Novel Reformulations for Modeling Uncertainty and Variations in a Framework for Chemical Process Design

(article starts on next page)

Article

Novel Reformulations for Modeling Uncertainty and Variations in a Framework for Chemical Process Design

Christian Langner,* Elin Svensson, Stavros Papadokonstantakis, and Simon Harvey

Read Online

ACCESS | 📊 Metrics & More | 📑 Article Recommendations | SI Supporting Information

ABSTRACT: In this work, we present a framework to support designers in the early-stage screening and design of chemical processes under uncertainty. The suggested framework is a stepwise approach which identifies: (1) if structural design proposals are structurally feasible, (2) the necessary and most cost-efficient overdesign of equipment size to allow for steady-state flexible operation, and (3) a basis to compare different structural design proposals with respect to a given objective. In contrast to previous work, the suggested framework utilizes advanced modeling of the expected uncertainty space to allow for considering dependencies in the uncertain parameters as well as independent operating periods. Theoretical examples illustrate how the advanced modeling can enable a better representation of the actual flexibility of a process compared to using more traditional representations of the uncertainty space. To illustrate the benefits of the proposed framework, the paper also presents an industrial case study of a heat exchanger network retrofit.
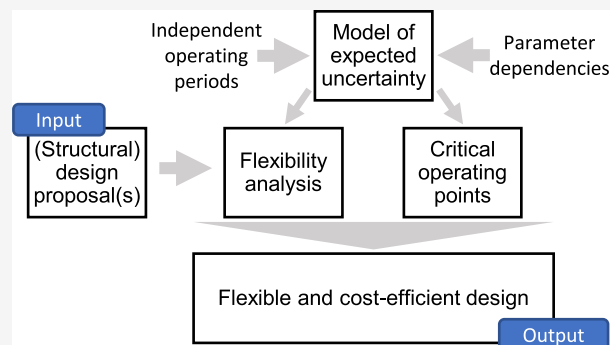
1. INTRODUCTION

Chemical process design is usually subject to uncertainty. This uncertainty can have different origins; for example, process parameters such as flow rates or heat-transfer coefficients may vary (uncontrolled) or may not be known exactly. Traditionally, as reported in the literature,[1−3] the approach to handle uncertainty in such parameters is to consider (only) nominal conditions during the design stage and use overdesign to compensate for the potential impact of the uncertainty. Some of this overdesign may be unnecessary, resulting in suboptimal design solutions. To identify the optimal solution to such a problem, specific methods and strategies have been developed, which are usually allocated to the field of optimization under uncertainty.

Grossmann and Sargent[4] defined the objective of the optimal design problem with uncertain parameters as designing processes "that are always able to meet the specifications for any feasible [or expected] values of the [uncertain] parameters and that, at the same time, are optimum with respect to a [···] cost function". Note that the authors explicitly considered the uncertain parameters to be continuous (and bounded), that is, the parameters can take any value between a lower bound and an upper bound, and that the optimal process design allows for feasible operation at any of these values. The authors concluded that design under uncertainty is an infinite problem. There exist two categories of approaches to solve such a problem, as pointed out by Steimel and Engell,[5] namely, sampling-based approaches and parametric approaches. In parametric approaches, the sampling of the uncertainty space is not required, meaning that

the expected value of the objective function is obtained without approximation. Examples have been presented by Pistikopoulos and Ierapetritou,[6] Acevedo and N. Pistikopoulos,[7] as well as Rooney and Biegler.[8] Difficulties may arise when the number of uncertain parameters increases due to the increased computational expenses. Sampling-based approaches are commonly based on the two-stage stochastic programming with recourse formulation and its transformation into a discretized deterministic equivalent problem. The idea behind two-stage stochastic programming for chemical process design is to optimize the design parameters (first-stage decisions) for a suitable cost function whose expected value is computed for an optimal choice of control variables (second-stage decisions). When transforming the problem to a discretized problem (sampling-based), the expected value of the cost function is approximated for a discrete set of realizations of the uncertain parameters. The interested reader is referred to the studies by Sahinidis[9] and Birge and Louveaux[10] for further information.
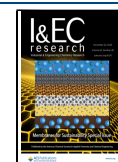
A major challenge with sampling-based approaches is that the number of samples necessary for accurate approximation of the expected objective function value as well as guaranteed

feasibility strongly increases with the number of uncertain parameters. Several studies in the literature suggest strategies to reduce the number of samples while maintaining a fair approximation of the expected objective value. Examples can be found in studies by Wei and Realff[11] and Pintarič et al.[12] Recently, especially in the field of power engineering, algorithms based on machine learning have been utilized to identify representative operating points/periods from a given distribution of operating points (see, e.g., Mozafari and Rosehart[13]). In addition to the expected value of the objective, the discretization needs to detect the minimum set of samples required to guarantee steady-state flexible operation, that is, feasible operation at any combination of uncertain parameter values within the expected uncertainty space[a]. In this context, flexibility analysis has evolved into a useful tool.

Flexibility analysis originates from the 1980s and aims to quantify a system's ability to cope with uncertainties. In this context, Halemane and Grossmann[14] formulated a flexibility test problem for fixed design specifications which enables the user to determine whether steady-state flexible operation of a process is possible given an expected uncertainty space. The authors further defined critical parameter values as those combinations of uncertain parameter values for which the feasible region of operation is the smallest. In this context, they postulated that a design that meets its operational targets at these critical operating points will also allow for steady-state flexible operation. To identify these critical operating points for processes which can be described by convex constraints, the authors utilized a procedure suggested by Grossmann and Sargent[4] which builds upon individual maximization of each constraint assuming monotonicity. Based on the work by Halemane and Grossmann, Pintarič and Kravanja suggested in two studies[15,16] several algorithms to identify those combinations of the uncertain parameter values that require the largest overdesign of process units for the expected uncertainty space. Although the authors presented no rigorous mathematical proof, they indicated that these algorithms can also identify critical operating points for nonconvex constraint functions. Based on the suggested algorithms to identify critical operating points, the authors developed frameworks for synthesis of chemical process design[17] and for designing heat exchanger networks[18] which are subject to uncertain operating data.

In order to quantify a system's ability to cope with uncertainties, several metrics were suggested in the 1980s, including the flexibility index developed by Swaney and Grossmann[19] which has gained much attention in the literature, and is still subject to research (see, e.g., Di Pretoro et al.,[20] da Silva et al.,[21] and Luo et al.[22]). The flexibility index is a scalar value that expresses the share of the expected uncertainty space that is feasible. More specifically, for each uncertain parameter, the flexibility index indicates the share of the expected disturbance range in which the respective parameter may vary while achieving feasible operation. For an extensive review on flexibility analysis, the interested reader is referred to the reviews by Grossmann et al.[2] and Zhang et al.[23] The flexibility index was utilized in several (stepwise) frameworks to design or redesign chemical processes which are subject to uncertain parameters (see, e.g., Floudas and Grossmann,[24] Papalexandri and Pistikopoulos,[25] and Pintarič and Kravanja[18]). In line with these studies, we previously presented a framework[26] to achieve optimal retrofit measures of heat exchanger networks considering parameter uncertainty. The referred framework is a stepwise approach which requires structural design proposals[b] as input

and utilizes flexibility analysis as well as identification of critical operating points. The optimal design specifications as well as the expected objective value are obtained by solving the discretized design under uncertainty problem considering representative and critical operating points.

