf{x}) \in [-1.53, 0.01]$ kg h^{-1} $d(\mathbf{x}) \in [-0.02, 2.39]$ kW	Catalyst activity assumed too low
S4	Catalyst activity	Mass flow of methanol-water Power of the synthesis	$d(\mathbf{x}) \in [-0.02, 0.76]$ kg h^{-1} $d(\mathbf{x}) \in [-1.33, 0.02]$ kW	Catalyst activity assumed too high
Cases for incremental drift				
I1	Catalyst activity	Mass flow of methanol-water	$d(\mathbf{x}) = -0.02$ kg h^{-1} per time step	Constant reduction in activity over time
I2	Catalyst activity	Mass flow of methanol-water	$d(\mathbf{x}) \in [-0.027, 0]$ kg h^{-1} per time step	Linear reduction in activity over time
I3	Catalyst activity	Mass flow of methanol-water	$d(\mathbf{x}) \in [-0.03, 0]$ kg h^{-1} per time step	Reduction depending on simulation

be initial characteristic maps, for example, when there are differences between the physical assumptions of the simulation model and the actual process plant. Fig. 11 illustrates the initial characteristic map and the ground truth characteristic map for two exemplary cases (Case S1 and Case S3).

The following assumptions were made for the adaptation of the initial characteristic map:

- The overall scheduling and adaptation horizon is 21 days.
- The characteristic maps are adapted and the schedule is recalculated every 24 h.
- During the first week, active learning is used to achieve the best possible adaptation of the characteristic map. To achieve this, the schedule optimizer is provided with up to three mandatory operating points that must be reached in the upcoming 24 h. It then independently schedules these points and defines any remaining operating points to calculate an optimal schedule. For the last two weeks, no active learning is used so that the optimizer can select the best possible operating points without any restrictions.
- Over the last two weeks, additional constraints have been introduced to alter the process's operating range. The goal is to ensure that a better-adapted characteristic map performs well even at previously unknown operating points, whereas a poorly adapted characteristic map does not. These are all process-related adjustments and are shown in Tab. 3.

To test the framework when an incremental drift is present, the three cases I1 - I3 in Tab. 2 are used. It is assumed that only the amount of methanol-water produced decreases, which is primarily due to a reduction in catalyst activity. For case I1, the output decreases by the same amount at every operating point, whereas for I2, it scales linearly depending on the operating point. In I3, the characteristic map gradually changes to one generated using a modified simulation model with reduced catalyst activity (see Case S3). For the sake of computational efficiency, the reduction in catalyst activity is considered here over a period of a few weeks, even though this process normally takes one to two years. The overall scheduling and adaptation horizon for the incremental drift cases is 49 d with an adaptation and schedule recalculation every 24 h.

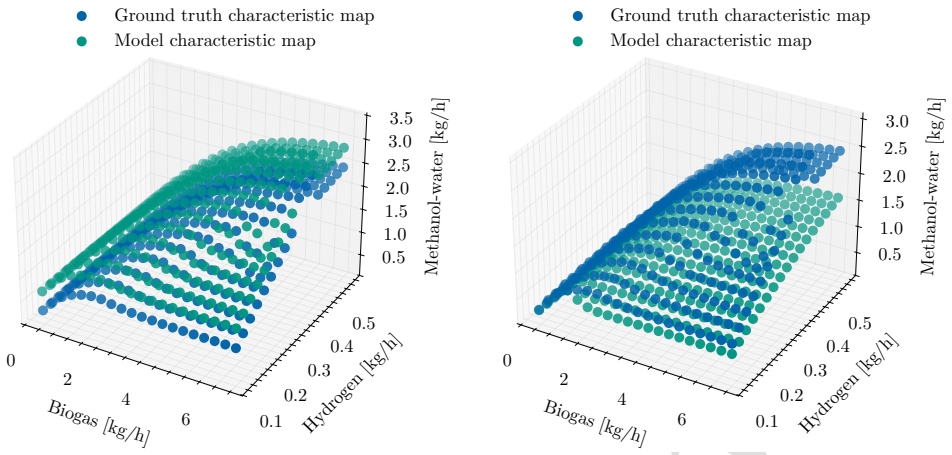
Table 3: Changes in process conditions after one week of process operation.

Process condition change	Description
PCC1	No process condition change.
PCC2	Increase in the average hydrogen-to-carbon dioxide ratio of the synthesis gas from 3 to 5.
PCC3	Change in the production capacity.

Fig. 12 shows the overall schematic overview of the scheduling with the adaptation of the characteristic maps. For all cases, the model characteristic map is used for scheduling, while the ground truth characteristic map is used for the subsequent emulation of the process operation. As a benchmark for the developed global adaptation algorithm, a local adaptation is used where the values of the characteristic map $M_k^v(\mathbf{x})$ at each measured operating point of iteration k are simply replaced by the new value.

$$M_{k+1}^v(\mathbf{x}) = \overline{y}_k^v(\mathbf{x}) \tag{28}$$

For local adaptation, no drift detection is done, and only the measured operating points of the characteristic map are updated, while all other operating points are not corrected. This



(a) Case S1: Constant offset.

(b) Case S3: Modified simulation model (Reduced catalyst activity).

Figure 11: Initial model and ground truth characteristic maps for different cases.

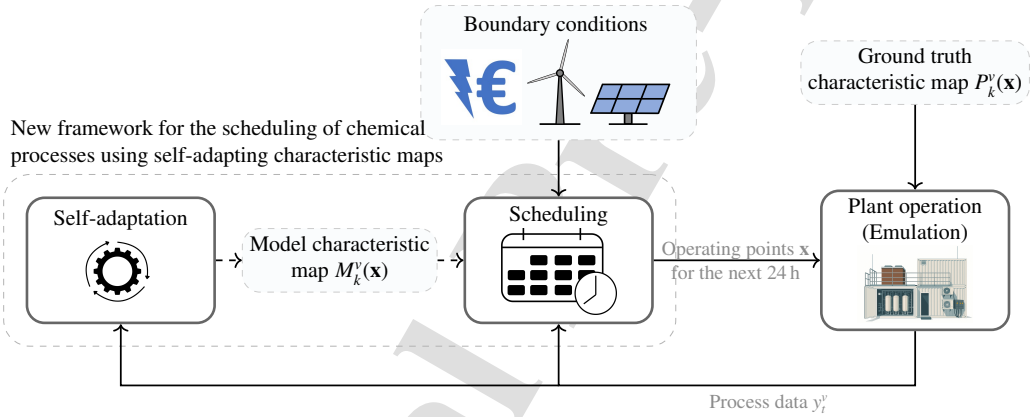


Figure 12: Schematic overview of the new framework

increases the probability that the characteristic map will eventually consist of both current and old data, unless all operating points are measured in each iteration. Since this is not possible or practical, especially with a large number of operating points, the characteristic map ends up being mixed with old and current data.

The optimization model for the scheduling of the PtM process consists of a PtM plant, a photovoltaic system and a wind turbine to model the power supply, a battery for energy storage and an electrolyser. The exact formulation as well as the constraints can be found in [48].