As a complement to flexibility analysis, the concept of process operability analysis has been suggested in the literature. The interested reader is referred to the work by Lima et al.[27] who presented a comparative review on the concepts of flexibility and operability analysis. The authors concluded that the concepts examine a process from different perspectives and provide valuable complementary information. More specifically, flexibility analysis aims to identify whether a given process design can operate feasibly over the entire expected uncertainty space. In comparison, the results obtained by operability analysis provide insights into whether a process controller can achieve its mission or if operability is limited by the process design. According to Lima et al.[27] such considerations are of special interest if the operational target of a process is defined by intervals and not as a single set point.

In the aforementioned studies[15−18,24−26] on chemical process design based on flexibility analysis, it is assumed that the expected distribution of the uncertain parameter values is independent, that is, no correlating trends between different uncertain parameters can be observed. The assumption of independent variation is reflected in the models used to express the expected uncertainty space, which are utilized for flexibility analysis. However, for real applications, this assumption is not always valid. Especially for industrial case studies, it is likely that when defining the system boundaries and thereby also the input parameters, which often are subject to some uncertainty, upstream dependencies between (some of) these input parameters can be missed (due to the size and complexity of the underlying system). Several studies[28−30] have investigated the impact of considering dependencies between uncertain parameters when modeling the expected uncertainty space for flexibility analysis. All of these studies conclude that ignoring dependencies may lead to the flexibility analysis metric underestimating the actual flexibility of the process, sometimes even significantly, as a consequence of bad resemblance between the modeled and the actual expected uncertainty space. Therefore, wrong conclusions may be drawn during a design process, for example, process design proposals may be discarded based on the flexibility metric although steady-state flexible operation for the actual uncertainty space would be possible[c]. Furthermore, Rooney and Biegler[32] reported differences in the obtained design parameter values, that is, equipment sizes, when solving the discretized design under uncertainty problem using realizations of uncertain parameters drawn from an independent distribution and from distributions which show a (positive or negative) correlation.

In this work, we build upon recent developments in flexibility analysis to propose a framework that can be utilized in the design and retrofit of chemical processes subject to uncertainty. In contrast to existing approaches, the proposed framework includes strategies for considering parameter dependencies and multiple independent operating periods (see Section 2.2 for further information). The proposed framework expands upon previous work[26] for investigating flexible and cost-efficient retrofit measures of heat exchanger networks, which utilizes flexibility analysis and identification of as well as consideration of critical operating points. The main motivation for this work is the need to consider parameter dependencies when modeling

the expected uncertainty space to avoid wrong conclusions being drawn from flexibility analysis. Additionally, we show that the application of traditional models may lead to a significant discrepancy between the modeled uncertainty space and the actual uncertainty space for problems characterized by the occurrence of (multiple) independent operating periods. To overcome these issues, we adopt new approaches to model the expected uncertainty space with high accuracy, based on our recently published studies on flexibility analysis considering parameter dependencies[31] as well as multiple independent operating periods.[33,34] One consequence of inexact modeling of the expected uncertainty space can be that identified critical operating points represent combinations of uncertain parameter values which are not expected to occur simultaneously, which in turn can result in unnecessary overdesign of the equipment. To avoid this, new approaches to model the expected uncertainty space can also be used to identify critical operating points. With the proposed developments, we aim to avoid unnecessary overdesign of the equipment and thereby reduce costs/increase revenue by ensuring that the modeled uncertainty space coincides well with the actual uncertainty space.

The paper is structured as follows: in Section 2, the theoretical and mathematical background of the different methodologies utilized in the suggested framework is provided. More precisely, the two-stage stochastic programming with recourse formulation and its deterministic equivalent problem are presented followed by essential background information on flexibility analysis, modeling of the expected uncertainty space and critical operating points. In Section 3, the proposed framework is introduced, followed by a detailed explanation of the different steps with a focus on the new additions to the framework. In Section 4, the applicability and benefits of the new additions to the framework are illustrated by an industrial case study. Finally, the findings of this work are discussed followed by concluding remarks.

2. THEORETICAL BACKGROUND

2.1. Design Problem under Uncertainty.
Steimel and Engell[5] and Pintarič and Kravanja[18] formulated the discretized equivalent to the stochastic two-stage design under uncertainty (with recourse) formulation, as given in problem 1.

$$\min_{z_s, d} G(d) + \sum_{s \in S} [w_s \cdot C_s(d, z_s, \theta_s)]$$

s. t.

$$
\left.
\begin{array}{l}
h_i(x_s, z_s, d, \theta_s) = 0;\ i \in I \\
g_j(x_s, z_s, d, \theta_s) \leq 0;\ j \in J \\
d \geq 0 \\
x_s, z_s, d, \theta_s \in \mathbb{R}
\end{array}
\right\} s \in S
$$

$$(1)$$

The objective function in problem 1 consists of two parts. The first term $G(d)$ describes the cost connected to investment in process units to provide the capacity necessary for steady-state flexible operation—this selection of the design parameters, d, is referred to as the first-stage decisions. The second term is the expected value of the operating costs, which is approximated via the summation of the costs for discrete scenarios, C_s, multiplied by the probabilities of the scenarios, w_s. The expected value of the operating costs is dependent on the selection of the design parameters, d, and the second-stage decisions which are the

selection of the control variables, z_s. The design problem is subject to equality constraints ($h_i \forall i \in I$) which often result from the models of the process units and the interconnections between them, and inequality constraints ($g_j \forall j \in J$) which commonly arise from product property constraints and capacity specifications.

Due to the discretization of the scenarios ($s \in S$) in problem 1, the numerical integration over the expected uncertainty space to approximate the expected value of the operating costs as well as the selection of the design parameter values can be performed simultaneously. In problem 1, the discrete choices of the uncertain parameters to represent the discrete scenarios are explicitly denoted by θ_s. Problem 1 can consequently be seen as a multiperiod (non-)linear program (NLP) which aims to minimize some objective value (e.g., the total annualized cost, TAC) of the design or retrofit of a chemical process considering its operation during discrete scenarios or operating periods. Note that state variables, x_s, are included in problem 1 which, in the mathematical sense, are similar to the control variables but are uniquely determined by the equality constraints.