The following assumptions are made for the scheduling:

- The power is supplied by the grid. Power supply from the renewable energy sources and the battery is included in the scheduling model, but is not used in this example of the characteristic map adaptation.
- The horizon of each schedule is set to 84 h with a one-hour time step resolution as a compromise between the forecast horizon and the computation time.
- The MILP is solved using Gurobi [49] with a 1 % optimality gap or a maximum solution time of 1000 s.
- The storage tank fill levels have to be equal after every 7 d so that the boundary conditions between the different adaptation strategies are the same every week.
- A specified amount of methanol-water (50 % plant capacity) has to be produced.
- The optimization function is the minimization of the electricity costs. Historical dynamic electricity prices of the

Journal Pre-proof

year 2024 of the German energy market are used [50].

- The schedule is used to calculate the actual process outputs using a plant emulation (see Fig. 12).

4. Results and Discussion

The following section presents and discusses the results, starting with the results of the scheduling in combination with the self-adaptation algorithm when considering a sudden drift. Next, the results considering an incremental drift of the process behavior are shown.

The results are evaluated on three different metrics. The first is the RMSE between the values of the respective output variable for the operating points of the characteristic map. It thus describes how well the model characteristic map matches the ground truth characteristic map at different iterations. Furthermore, both the production price relative to the production volume (price per liter of methanol-water) and the production amount of methanol-water are used to evaluate the results. A comparison of prices per liter of methanol-water only makes sense if the same amount of methanol-water is produced. Therefore, the price per liter is adjusted using a factor, so that the corrected price (C1)

$$P_{\mathrm{C1}} = \frac{P}{\frac{m}{m_{\mathrm{gt}}}} \tag{29}$$

is calculated based on the amount produced for the adaptation method under consideration (m) and the ground truth (m_{gt}).

Finally, constraint violations in the calculated schedule are used to validate the results. These include, for example, the molar ratio of H_2 to CO_2 in the synthesis gas, which must maintain a specific average ratio $T_{\overline{\mathrm{H_2-CO_2}}}$ and must exceed a specified threshold $T_{\mathrm{H_2-CO_2}}$ for each time step.

The various cases (Tab. 2) and the boundary conditions (Tab. 3) are compared with one another. As adaptation methods, the developed global adaptation is used, as well as local adaptation as a benchmark. Additionally, these results are compared with the ground truth (always a correct process model) and with no adaptation (the model is not adjusted at all).

4.1. Results of adaptation to sudden drift

For the adaptation of the characteristic map to sudden drift, all relevant output variables described in Section 3 are considered in the cases S1 - S4. While in cases S1 and S2 a random value is selected for the offset or the constant value for each output variable, in cases S3 and S4 the operating point values for the characteristic maps are specified by the modified Aspen Plus simulation.

Evaluation of model adaptation

The first metric examined is the RMSE between the model characteristic map and the ground truth characteristic map. Fig. 13 shows the RMSE for the various adaptation strategies for the individual cases S1 – S4 for the output variable of the volume flow of produced methanol-water for the process condition change PCC1 after 21 d. It is obvious that the RMSE between the model characteristic map and the ground truth characteristic map is, in most cases, lower when using the global adaptation algorithm, demonstrating the advantage of the proposed approach. While it performs really well for constant offsets and constant initial maps (cases S1 and S2) as well as for the modified simulation case S3, the performance seems not as good for the modified simulation case S4, where the RMSE using the local adaptation strategy is slightly lower. The residual function between measurements and the model characteristic map follows a simple function for constant offsets; therefore, it can be easily calculated. As a result, the global adaptation is very fast and effective. For the slightly worse RMSE in case S4, the reason for this lies in the characteristic map of the variable under consideration. The values of the operating points between the initial characteristic map and the ground truth characteristic map are very similar for the majority of operating points and differ only within a specific range. In this range, however, few operating points were actually reached; instead, the system operated only in the surrounding operating points, where the error was already low to begin with. As a result of global adaptation, the characteristic map in this range deteriorated slightly because no measurements were available in this range, whereas the characteristic map in this range remained unchanged with local adaptation. By specifically targeting operating points — for example, through extended active learning — the error resulting from global adaptation can be significantly reduced. It appears that selecting the correct operating points for the adaptation is essential for reducing the error between the model and the ground truth characteristic map.

This can further be confirmed when looking at the RMSE of every operating point. Fig. 14 shows the RMSE for each operating point for case S4 for the local and global adaptation algorithms after 21 days. The red dots indicate the operating points reached, while the background color indicates the RMSE. While local adaptation only adjusts the operating points that are actually reached — resulting in a low RMSE at those points — global adaptation adjusts the entire characteristic map, thereby reducing the RMSE across the entire range. Unfortunately, the active learning did not cover the area (marked with a green rectangle) where the difference between the model characteristic map and the ground truth characteristic map is the highest, resulting in the slightly lower RMSE when using the local adaptation. It should also be noted that the RMSE is already very low at many operating points, so global adaptation offers little benefit in this case.

In addition to the RMSE at the end of the 21 d, the RMSE over the iterations is also crucial, as it is a measure of how fast the characteristic map is adapted. Fig. 15 shows the RMSE over the iterations for all four cases S1 - S4 for the output variable of the volume flow of methanol-water. The solid lines represent the RMSE of all operating points, while the dashed lines represent the RMSE of the operating points used in each iteration. For all cases, the RMSE for no adaptation remains, of course, constant, since the characteristic map is not adapted at

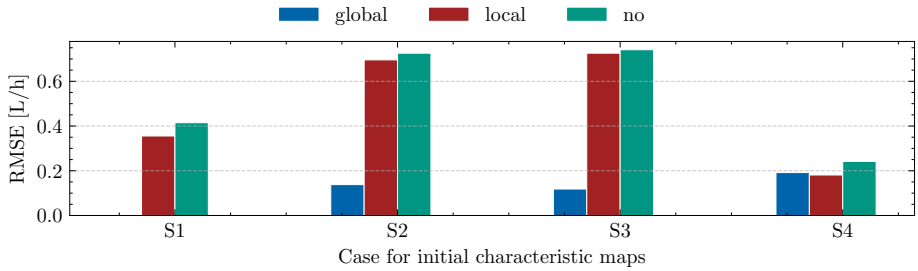
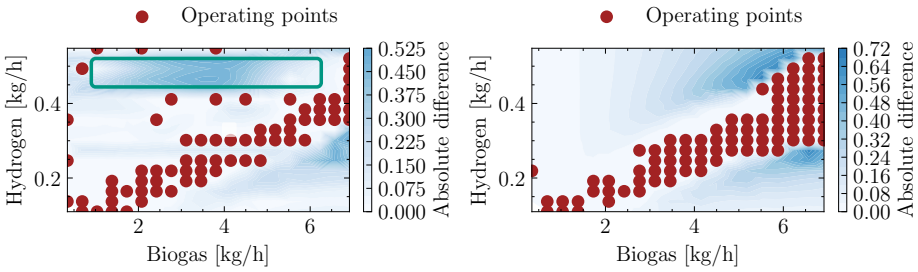


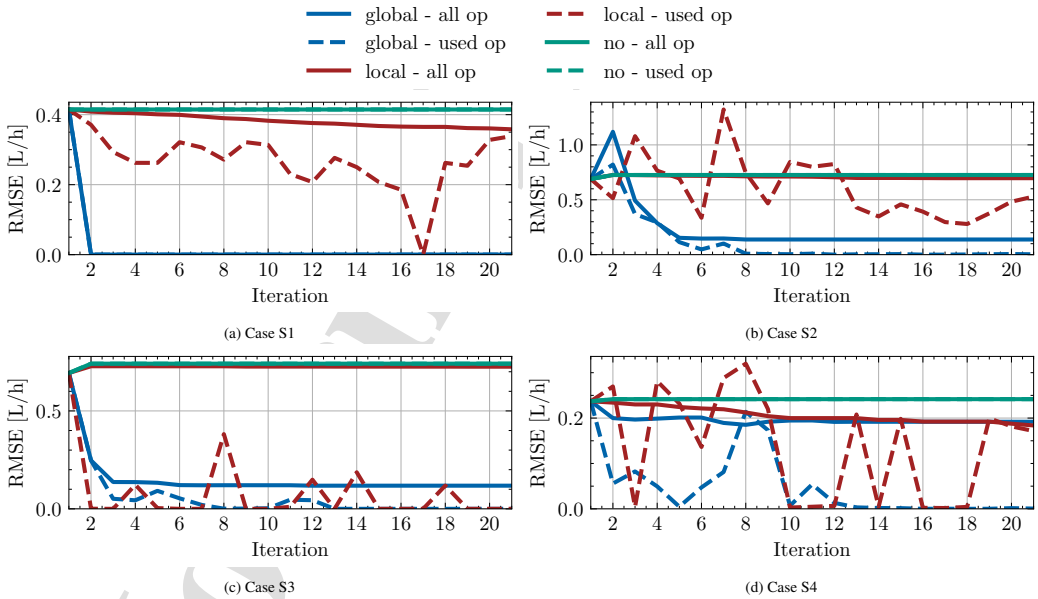
Figure 13: RMSE between the model characteristic map and the ground truth characteristic map for all cases after 21 d for boundary condition PCC1 using different adaptation methods.



(a) Case S4 - global adaptation

(b) Case S4 - local adaptation

Figure 14: RMSE per operating point between the model characteristic map and the ground truth characteristic map for the final characteristic map for different adaptation methods for case S4.



(a) Case S1

(b) Case S2

(c) Case S3

(d) Case S4

Figure 15: RMSE between the model characteristic map and the ground truth characteristic map over the iterations for different adaptation methods.

Journal Pre-proof

all. For Case S1, the map is perfectly adapted after one iteration, since the initial map has only an offset. For Case S2, the RMSE rises briefly in the first iteration, which is due to the constant initial map. All operating points are "equivalent" for the schedule calculation regarding the production amount, and the optimizer therefore uses only one. Consequently, after the first iteration, the map is constantly adapted with global adaptation. With additional operating points in the subsequent iterations, the RMSE for global adaptation decreases significantly, while that for local adaptation remains nearly constant and the RMSE of the points used oscillates strongly. In the cases S3 and S4, it is clear that the RMSE across all operating points for global adaptation is lower than that for local adaptation in almost all iterations. This is even more pronounced when looking at the RMSE of the operating points used.

Fig. 16 shows the RMSE between the ground truth characteristic map and the model characteristic map after the end of the three weeks for the various adaptation methods and the different variables for the process condition change PPC1. The mean RMSE was calculated across the cases S1 - S4. The RMSE is lowest for global adaptation across all variables considered. While it is between 76.7 % and 97.6 % lower compared to no adaptation, the local RMSE is only between 5 % and 12.7 % lower. This clearly demonstrates the improved adaptation of the characteristic maps and the reduction in plant-model mismatch achieved through global adaptation. The RMSE results for process condition changes PCC2 and PCC3 are similar to the results for PCC1. Therefore, these are not listed or discussed separately here.

Accurate and rapid adjustment of the characteristic map is particularly important to ensure that it remains consistent with the actual process and that the model characteristic map can be reliably used for tasks such as production scheduling and monitoring. In both cases, a correct characteristic map is indispensable. During monitoring, current measured values of the process variables are compared with the model characteristic map values. If there is a discrepancy, this indicates an error in the process. This can only be done reliably if the model map is as accurate as possible. At the same time, accurate production planning that complies with various process conditions requires a correct model of the process.

Evaluation of scheduling results

First, the amounts of methanol-water produced are compared. Fig. 17 shows this in a box plot for the produced amount of methanol-water in the cases S1 - S4, with the x-axis representing the different adaptation strategies for each of the three process condition changes. It is interesting to note here that only the global adaptation method was able to produce nearly the specified amount in every case, while the other methods exhibit significantly greater variations depending on the case. Thus, only the developed method ensures that the specified amount of methanol-water is always produced. This is advantageous because producing a smaller quantity means the production target is not met. It would also be problematic if, for example, the downstream storage buffer tank were to run empty, thereby preventing subsequent processes from operating optimally.

Fig. 18 depicts the mean difference in methanol-water over all cases for the different process condition changes between the ground truth and the adaptation strategies. It is clear that the difference is greatest for no adaptation, followed by the local adaptation strategy. Using the global adaptation strategy, the difference is almost negligible.

To account for the different amounts of methanol-water produced, the corrected costs using a correction factor (C1) are used. Fig. 19 shows exemplarily the corrected costs (C1) for the Cases S1 - S4 with process condition change PCC3. The costs per produced liter of methanol-water over the whole three weeks are, of course, lowest for the ground truth model, since it is error-free. In contrast, the non-adapted model performs significantly worse in all cases. This is straightforward, since with no adaptation, the process model used for the scheduling is always incorrect. However, the difference in the costs depends heavily on the specific case. While the price for no adaptation compared to global adaptation is only slightly higher for cases S1 and S4, it is already significantly higher for S3 and highest for S2. The reason for this is that, in case S2, the initial characteristic maps consist only of a constant value, and thus the difference from the actual characteristic map is the greatest. In

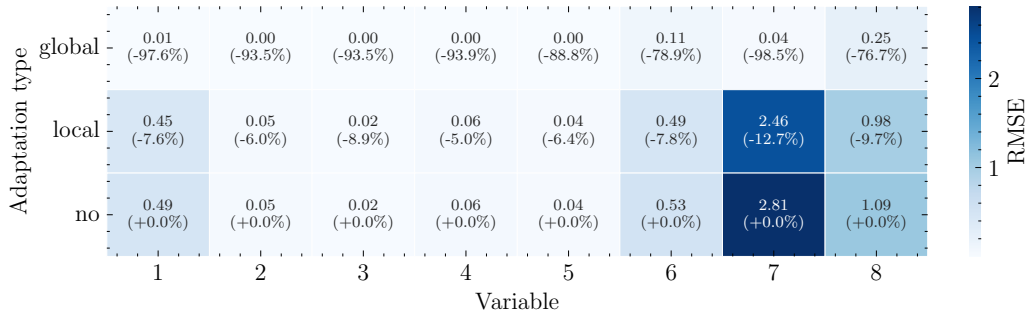
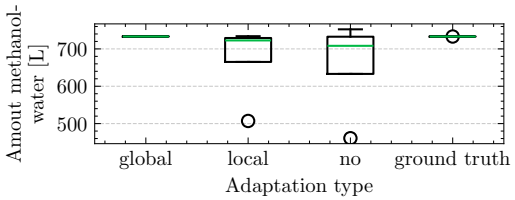
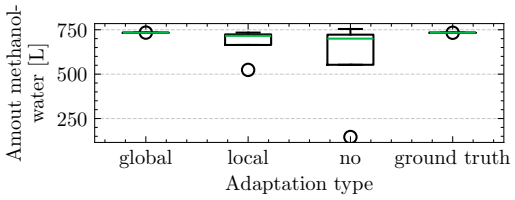