As outlined in Section 1, when designing or retrofitting a chemical process which is subject to uncertainty, a sufficient number of realizations of the uncertain parameters, that is, scenarios, need to be considered to allow for steady-state flexible operation as well as a fair approximation of the objective function value, for example, TAC. A common approach for defining the scenarios to be included in the discretized design problem is to identify the most representative scenarios for operation of the process, that is, scenarios that are expected to be a good representation of typical conditions that occur during normal operation. Such representative operating scenarios should represent the operating conditions during specific time periods (e.g., different seasons or different production campaigns) and should allow for adequate approximation of the expected objective value.

To also ensure steady-state flexible operation within the expected uncertainty space, critical operating points need to be identified and included in the constraints of problem 1 (i.e., $h_i \forall i \in I$ and $g_j \forall j \in J$). Note that critical operating points usually represent extreme operating conditions and not typical conditions that occur during normal operation. Therefore, critical operating points should not be considered when approximating the expected value of the objective function, that is, the operating costs, C_s, for operation at the critical operating points should not be considered in the objective function of problem 1. When denoting the set of representative scenarios with OP and the set of critical operating points with CP, the operating costs, C_s are calculated for all scenarios in OP while the constraints of problem 1 are evaluated at all scenarios/points in the union (OP ∪ CP).

The concept of critical operating points originates from flexibility analysis (see Section 2.2) and refers to those realizations of the uncertain parameters within the expected uncertainty space that require the largest overdesign of process units to allow for steady-state flexible operation. In the next section, a comprehensive overview on flexibility analysis is given followed by further information on critical operating points (see Section 2.3).

2.2. Flexibility Analysis and Modeling of the Uncertainty Space.
As mentioned in Section 1, a well-established concept for performing flexibility analysis of chemical processes is the flexibility index which was introduced by Swaney and Grossmann.[19] The core idea of the flexibility index is to provide

a scalar value ≥ 0 which scales a geometric shape that is used to model the expected uncertainty space in the (hyper-)space of the uncertain parameters. More specifically, the returned value of the flexibility index problem is the scaling factor for which the scaled geometric shape intersects with the boundary of the feasible region, meaning that for any higher value of the scaling factor parts of the shape would be outside the feasible region. Swaney and Grossmann[19] modeled the expected uncertainty space using the expected extreme values, which they expressed as expected deviations ($\Delta\theta^-$ and $\Delta\theta^+$) from a nominal or mean operating point (θ_N). Consequently, the modeled uncertainty space can be imagined as a hyperrectangle, that is, a multidimensional rectangle such as a cuboid for three dimensions. This means that the flexibility index can be imagined as the ratio between the largest scaled hyperrectangle within the feasible region and the hyperrectangle defined by the expected extreme values. The mathematical formulation of the scaled hyperrectangle is given in eq 2. Note that the scaling factor is depicted by δ.
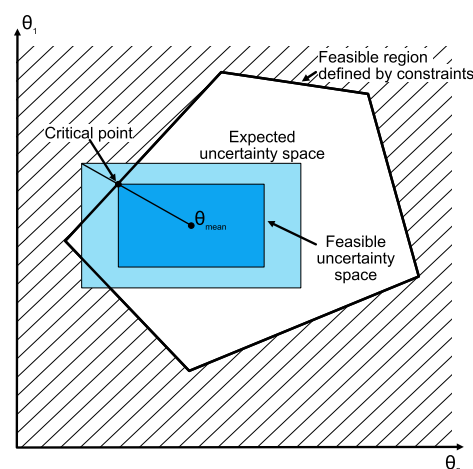
$$T_{\text{hyperrectangle}}(\delta) = \{\theta_i \mid \theta_{i,N} - \delta\Delta\theta_i^- \leq \theta_i \leq \theta_{i,N} + \delta\Delta\theta_i^+\}$$
$$\forall\ \theta_i \in \theta \tag{2}$$

$$T_{\text{boundary}}(\delta)$$
$$= \begin{cases} \{\theta_i \mid \theta_{i,N} - \delta\Delta\theta_i^- \leq \theta_i \leq \theta_{i,N} + \delta\Delta\theta_i^+\}\ \forall\ \theta_i\ \text{in}\ \theta_{\text{ind}} \\ \{\theta_j \mid f_l(\theta_{\text{ind}}) \leq \theta_j \leq f_u(\theta_{\text{ind}})\}\ \forall\ \theta_j\ \text{in}\ \theta_{\text{dep}} \end{cases}$$
$$\tag{3}$$

To obtain the flexibility index, Swaney and Grossmann[19] defined the flexibility index problem. The aim of the problem is to identify the maximum value of δ for which none of the constraint functions describing the process (i.e., $h_i \forall i \in I$ and $g_j \forall j \in J$ in problem 1) is violated, that is, the maximum value of δ for which the operational targets of the process are met given the (fixed) design specifications of the process. Consequently, if the index is greater than or equal to 1, the process can be operated for all expected variation. For two uncertain parameters, this search for the largest scaled rectangle within the feasible region (i.e., feasible uncertainty space) is visualized in Figure 1a. In Figure 1a, the feasible uncertainty space is smaller than the expected uncertainty space, that is, the rectangle corresponding to the expected variations. Thus, feasible operation with respect to all expected variations cannot be guaranteed. Note that in Figure 1a, the point at which the feasible uncertainty space coincides with the boundary of the feasible region has been marked as critical point. This is discussed further in Section 2.3.

Different algorithms have been suggested to solve the flexibility index problem, including the active set approach proposed by Grossmann and Floudas.[28] The advantage of the active set approach is that an iterative evaluation of the vertices of the expected uncertainty space can be avoided. As a result, global solution schemes have been developed for this approach (see, e.g., the work by Floudas et al.[35]).

2.2.1. Flexibility Analysis Considering Dependencies in the Uncertain Parameters. Expressing the uncertainty space as a hyperrectangle is often accurate when the uncertain parameters are independent of each other. On the other hand, it has been reported in the literature[28−30] that if dependencies occur in the uncertain parameters, the flexibility index (δ) based on the hyperrectangle uncertainty space may underestimate the flexibility of the process. To consider these parameter

(a) (hyper-)rectangle with respect to expected variations and maximum scaled (hyper-)rectangle inscribed within the feasible region in the space of two uncertain parameters.
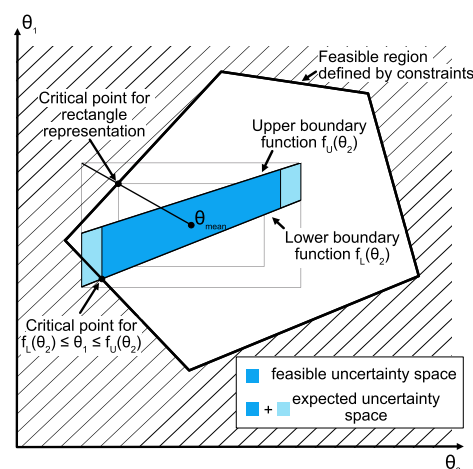
(b) Reformulated model of the uncertainty space based on boundary functions to capture a dependency between the two uncertain parameters.