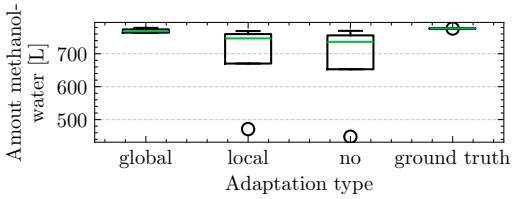
Figure 16: RMSE between the ground truth characteristic map and the model characteristic map for all variables and adaptation types for process condition change PPC1.



(a) Process condition change PCC1



(b) Process condition change PCC2



(c) Process condition change PCC3

Figure 17: Produced amount of methanol-water for different process conditions and cases using different adaptation methods.

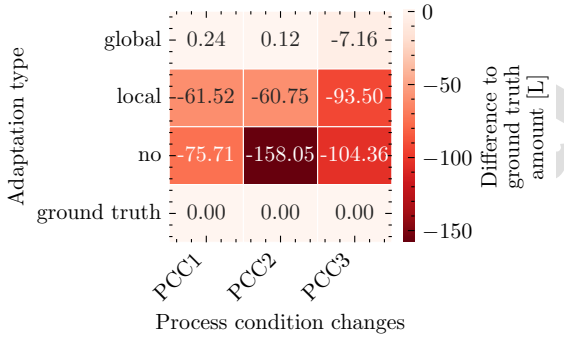


Figure 18: Mean difference between the produced methanol-water and the ground-truth amount over all cases for different process condition changes using different adaptation methods.

cases S1 and S4, the difference between the initial characteristic map and the actual characteristic map is smaller, which also results in a smaller difference in costs per liter. Comparing this with the RMSE between the model and the ground-truth characteristic map per case (Fig. 13), a correlation emerges between the corrected price and the RMSE.

The costs when using the global adaptation algorithm are lower than using the local adaptation for cases S1, S2 and S3 and slightly worse for case S4. The reason for the higher costs in case S4 is, on the one hand, the small difference between the initial characteristic map and the ground truth characteristic map, and, on the other hand, the active learning employed during the first seven days. As a result, suboptimal operating points are targeted to improve model adaptation, at the cost of higher operating costs. In summary, it is clear that operating costs depend heavily on the model mismatch between the initial and ground truth characteristic maps. If the characteristic maps are similar, this has virtually no effect on costs. However, if, as in case S2, a constant characteristic map is simply assumed, the global adaptation is significantly better. This is also a major advantage of the algorithm, as it achieves good characteristic map adaptation and low costs even with extremely little prior knowledge.

Under normal circumstances, it is of course unclear to what extent the initial characteristic map differs from the true characteristic map. For this reason, the following analysis always uses the average value of the cases, which constitute a small but representative sample of possible initial characteristic maps. In addition, the cost per liter of methanol-water is compared for the entire three weeks, as well as for the first week and the last two weeks. This allows us to examine the extent to which active learning and a change in the process conditions influence the costs.

To illustrate the costs associated with the other process condition changes (PCC1 - PCC3), Fig. 20 depicts the mean corrected costs (C1) over all cases for the different process condition changes and weeks. The darker the blue field, the higher the costs. The following points stand out:

- Over the whole time period, the costs using the global adaptation are slightly lower compared to the local adaptation for all process condition changes and even lower compared to no adaptation, while they are slightly higher compared to the ground truth. The cost savings for the adjusted price (C1) range from 11 pp to 18 pp compared to local adaptation and from 21 pp to 42 pp compared to no adaptation, depending on the process condition change.

- Although active learning is used in the first week, the corrected costs for global adaptation are significantly lower than those for local adaptation (12 pp). This suggests that significantly less methanol-water was produced during local adaptation in the first 7 d. It is also interesting to note that the no-adaptation strategy in the first week outperforms local adaptation. One reason for this could be that, due to local adaptation and the adjustment of only individual operating points, the optimizer remains stuck in local minima, since the characteristic map consists of both current and old values.

- After the first week, the scheduling can choose the optimal operating points over the whole horizon without active learning. There is still adaptation happening in the global adaptation, but no operating points are suggested using active learning. The cost savings range from 10 pp

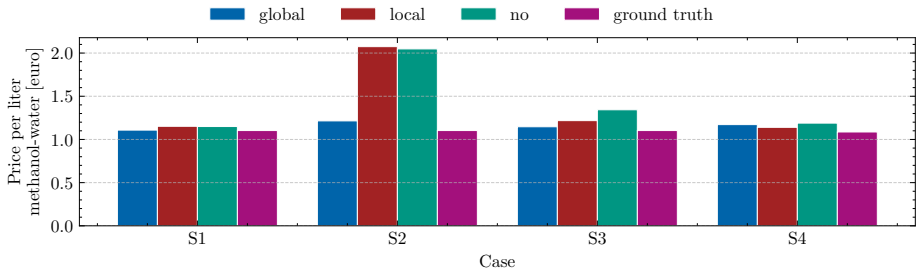


Figure 19: Corrected costs (C1) per produced liter of methanol-water for different cases (Case S1-S4) and different adaptation types for the process condition change PCC3.

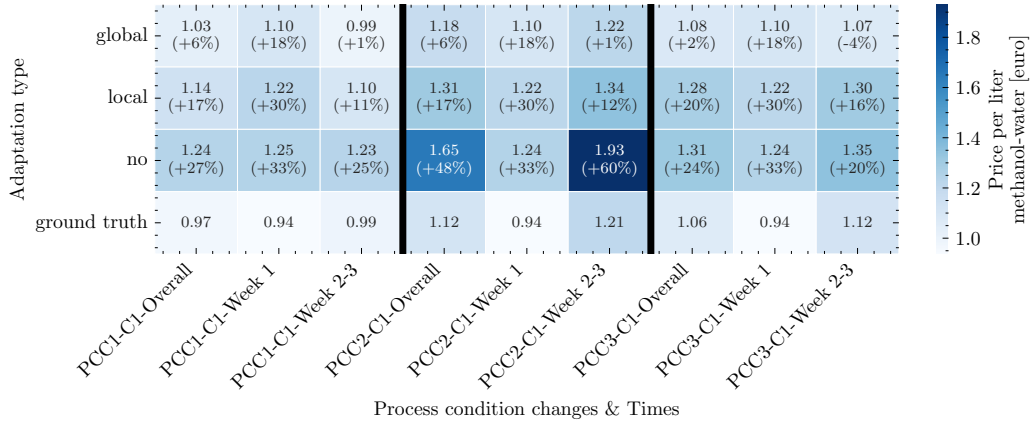


Figure 20: Mean corrected costs (C1) over all cases and the difference to the ground truth in percent for the different process condition changes and times using different adaptation methods.

to 20 pp compared to local adaptation and from 24 pp to 59 pp compared to no adaptation.

- Process condition change PCC2 has the biggest difference in price per liter of methanol-water for no adaptation, since one case produced a significantly smaller amount of methanol-water, which significantly increases the corrected price. For global and local adaptation, all process condition changes don't make much of a difference in the costs, whereas a change in the process condition increases the price per liter of methanol-water for the ground truth slightly.

For a better overview, Tab. 4 summarizes the most important economic performance indicators over all cases for every adaptation strategy. The computation times for scheduling and the characteristic map update are shown in Table A.6 in the appendix. With a planned schedule update interval of 24 h (see Table A.5), the computation times are more than sufficient for practical implementation.

Evaluation of active learning

Fig. 21 depicts the chosen operating points from active learning for case S1, showing their deviation from the economic optimum. This is defined as the operating point with the highest yield at the lowest energy consumption. The difference to the economic optimum ranges between 43.9 and 652.9 %, with

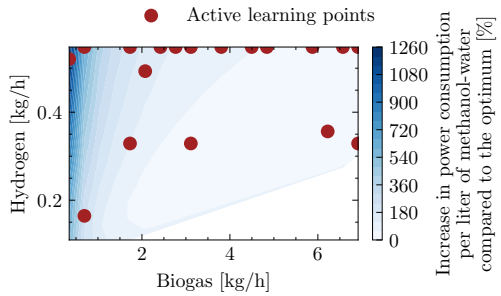


Figure 21: Active learning points for case S1 and their difference from the economic optimum.

Journal Pre-proof

Table 4: Overall mean economic comparison for all cases, process condition changes and adaptation strategies.

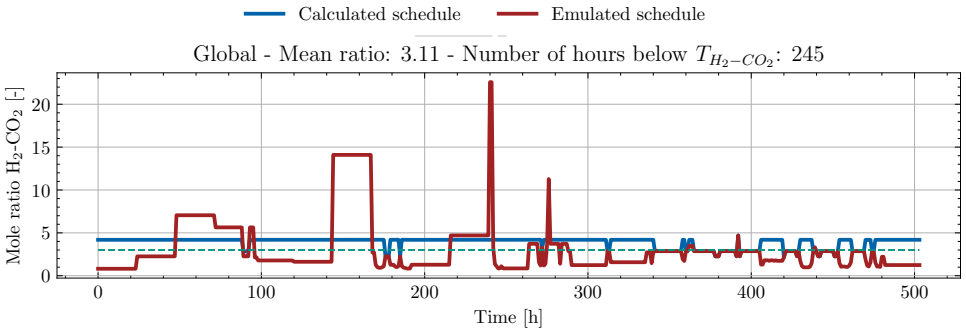
Strategy	Cost per liter [euro]	Corrected cost per liter [euro]	Actual production amount [L]	Production-target shortfall [L]
Global adaptation	1.09	1.10	745.46	2.27
Local adaptation	1.12	1.24	675.80	71.92
No adaptation	1.18	1.40	635.02	112.71
Ground truth	1.05	1.05	747.73	0

a median value of 153.4 %. It can clearly be seen that most points are at the boundary of the operating region, since exploration is weighted more heavily than exploitation. The total number of active learning points for the other cases during the first week is 18. The significant increase in costs is primarily due to the fact that very little methanol-water is produced at many of the active learning points. One way to improve the selection of active learning points would therefore be to find operating points that are closer to the economic optimum but are still near the proposed points. Another option would be to remove these points entirely from the operating range so that they are not approached at all.

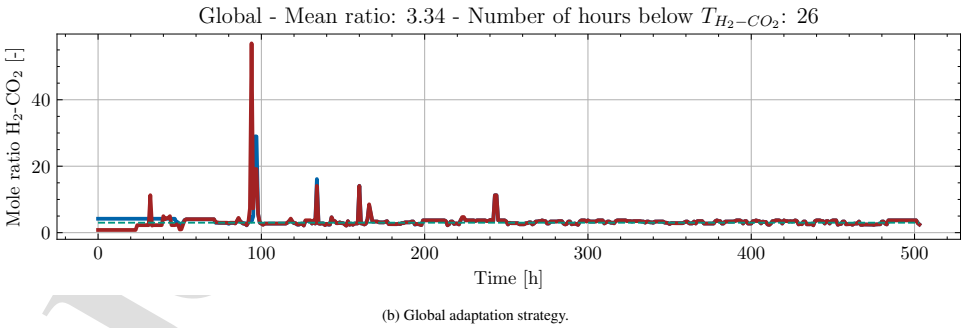
Evaluation of production constraints

Finally, constraint violations in the calculated schedule are used to validate the results. Fig. 22 depicts the calculated and emulated schedule of the molar ratio of H_2 to CO_2 in the synthesis gas for the local adaptation (top) and global adaptation (bottom) strategy for case S2 for PPC1. In both cases, the average ratio is above the specified threshold of $T_{\overline{H_2-CO_2}} = 3$.

However, in the local adaptation, there are 245 h with a ratio below the threshold $T_{H_2-CO_2}$, which is a 48.61 % of the time. In contrast, in the global adaptation, there are only 26 h, resulting in 5.16 % of the time. The mean deviation of the ratio is 0.69, with a maximum of 1.19 below the specified threshold for the local adaptation, while it is 1.17 and 1.19 for the global adaptation. Although the mean deviation is higher, there are still more time points with a larger deviation in the local adaptation, indicating that the violation severity is also lower for the global adaptation. The higher mean is only due to the fact that there are significantly fewer time points below the threshold for global adaptation. The constraint violations still occur, even



(a) Local adaptation strategy.



(b) Global adaptation strategy.

Figure 22: Molar ratio of H_2 to CO_2 in the synthesis gas for the calculated and emulated schedule for the local and global adaptation strategy.

if these constraints are included in the scheduling formulation, since the schedule is calculated with the adapted characteristic map, while it is emulated with the ground truth characteristic map. However, since the global adaptation adapts the characteristic map more quickly, there are hardly any time steps left with a value less than the threshold $T_{H_2-CO_2}$ after the first iteration. Since falling below the threshold value indicates degraded product quality as well as potentially higher wear on the catalyst in the reactor, it is necessary to reduce such occurrences to a minimum.