Figure 1. Visualization of the flexibility index for two uncertain parameters using the (hyper-)rectangle approach (a) and the approach based on boundary functions (b) to model the expected uncertainty space.

dependencies when modeling the expected uncertainty space for flexibility analysis, we[31] recently proposed to utilize upper and lower boundary functions. To define boundary functions, we grouped the uncertain parameters (θ) into independent uncertain parameters (θ_{ind}) and dependent uncertain parameters (θ_{dep}) and reformulated the hyperrectangle uncertainty space to eq 3.

A conceptual illustration of the flexibility index based on upper and lower boundary functions in comparison to the (hyper-)rectangle uncertainty space is shown in Figure 1b. Figure 1b shows that similar to the (hyper-)rectangle uncertainty space, the feasible uncertainty space based on boundary functions is smaller than the expected uncertainty space. However, the flexibility index in the example shown in Figure 1b is closer to 1 compared to that in the example shown in Figure 1a. Consequently, Figure 1 illustrates that the value of the flexibility index is strongly dependent on the chosen model or representation of the expected uncertainty space.
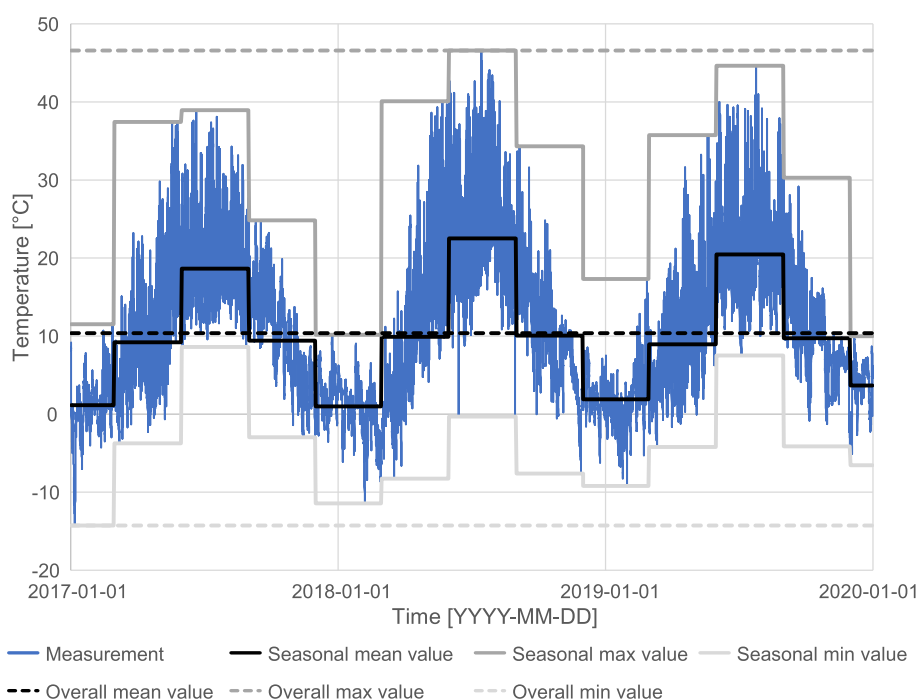
Figure 2. Measurement data of the inlet temperature of the drying air flow at a Swedish pulp mill over a period of three years 2017−2019.

Note that the extreme values (maximum and minimum values of the uncertain parameters, θ_1 and θ_2) are similar in Figure 1a,b, while not all possible combinations are expected to occur in Figure 1b. Consequently, Figure 1 shows that the presence of parameter dependencies requires a different geometric shape to model the expected uncertainty space with good accuracy compared to the hyperrectangle representation (which is a good representation for independent parameter variation). When using boundary functions for modeling the expected uncertainty space, the resulting geometric shape in Figure 1b is an irregular square which can be imagined as a hyperpolygon in higher dimensions. Note that this irregular square shown in Figure 1b is defined by the extreme values of the independent parameter θ_2 and the boundary functions. Our approach based on boundary functions offers additional degrees of freedom when modeling the expected uncertainty space which allows for excluding regions or subspaces of the modeled hyperrectangle uncertainty space in which no operating points are observed/expected. Note that more than a single upper and lower boundary function may be utilized to enhance the modeling of the uncertainty space. We can conclude that a good model representation of the actual uncertainty space is essential to obtain a value of the flexibility index which is a good indicator for the actual flexibility of the process.

2.2.2. Flexibility Analysis Considering Independent Operating Periods. When formulating the flexibility index problem, the uncertainty space should reflect all operating conditions expected during the lifetime of the process. Usually these different realizations of the uncertain parameters are aggregated in a single model/representation, for example, a hyperrectangle for independent uncertain parameters or a hyperpolygon when considering parameter dependencies. The modeled uncertainty space is, thus, independent of time, potentially ignoring that certain realizations of the uncertain parameters may only occur during certain periods of the expected lifetime. A common example where certain realizations of the uncertain parameters occur only during certain periods is the seasonal variation over a

year. Another example is differences in day-time and night-time operation. In these, two examples (some of) the uncertain parameters show a systematic pattern which is also known as seasonality.[36] Note that seasonality can be observed in available data (retrofit problem), while it can also be anticipated (greenfield problems). Since seasonality occurs with a certain frequency (which may be known or anticipated), it allows the data to be divided into several (independent) intervals. For illustration, Figure 2 shows the inlet temperature of the air used for drying at a Swedish pulp mill over a three year period. The data show a clear indication of seasonality. To exemplify the impact of this seasonality, the operating data have been divided into seasonal periods of three months each[d]. Figure 2 shows that (during the respective seasonal periods) the deviations from the seasonal mean values are significantly smaller compared to the deviation from the overall mean value to the (overall) maximum and minimum temperature.

The aforementioned phenomenon can also occur if the system of interest is part of a plant that is adjusted to produce different products, for example, during different production campaigns. An example for such a system is given in the case study section of this work (see Section 4). To summarize, in certain situations (see above), certain parameter values may be expected only during certain operating periods of the entire data sequence. Consequently, when modeling the expected uncertainty space, such independent operating periods should be considered. In our previous studies,[33,34] we proposed to divide the time-horizon of the analysis into several instances representing the different independent operating periods. The expected combinations of the uncertain parameter values can then be allocated to the respective (independent) operating periods, defining an individual uncertainty space for each period.