The case considered above represents an extreme scenario, since a constant initial characteristic map was used here. Nevertheless, the number of hours below the threshold per case is always lower for global adaptation than for local adaptation. On average across all cases, the threshold is undershot 1.49 % of the time with global adaptation and 12.79 % of the time with local adaptation for PPC1. Without any adaptation, the figure is as high as 17.9 %. Similar results are obtained for PPC2 and PPC3. Thus, global adaptation clearly demonstrates its advantages here. Regarding the production constraint on the quality of biogas, both the local and global adaptation have a mean molar fraction of carbon dioxide of about 6 % and thus are under the required threshold of 10 %.

4.2. *Results of adaptation to incremental drift*

The incremental drift cases from Tab. 2 are used to investigate characteristic map adaptation under incremental drift. It is assumed that, for the incremental drift, only the produced methanol-water is drifting over time. Fig. 23 shows the RMSE for each case and for each adaptation strategy. It is clear that the RMSE is significantly lower with global adaptation than with local adaptation or no adaptation at all.

The constant drift and linear drift per operating point are captured the best, since these are the easiest to detect and model, resulting in the lowest final RMSE after 49 d. Case I3 has the lowest difference in RMSE between the global and the local adaptation. This is obvious, since in this case the drift is not the same at every operating point but depends on the specific operating point. As a result, the drift characteristic map is a more complex function that requires more measurements and operating points to robustly calculate the drift across all operating points per time step. Nevertheless, it is better than no adaptation or only a local adaptation.

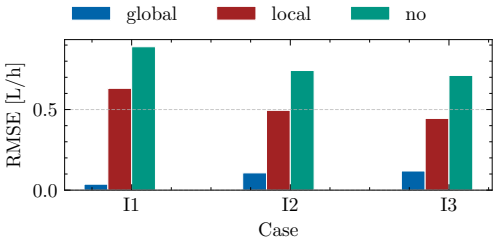


Figure 23: RMSE between the ground truth characteristic map and the model characteristic map for all cases I1 - I3 and adaptation types.

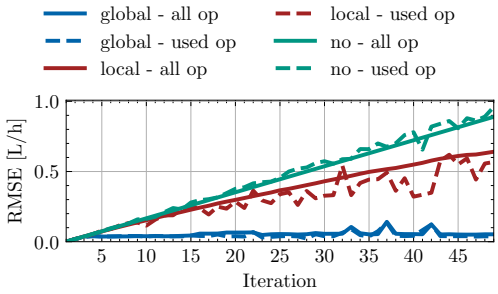


Figure 24: RMSE over the iterations between the ground truth characteristic map and the model characteristic map for case I1 for all adaptation types.

The RMSE over the number of iterations for case I1 is depicted in Fig. 24. For all adaptation strategies, the RMSE starts at zero, since it is assumed that the model characteristic map and the ground truth characteristic map are identical at the beginning. It becomes apparent that the RMSE remains nearly constant with global adaptation, whereas it otherwise increases. By detecting the drift and calculating a drift characteristic map — which is taken into account during adaptation — global adaptation makes it possible to keep the error between the ground truth and model characteristic map low, whereas this is not achieved otherwise.

The corrected prices per liter of methanol-water (C1) are visualized in Fig. 25. It is evident that they are lowest for global adaptation in all considered cases. The reduction in costs per liter of methanol-water for global adaptation is about 4.4 % lower compared to local adaptation and 5.8 % lower compared to no adaptation. At the same time, the amount of methanol-water produced is closest to the ground truth quantity with global adaptation, whereas it deviates more significantly with local adaptation or no adaptation.

Global adaptation is therefore advantageous over local adaptation or no adaptation at all, even with an incremental change in process behavior. This allows for a reduction in the plant-model mismatch and, consequently, in production costs as the process behavior changes.

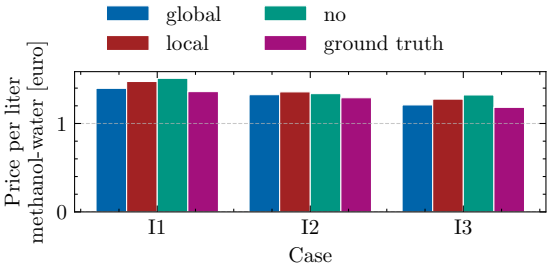


Figure 25: Costs per produced liter of methanol-water for cases I1 - I3 and different adaptation types.

Journal Pre-proof

5. Conclusion

This paper presents a new framework for the scheduling of chemical processes using self-adaptive characteristic maps as surrogate process models on the example of a novel PtM process. A new global adaptation algorithm is developed, which adapts the characteristic maps of process variables to overcome plant-model mismatch due to sudden and incremental drift in the process behavior. The global adaptation detects sudden and incremental drifts in the process variables using process measurements and calculates a residual characteristic map to update the model characteristic map at regular time intervals.

The results indicate a rapid and effective adaptation, as evidenced by a rapidly decreasing and significantly lower RMSE compared to a local adaptation strategy and no adaptation. As a result, the plant-model mismatch is significantly reduced, and the model is more reliable for tasks such as schedule calculation and process monitoring. It appears that the adaptation depends heavily on the difference between the initial and ground truth characteristic maps. An advantage here is that global adaptation performs significantly better when simple initial characteristic maps are specified without expert knowledge. This can significantly reduce the effort required to create the initial characteristic maps, since no time-consuming and complex simulations need to be performed to generate them.

When considering the economic costs calculated by the schedule emulation, the cost per liter of methanol-water produced is reduced by over 11.8 % when the global adaptation strategy is used, compared to local adaptation, and over 21.4 % compared to no adaptation. At the same time, only with global adaptation is the specified amount of methanol-water produced within the time horizon under consideration. It appears that the operating characteristics adjusted through global adaptation respond better to process changes, since in this case the entire operating characteristic map is up to date, whereas with local adaptation only certain areas are adjusted.

Furthermore, global adaptation also results in significantly fewer constraint violations regarding certain production conditions, such as molar ratios or proportions in process streams of the calculated schedules. Depending on the nature and severity of the violation, these can significantly restrict the process, making compliance with these constraints critical for operation.

Ultimately, it becomes clear that in this case there is always a trade-off between model accuracy and reduced production costs. The more operating points are adjusted through active learning, the more accurate the model becomes — but at the cost of higher operating costs. At the same time, it is also evident that if only a few operating points are sufficient for optimal process operation, the difference between local and global adaptation becomes minimal. The advantage of global adaptation increases significantly as the number of operating points rises.

Future research will include examining the algorithm's transferability to other processes. Furthermore, the implementation and validation of the algorithm in the actual plant with real measurements is an important aspect for the future.

Acknowledgment



Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or, neither the European Union nor the granting authority can be held responsible for them.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101083323.

References

[1] D. Stanelyte, N. Radziukyniene, V. Radziukynas, Overview of Demand-Response Services: A Review, Energies 15 (5) (2022) 1659. doi:10.3390/en15051659.

[2] Y. Zou, X. Ma, Y. Yang, S. Li, An overview of chemical process operation-optimization under complex operating conditions, Digital Chemical Engineering 16 (2025) 100249. doi:10.1016/j.dche.2025.100249.

[3] L. C. Brée, K. Perrey, A. Bulan, A. Mitsos, Demand side management and operational mode switching in chlorine production, AIChE Journal 65 (7) (2019) e16352. doi:10.1002/aic.16352.

[4] M. T. Kelley, R. C. Pattison, R. Baldick, M. Baldea, An MILP framework for optimizing demand response operation of air separation units, Applied Energy 222 (2018) 951–966. doi:10.1016/j.apenergy.2017.12.127.

[5] Y. Zheng, S. You, X. Li, H. W. Bindner, M. Münster, Data-driven robust optimization for optimal scheduling of power to methanol, Energy Conversion and Management 256 (2022) 115338. doi:10.1016/j.enconman.2022.115338.

[6] A. Boldrini, D. Koolen, W. Crijns-Graus, M. V. D. Broek, The demand response potential of a hydrogen-based iron and steel plant, in: 2022 18th International Conference on the European Energy Market (EEM), IEEE, Ljubljana, Slovenia, 2022, pp. 1–6. doi:10.1109/EEM54602.2022.9921013.

[7] D. Krishnamoorthy, S. Skogestad, Real-Time optimization as a feedback control problem – A review, Computers & Chemical Engineering 161 (2022) 107723. doi:10.1016/j.compchemeng.2022.107723.

[8] K. McBride, K. Sundmacher, Overview of Surrogate Modeling in Chemical Process Engineering, Chemie Ingenieur Technik 91 (3) (2019) 228–239. doi:10.1002/cite.201800091.

[9] D. Šturek, S. Lazarova-Molnar, Surrogate Modeling: Review and Opportunities for Expert Knowledge Integration, Procedia Computer Science 257 (2025) 826–833. doi:10.1016/j.procs.2025.03.106.

[10] O. Keremer, F. C. Malay, B. Deveci, P. Kirci, P. Unal, Optimisation of a Chemical Process by Using Machine Learning Algorithms with Surrogate Modeling, in: M. Younas, I. Awan, T.-M. Grønli (Eds.), Mobile Web and Intelligent Information Systems, Springer Nature Switzerland, Cham, 2023, pp. 187–201. doi:10.1007/978-3-031-39764-613.

[11] M. Urquhart, E. Ljungskog, S. Sebben, Surrogate-based optimisation using adaptively scaled radial basis functions, Applied Soft Computing 88 (2020) 106050. doi:10.1016/j.asoc.2019.106050.

[12] A. Bhosekar, M. Ierapetritou, Advances in surrogate based modeling, feasibility analysis, and optimization: A review, Computers & Chemical Engineering 108 (2018) 250–267. doi:10.1016/j.compchemeng.2017.09.017.

Journal Pre-proof

[13] M. Li, Y. Zhuang, W. Li, Y. Dong, L. Zhang, J. Du, S. Shengqiang, A surrogate-based optimization framework for simultaneous synthesis of chemical process and heat exchanger network, Chemical Engineering Research and Design 170 (2021) 180–188. doi:10.1016/j.cherd.2021.04.001.

[14] A. Obermeier, N. Vollmer, C. Windmeier, E. Esche, J.-U. Repke, Generation of linear-based surrogate models from non-linear functional relationships for use in scheduling formulation, Computers & Chemical Engineering 146 (2021) 107203. doi:10.1016/j.compchemeng.2020.107203.

[15] C. Nentwich, S. Engell, Application of surrogate models for the optimization and design of chemical processes, in: 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 1291–1296, iSSN: 2161-4407. doi:10.1109/IJCNN.2016.7727346.

[16] A. Di Pretoro, B. Bruns, S. Negny, M. Grünewald, J. Riese, Demand response scheduling using derivative-based dynamic surrogate models, Computers & Chemical Engineering 160 (2022) 107711. doi:10.1016/j.compchemeng.2022.107711.

[17] A. Shokry, P. Baraldi, E. Zio, A. Espuña, Dynamic Surrogate Modeling for Multistep-ahead Prediction of Multivariate Nonlinear Chemical Processes, Industrial & Engineering Chemistry Research 59 (35) (2020) 15634–15655. doi:10.1021/acs.iecr.0c00729.

[18] K. Bakon, T. Holczinger, Z. Süle, S. Jaskó, J. Abonyi, Scheduling Under Uncertainty for Industry 4.0 and 5.0, IEEE Access 10 (2022) 74977–75017. doi:10.1109/ACCESS.2022.3191426.

[19] L. S. Dias, M. G. Ierapetritou, Integration of scheduling and control under uncertainties: Review and challenges, Chemical Engineering Research and Design 116 (2016) 98–113. doi:10.1016/j.cherd.2016.10.047.

[20] A. Baghban, P. M. Castro, F. Oliveira, Data-driven robust optimization for pipeline scheduling under flow rate uncertainty, Computers & Chemical Engineering 193 (2025) 108924. doi:10.1016/j.compchemeng.2024.108924.

[21] B. Chachuat, B. Srinivasan, D. Bonvin, Adaptation strategies for real-time optimization, Computers & Chemical Engineering 33 (10) (2009) 1557–1567. doi:10.1016/j.compchemeng.2009.04.014.

[22] J. O. A. Matias, G. A. C. Le Roux, Real-time Optimization with persistent parameter adaptation using online parameter estimation, Journal of Process Control 68 (2018) 195–204. doi:10.1016/j.jprocont.2018.05.009.

[23] S. Skogestad, Self-optimizing control: the missing link between steady-state optimization and control, Computers & Chemical Engineering 24 (2-7) (2000) 569–575. doi:10.1016/S0098-1354(00)00405-1.

[24] B. Ghojogh, A. Ghodsi, F. Karray, M. Crowley, KKT Conditions, First-Order and Second-Order Optimization, and Distributed Optimization: Tutorial and Survey, arXiv:2110.01858 [math.OC] (Oct. 2021). doi:10.48550/arXiv.2110.01858.

[25] A. Marchetti, G. François, T. Faulwasser, D. Bonvin, Modifier Adaptation for Real-Time Optimization - Methods and Applications, Processes 4 (4) (2016) 55. doi:10.3390/pr4040055.

[26] J. Ehlhardt, A. Ahmad, I. Wolf, S. Engell, Real-Time Optimization Using Machine Learning Models Applied to the 4,4′ -Diphenylmethane Diisocyanate Production Process, Chemie Ingenieur Technik 95 (7) (2023) 1096–1103. doi:10.1002/cite.202200244.

[27] M. Vaccari, G. Pannocchia, A Modifier-Adaptation Strategy towards Offset-Free Economic MPC, Processes 5 (1) (2017) 2, publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/pr5010002.

[28] A. Papasavvas, T. de Avila Ferreira, A. G. Marchetti, D. Bonvin, Analysis of output modifier adaptation for real-time optimization, Computers & Chemical Engineering 121 (2019) 285–293. doi:10.1016/j.compchemeng.2018.09.028.

[29] F. J. Serralunga, M. C. Mussati, P. A. Aguirre, Model Adaptation for Real-Time Optimization in Energy Systems, Industrial & Engineering Chemistry Research 52 (47) (2013) 16795–16810. doi:10.1021/ie303621j.

[30] J. Matias, J. Jäschke, Online model maintenance in real-time optimization methods, Computers & Chemical Engineering 145 (2021) 107141. doi:10.1016/j.compchemeng.2020.107141.