Based on a second theoretical example (see Figure 3), we wish to illustrate how the structuring of data into independent operating periods, for example, due to seasonality, can impact the result of flexibility analysis, that is, the flexibility index. In Figure 3a, the feasible uncertainty space of the theoretical
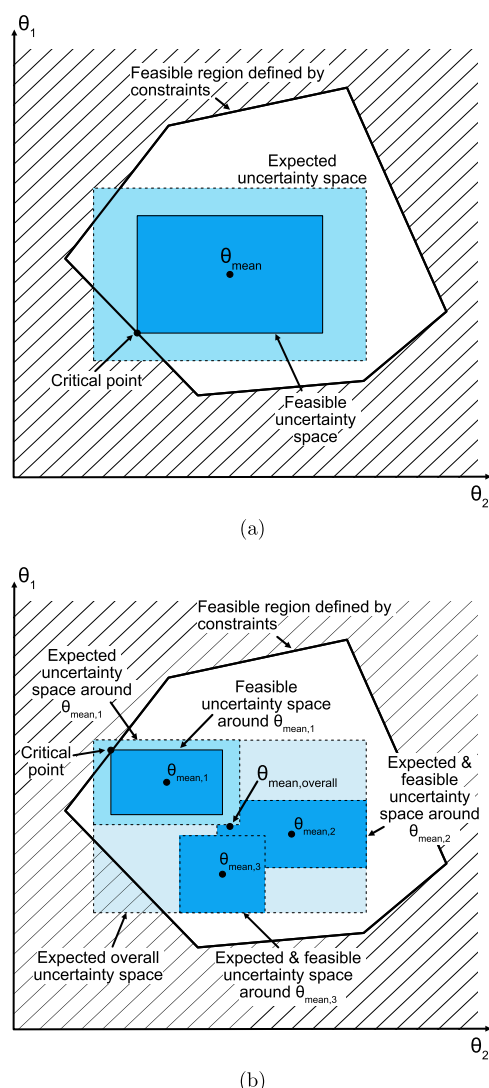
(a)

(b)

Figure 3. Visualization of the flexibility index for two uncertain parameters considering a single overall expected uncertainty space (a) and three independent operating periods observed within the overall expected uncertainty space (b). Note that the (hyper-)rectangle approach was used to model the expected uncertainty space(s).

example is determined by using the expected uncertainty space as defined by all operating periods (modeled as a rectangle). On the other hand, in Figure 3b, this overall expected uncertainty space is divided into three independent (or uncoupled) uncertainty spaces, resulting from independent operating periods. Note that the overall extreme values (maximum and minimum values of the uncertain parameters, θ_1 and θ_2) are still expected to occur also for the representation of uncertainty shown in Figure 3b, but their occurrences are expected in different periods. It should also be noted that similar to the boundary function approach for modeling the expected uncertainty space in the presence of parameter dependencies (see Section 2.2.1), the approach shown in Figure 3b excludes subspaces in the overall uncertainty space in which no operation is expected.

Generally, different approaches can be chosen to model the expected uncertainty spaces of independent operating periods, for example, the hyperrectangle model for independent uncertain parameters or the hyperpolygon model when expecting parameter dependencies. For simplicity, in Figure

3b each of the three uncertainty spaces is modeled using the hyperrectangle approach. Note that operation of the process would be feasible for all expected variations in two of the three periods, while for the period around $\theta_{mean,1}$, the feasible uncertainty space is smaller than the expected uncertainty space.

In a case such as that illustrated in Figure 3b, the flexibility can be assessed by first determining the flexibility index for each independent operating period and then calculating the overall flexibility index, as proposed by Marton et al.[33] For N independent operating periods, the overall flexibility index is found by

$$FI = \min(FI_n) \quad \forall\ n \in 1, 2, ..., N \tag{4}$$

Another cause of independent operating periods can be that the nominal operating point changes due to a (future) uncertain singular/rare event, as discussed by Marton et al.[33] and Langner et al.[34] Examples of such a singular/rare event are a permanent switch of feedstock, a change of operational parameters required to comply with new emission legislation and/or a change in the production rate. All these events may have a lasting effect on the operation of the process in question, and a possible consequence is that the nominal operating conditions change temporarily or even permanently. Additionally, disturbances around the nominal point (current and future) may be present which are comparable to the traditional interpretation of operational uncertainty (which Swaney and Grossmann[19] aimed to analyze by means of the flexibility index). In line with Marton et al.[33] and Langner et al.,[34] we refer to this (more traditional) interpretation of uncertainty as short-term operational disturbances. Note that these short-term operational disturbances may also be affected when the nominal operating point changes, for example, due to a singular event.

In terms of modeling the expected uncertainty space, these types of changes caused by an uncertain singular/rare event show similar characteristics as the independent operating periods caused, for example, by seasonality. More specifically, an uncertain singular or rare event divides the time-horizon of the flexibility analysis into two independent operating periods. If the expected uncertainty space before and after the change can be quantified with high certainty, the flexibility of the respective process can be evaluated by calculating the flexibility index for each individual uncertainty space to eventually determine the overall flexibility index (see eq 4). On the other hand, such singular/rare events usually come with high levels of uncertainty, and the impact on both the nominal operating conditions as well as the expected uncertainty space before and after the change must be estimated.

To illustrate such an uncertain singular/rare event, in Figure 4, a third theoretical example shows a situation in which the nominal value of one of two uncertain parameters is expected to change. The new nominal value after the singular/rare event is uncertain, but we assume that the change has an upper limit. Furthermore, Figure 4 shows that the expected uncertainty space for the short-term operational disturbances around the current nominal point is within the feasible region, that is, steady-state flexible operation is possible at the current nominal operating point. On the other hand, Figure 4 shows that the change of the nominal value can result in a situation where the process cannot handle the expected short-term operational disturbances if the (absolute) deviations ($\Delta\theta^-$ and $\Delta\theta^+$) remain constant even when the operating point changes. To identify the maximum feasible change of the nominal value, the flexibility index problem needs to be formulated for different discrete
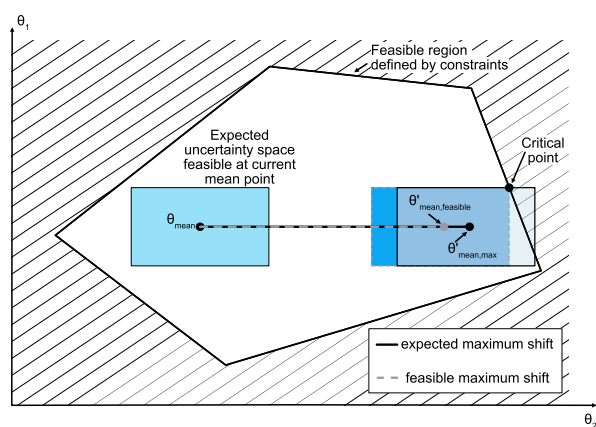
Figure 4. Theoretical example illustrating the flexibility index calculation suggested by Langner et al.[34] to identify the maximum feasible change of the nominal operating point due to an uncertain singular/rare event.

values, that is, discrete points on the line illustrating the expected (maximum) change of the nominal value of θ_2 in Figure 4, and the resulting formulations can then be solved in an iterative scheme. To avoid such a time-consuming iterative scheme, we[34] have previously presented a reformulation of the original flexibility index problem (formulated by Swaney and Grossmann[19]), which allows for considering uncertainty in the nominal value, and which yields the maximum feasible change of the nominal value such that all expected short-term operational disturbances are (exactly) feasible. The solution obtained when applying this reformulation to the flexibility index problem for the theoretical example is shown in Figure 4 and marked as the feasible maximum shift.

In comparison to the first and the second theoretical example, the example shown in Figure 4 includes a new dimension of uncertainty which is the combination of the uncertain change of the nominal value and the short-term operational disturbances (see also Langner et al.[34] where this is referred to as overlaying uncertainty sources). Even for the case of a retrofit where historical measurement data may be available to describe current short-term operational disturbances, this kind of analysis relies on assumptions regarding how the operation of the process will be affected by the rare event; that is, uncertainty in data is added to the uncertainty in operation.

To summarize this section, we have established that certain realizations of the uncertain parameters may occur only during certain periods of the time-horizon considered for the flexibility analysis. By allocating realizations of the uncertain parameters to different independent operating periods and modeling those periods as individual uncertainty spaces, some time-dependency can be enabled in the flexibility analysis.

2.3. Critical Operating Points. In Figures 1a,b and 3a,b, the points at which the feasible uncertainty space coincides with the boundary of the feasible region are marked as critical points. These points define the sizes of the feasible uncertainty spaces (e.g., the rectangle in Figure 1a and the irregular square in Figure 1b). In principle, if there are several independent periods, the feasible uncertainty space of each independent period is limited by different critical points. For example, three critical points could have been marked in Figure 3b, although for reasons of clarity only the critical point determined by the period with the smallest flexibility index was marked. Finally, in Figure 4, the critical point marks the point which defines the maximum

feasible change of the nominal value when requiring that the short-term operational disturbances are also feasible at the nominal operating point after the uncertain event occurred.

Note that at the marked critical points, feasible operation is possible, but even a small deviation from the critical parameter values may lead to infeasibility[e]. This observation is of special interest if some of the constraints forming the feasible region are functions of some design parameters, d, which, for example, can represent the size of certain equipment. If the size of the feasible uncertainty space is limited by constraint(s) depending on d, debottlenecking may be possible by manipulating the respective design parameter(s) enabling a larger feasible uncertainty space (if desired, e.g., if the flexibility index is < 1). Note that generally not all constraint functions which form the feasible region (i.e., $h_i \forall i \in I$ and $g_j \forall j \in J$ in problem 1) are dependent on d. Those constraints, which are dependent on d and therefore can be manipulated by increasing (or decreasing) the equipment size, are hereafter denoted design constraints.

When conducting structural flexibility analysis, design constraints are discarded, and only structural constraints are considered, that is, unlimited equipment size is assumed to be available (see, e.g., Langner et al.[26]). This implies that the value of the structural flexibility index is an upper bound for the general flexibility index, which also considers the design constraints. This is illustrated by an example in Figure 5. For
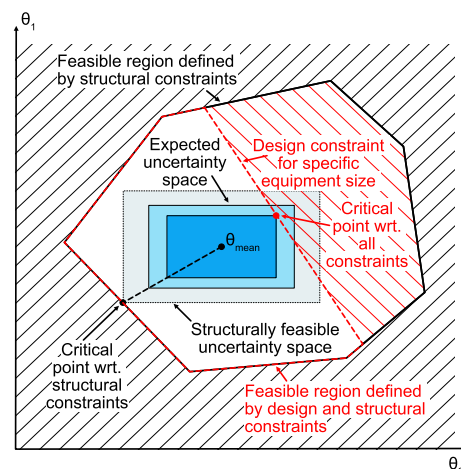

Figure 5. Theoretical example to illustrate the difference between the structural flexibility index where design constraints are discarded and the (general) flexibility index where design constraints are included.

this example, the feasible uncertainty space which is limited by structural constraints is larger than the expected uncertainty space, that is, the structural flexibility index is >rbin1. However, when including the design constraint for the installed equipment size, the feasible uncertainty space is smaller than the expected uncertainty space resulting in a flexibility index which is < 1.

Figure 5 shows that if design constraints are included in the flexibility analysis and the resulting flexibility index is smaller than the structural flexibility index, the design constraints and thereby the corresponding equipment size limit the value of the flexibility index. In this context, Pintarič and Kravanja[15,16] assumed that if, for a given structural design proposal, the structural flexibility index is greater than 1, it is possible to determine at least one set of values for the design parameters which yields a (general) flexibility index of exactly 1. To identify this set of values for the design parameters, Pintarič and Kravanja
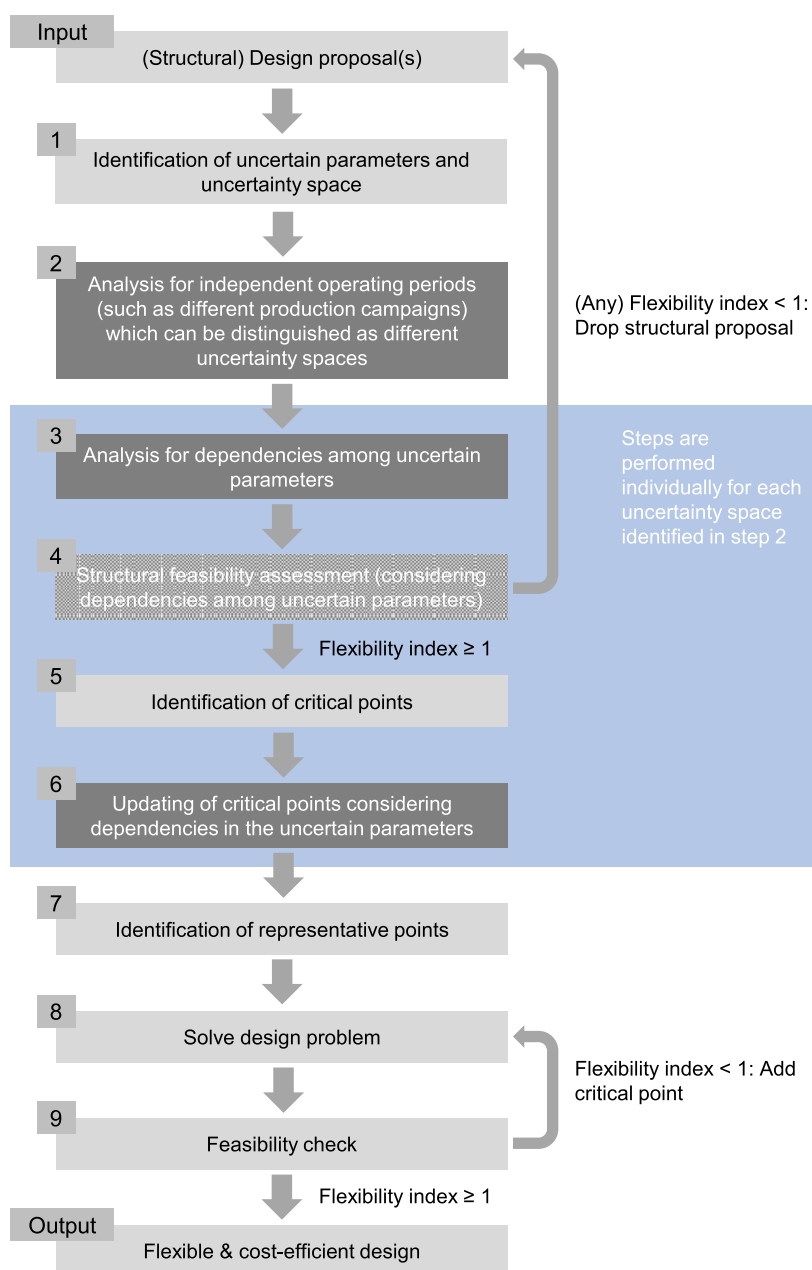
Figure 6. Extended framework for addressing design under uncertainty problems when designing or redesigning chemical processes/plants. Steps 2, 3, and 6 are highlighted since they represent additional steps to the original framework. Additionally, step 4 is highlighted since it involves the calculation of the flexibility index which requires modeling techniques which were not considered in the original framework.

suggested identifying critical operating points, that is combinations of the uncertain parameter values within the expected uncertainty space that require the largest overdesign of process units in order to allow for steady-state flexible operation[f].

In the example shown in Figure 5, we can graphically identify the necessary increase of the equipment size by parallel shifting the design constraint to the right so that the design constraint intersects with the upper right corner point of the expected uncertainty space. This is possible since in this two-dimensional example, the identification of the critical operating point which demands the largest equipment size (upper right corner point) is trivial. However, as Pintarič and Kravanja[15,16] also pointed out, in more complex cases, that is, if more than two uncertain parameters and/or more design parameters are present, several combinations for the design parameter values may yield a

flexibility index of 1 (when design constraints are included). In such cases, the aim should be to identify the unique combination which simultaneously yields the smallest cost and a flexibility index of 1. To achieve this, the authors suggested solving the bilevel optimization problem in which each design variable is individually maximized while simultaneously minimizing a given cost function. Furthermore, the authors suggested several algorithms to solve this bilevel optimization problem, and the interested reader is referred to their work for further information. Note that critical operating points are dependent on the nature and placement of the equipment; that is, for each structural layout, a unique set of critical operating points can be identified.

Finally, note that the algorithms by Pintarič and Kravanja[15,16] were defined for a hyperrectangle representation of the

uncertainty space. However, the examples in Section 2.2 illustrate that the combination(s) of uncertain parameter values which are identified to be the critical operating point(s) strongly depends on the chosen model of the expected uncertainty space. Thus, ignoring factors such as parameter dependencies and independent operating periods when identifying critical operating points could imply that equipment is designed for combinations of uncertain parameter values which are not expected to occur. Particularly conservative modeling of the expected uncertainty space may result in equipment being oversized. To avoid these problems, we suggest a strategy in Section 3 to consider these factors when identifying critical operating points.

3. METHODOLOGY

The framework proposed in this work can be seen as a tool to analyze the results of early design stage screening processes for chemical processes subject to uncertain parameters. The suggested framework extends our previously developed framework[26] to identify flexible and cost-efficient retrofit measures of heat exchanger networks. The different steps of the extended framework are visualized in Figure 6. In the following sections, we present the individual steps of the framework with a special focus on the novel steps suggested in this work.

The proposed framework can be used to evaluate and improve one or several structural design proposals that are provided by the designer. In this way, it is possible to utilize proven design methodologies available for single steady-state operating conditions to generate the basic structure of the design. This setup also allows for incorporating nonquantifiable knowledge such as experience-based heuristics in the design proposals, which can be advantageous for complex industrial case studies. Since the (structural) design proposals are defined prior to the application of the framework, the framework can be utilized to guide both greenfield and retrofit design projects. When following the suggested framework, the design proposals are first checked for feasibility for varying operating conditions (i.e., flexibility analysis) to establish a basis for comparison. Second, the design proposal which achieves the lowest TAC during the entire operating period is identified.

3.1. Identification of Uncertain Parameters. In this work, we consider uncertainty in the input parameters of a given process or system. Since these parameters are determined outside the system's boundary, they cannot be affected by measures within the system. However, recourse actions can be taken to ensure that the operational target of the system of interest is met. These recourse actions are limited by the structure of a design and/or the size and/or capacity of the equipment. Once the uncertain parameters are identified, the expected uncertainty space needs to be modeled, that is, the space in which all possible combinations of operating values for these parameters are expected. As shown in Section 2.2, dependencies in the uncertain parameters as well as independent operating periods should be considered to ensure a good resemblance between the modeled and the actual uncertainty space. Therefore, additional data analysis is necessary, that is, checking for such dependencies and independent operating periods, to model the uncertainty space with good accuracy.

3.2. Analysis for Independent Operating Periods. In this step, the existence of potential independent operating periods is identified. These independent operating periods can be identified manually by analyzing if (potential) causes for independent operating periods are present such as

- seasonality in the operating data (e.g., differences in day-time and night-time operation or seasonal variation),
- different production campaigns,
- uncertain singular/rare events resulting in changes in the nominal operating conditions (overlaying uncertainty sources).

Additionally, automated approaches such as data clustering may be utilized to identify whether the operating points can be divided into separate subsets. Note that automated approaches rely to a larger extent on the availability of (good-quality) operating data, that is, measurements of the uncertain parameters, which is often only the case in retrofit projects. On the other hand, anticipating different operating cases such as those described in the list above is commonly carried out for greenfield design projects where no historic operating data are available. In such cases, also the expected disturbance ranges of the different (presumably) uncertain parameters need to be anticipated.

Since the identified operating periods or data clusters are (ideally) independent of each other, a separate uncertainty space should be defined for each period/cluster, for example, as shown in Figure 3b. The following steps of the framework (steps 3−6) are then performed for each identified uncertainty space.

3.3. Analysis for Dependencies among Uncertain Parameters. For each identified uncertainty space (overall or for each individual independent operating period), the shape of the uncertainty space needs to be approximated. If the uncertain parameters are noncorrelated, the shape can be described by the expected lower and upper bound values yielding several hyperrectangle spaces in the overall space of uncertainty[g]. If dependencies (correlations) are observed or expected, the real uncertainty space would be smaller than the space approximated by the hyperrectangle representation since there will exist regions within the hyperrectangle in which no operating points are expected. By considering dependencies when modeling the uncertainty (sub)spaces, such "empty" regions may be identified and removed, achieving a more accurate approximation of the actual uncertainty (sub)spaces. To identify dependencies between different uncertain parameters, we propose two alternative approaches in this work:

1. Calculation of Pearson's correlation coefficient.
2. Ratio of the area of the polygon convex hull around the 2-d projections of the observed operating points to the area of the rectangle defined by the expected lower and upper bound values.

Pearson's correlation coefficient determines the strength and the direction of the linear relationship between two random variables.[37] When nonlinear dependencies occur, Pearson's coefficient may fail to detect the dependency. For this reason, we developed a complementary methodology (see the second alternative in the list above) to identify dependencies based on the abovementioned observation that the actual uncertainty space reduces (significantly) in volume compared to the hyperrectangle model if parameter dependencies are present. This means that when a dependency occurs between two uncertain parameters, the area of the tightest polygon which covers all observed or expected operating points, that is, the polygon convex hull, should be considerably smaller than the rectangle formed by the expected lower and upper bound values, for example, as illustrated in Figure 7.

The determination of the polygon convex hull can be automated by means of the Quick Hull algorithm (see, e.g.,
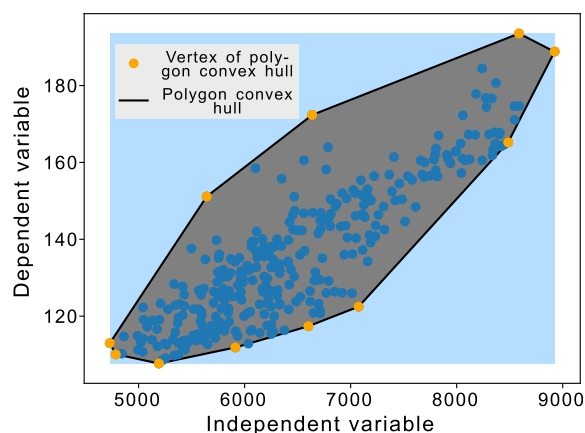
Figure 7. Visual example for two dependent parameters whose polygon convex hull encloses a significantly smaller area than the rectangle defined by the most extreme operating points.

Barber et al.[38]). Once there is an indication that a dependency exists between two or more uncertain parameters, it has to be determined if the dependency indicated by data should be considered when modeling the uncertainty space[h]. Such a decision can be made based on different criteria, for example, the strength of the dependency. Next, for each identified pair of uncertain parameters (which show a dependency/correlation), there is a need to determine which of the two parameters is to be classified as dependent and consequently modeled in terms of the other (independent) parameter, for example, by using boundary functions which are formulated based on the independent parameters. This decision can sometimes be enhanced by background knowledge about the process to identify the origin of the dependencies. Finally, the identified dependencies which should be considered when modeling the expected uncertainty space can be formulated in the following general form

$$\theta_j = f(\theta_{i,1}, \theta_{i,2}, ...), \qquad \theta_i \in \theta_{ind}, \ \theta_j \in \theta_{dep} \qquad (5)$$

3.4. Structural Flexibility Analysis. In the next step, it is analyzed if the structural layout of each design proposal allows for steady-state flexible operation. As previously outlined, this can be done by calculating the flexibility index, discarding all design constraints and considering only structural constraints. If design proposals can be identified that are structurally infeasible for at least some operating points within the uncertainty space (i.e., flexibility index smaller than 1), these design proposals are discarded.

If independent operating periods and/or dependencies in the uncertain parameters are identified in steps 2 and 3, it is necessary to adjust the calculation procedure of the flexibility index compared to the conventional procedure based on a single hyperrectangle uncertainty space. In order to assess the structural feasibility in the presence of independent operating periods, the identified uncertainty spaces of the different periods need to be analyzed individually (as discussed in Section 2.2.2) to eventually determine the overall flexibility index following eq 4. In the special case where an uncertain singular/rare event is expected to lead to permanent or temporary changes in the nominal operating conditions, there is an inevitable need for assumptions regarding the future change of operation (see also Section 2.2.2). If the change of the nominal operating point and short-term disturbances around the new nominal operating point can be predicted with high certainty, the situation is similar

to the other cases of independent operating periods, and the overall flexibility index can be calculated following eq 4. However, in the case of high uncertainty with respect to how the nominal operating point and/or the short-term disturbances are affected by the uncertain future event, it may be desirable to identify the maximum change of the nominal operating point which is structurally feasible (compare Section 2.2.2 and Figure 4). In such a case, our reformulation of the deterministic flexibility index problem can be utilized avoiding the otherwise necessary iterative assessment of different (nominal) operating points (see also our previous work[34] for further information).

If parameter dependencies (correlations between uncertain parameters) have been identified in step 3 (see Section 3.3), then those should be considered when calculating the structural flexibility index. Considering parameter dependencies by means of boundary functions requires more advanced modeling of the uncertainty space compared to the hyperrectangle approach (see Section 2.2.1), which can lead to more complex problems to solve. Therefore, it can be advantageous to know beforehand if considering a given dependency when computing the flexibility index is likely to have an influence on the results. Such knowledge is usually very much dependent on the process, but it is possible to establish guidelines for specific types of processes, such as heat exchanger networks (see Section 2 in the Supporting Information file for further information).

3.5. Identification of Critical Operating Points. For each of the remaining structural design proposals, we suggest identifying the respective critical operating points. Including the critical operating points in the discretized design under uncertainty problem (i.e., problem 1, see also Section 2.1) ensures that each piece of equipment is of sufficient size to allow for steady-state flexible operation within the expected uncertainty space. As shown in Figure 6, we propose to identify the critical operating points for each independent operating period, individually. For this identification procedure, we propose two steps. In the first step (step 5), which is described in this section, we propose to identify the critical operating points for the hyperrectangle representation of the uncertainty space, that is, ignoring potential parameter dependencies. If parameter dependencies are expected, the subsequent step 6 of the framework (see Section 3.6) allows the obtained critical operating points to be updated considering these parameter dependencies.

In this work, we propose a new strategy to identify the critical operating points of a hyperrectangle representation of the uncertainty space. The proposed strategy builds upon the theory of critical operating points described in Section 2.3. More specifically, for the example shown in Figure 5, the solution of the flexibility index problem, θ^*, considering an initial equipment size (installed or estimated) and the critical operating point, θ_c, (demanding the necessary change in equipment size) are on the same diagonal of the rectangular expected uncertainty space. Consequently, it would be possible to identify θ_c by projecting θ^* from the feasible uncertainty space (given the initial equipment size) to the expected uncertainty space, that is, following the aforementioned diagonal from the feasible to the expected uncertainty space. Based on this conclusion, we developed and utilized a two-stage iterative algorithm for identifying the critical operating points of a hyperrectangle representation of the uncertainty space (see also Figure 8):