[31] W. Gao, S. Wenzel, S. Engell, A Reliable Modifier-Adaptation Strategy for Real-Time Optimization, Computers & Chemical Engineering (Aug. 2016). doi:10.1016/j.compchemeng.2016.03.019.

[32] B. Tang, E. B. Khalil, J. Drgoňa, Learning to Optimize for Mixed-Integer Non-linear Programming, arXiv:2410.11061 [cs.LG] version: 1 (Oct. 2024). doi:10.48550/arXiv.2410.11061.

[33] A. G. Marchetti, M. Singhal, T. Faulwasser, D. Bonvin, Modifier adaptation with guaranteed feasibility in the presence of gradient uncertainty, Computers & Chemical Engineering 98 (2017) 61–69. doi:10.1016/j.compchemeng.2016.11.027.

[34] C. A. Floudas, Nonlinear and mixed integer optimization : fundamentals and applications, Topics in chemical engineering, Oxford Univ. Press,, New York, NY [u.a.] :, 1995.

[35] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, G. Zhang, Learning under Concept Drift: A Review, IEEE Transactions on Knowledge and Data Engineering 31 (12) (2019) 2346–2363, conference Name: IEEE Transactions on Knowledge and Data Engineering. doi:10.1109/TKDE.2018.2876857.

[36] K. P. Murphy, Machine learning: a probabilistic perspective, 4th Edition, Adaptive computation and machine learning series, MIT Press, Cambridge, Mass., 2013.

[37] C. E. Rasmussen, C. K. I. Williams, Gaussian processes for machine learning, 3rd Edition, Adaptive computation and machine learning, MIT Press, Cambridge, Mass., 2008.

[38] J. Wang, An Intuitive Tutorial to Gaussian Process Regression, Computing in Science & Engineering 25 (4) (2023) 4–11. doi:10.1109/MCSE.2023.3342149.

[39] A. A. Neath, J. E. Cavanaugh, The Bayesian information criterion: background, derivation, and applications, WIREs Computational Statistics 4 (2) (2012) 199–203, _eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wics.199. doi:10.1002/wics.199.

[40] A. Sóbester, S. J. Leary, A. J. Keane, On the Design of Optimization Strategies Based on Global Response Surface Approximation Models, Journal of Global Optimization 33 (1) (2005) 31–59. doi:10.1007/s10898-004-6733-1.

[41] H. Liu, Y.-S. Ong, J. Cai, A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design, Structural and Multidisciplinary Optimization 57 (1) (2018) 393–416. doi:10.1007/s00158-017-1739-8.

[42] D. Wu, C.-T. Lin, J. Huang, Active learning for regression using greedy sampling, Information Sciences 474 (2019) 90–105. doi:10.1016/j.ins.2018.09.060.

[43] T. M. Cover, J. A. Thomas, Elements of information theory, 2nd Edition, Wiley-Interscience, Hoboken, N.J, 2006.

[44] C. M. Bishop, Pattern recognition and machine learning, Information science and statistics, Springer, New York, 2006.

[45] M. Kollmer, M. Vogelbacher, M. Terzi, M. Singh, F. Vidal-Vazquez, J. Matthes, Demand Response for a Novel Decentralized Power-To-Methanol Process, in: 2024 International Conference on Smart Energy Systems and Technologies (SEST), IEEE, Torino, Italy, 2024, pp. 1–6. doi:10.1109/SEST61601.2024.10694417.

[46] F. Vidal-Vazquez, R. Dittmeyer, P. Pfeifer, Vorrichtung und Verfahren zur Herstellung von Methanol aus Kohlendioxid, DE patent (Oct. 2022).

[47] AspenTech, Aspen Plus | Leading Process Simulation Software | AspenTech (Nov. 2025).

[48] M. Kollmer, M. Vogelbacher, M. Terzi, M. Singh, F. Vidal-Vazquez, J. Matthes, Energy Flexibilization and Demand Response for a Novel Decentralized Power-to-Methanol Process, Chemie Ingenieur Technik n/a (n/a) (2026) e70103, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cite.70103. doi:10.1002/cite.70103.

[49] L. Gurobi Optimization, Gurobi Optimizer Reference Manual (2026). URL `https://www.gurobi.com/`

[50] B. Burger, Fraunhofer ISE (Energy Charts) (2024).

Journal Pre-proof

Appendix A. Supplementary material

Table A.5: Implementation settings, algorithmic parameterization and solver configuration.

Category	Parameter description	Parameter value
Sudden drift detection	Threshold for Mahalanobis distance between map and measurement values	$t_{\mathrm{M,map}}$ (from χ^2 distribution)
	Threshold for Mahalanobis distance between two measurements of different iterations of one operating point	$t_{\mathrm{M,pts}}$ (from χ^2 distribution)
Incremental drift detection	Threshold for number of failed KS-tests per operating point	$\zeta = 3$
	Threshold for number of conspicuous operating points	$\xi = 2$
	Oberservation window for incremental drift detection	$T_{\mathrm{inc}} = 40$ iterations
Gaussian process regression	Kernel type	RBF, Matérn, constant, linear
	Kernel selection	Voting using BIC, LML and MRMSE
	Hyperparameter optimization	Maximum Likelihood Estimation (MLE)
	Noise assumptions	Gaussian (σ_n^2 fixed depending on output variable, estimated with real process noise data)
Active learning	Candidate points	$N_{\mathrm{cand}} = 3$
	Exploitation weight	$w_1 = 0.3$
	Exploration weight	$w_2 = 0.7$
	Batch diversity	$\gamma = 0.5$
Data acquisition & updates	Data acquisition interval	$T_{\mathrm{sample}} = 24\,\mathrm{h}$
	Update interval	$\Delta t_{\mathrm{update}} = 24\,\mathrm{h}$
	Measurements per update	Depending on the calculated schedule (between 0 and 24 because of update interval)
Solver configuration	Optimization solver	Gurobi (MILP)
	Convergence tolerance	0.01 (1 %)
	Maximum time	1000 s

Table A.6: Mean runtimes and 95 % confidence interval (CI) per case and adaptation strategy for scheduling and characteristic map update.

Case #	Adaptation strategy	Mean scheduling time [s]	95 % CI	Mean update-time [s]	95 % CI
S1	Global	157.22	93.73	0.3165	0.0106
	Local	131.00	88.99	0.000217	0.000010
	No	169.81	132.29	-	-
	Ground truth	210.49	139.82	-	-
S2	Global	167.61	123.01	0.3503	0.0180
	Local	225.71	173.88	0.000242	0.000024
	No	186.40	164.27	-	-
	Ground truth	212.54	139.90	-	-
S3	Global	154.28	104.55	0.3276	0.0118
	Local	555.15	222.67	0.000278	0.000015
	No	423.03	155.18	-	-
	Ground truth	215.95	140.60	-	-
S4	Global	130.76	103.46	0.3948	0.0197
	Local	208.61	135.60	0.000278	0.000018
	No	220.79	139.89	-	-
	Ground truth	242.12	143.07	-	-

Journal Pre-proof

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof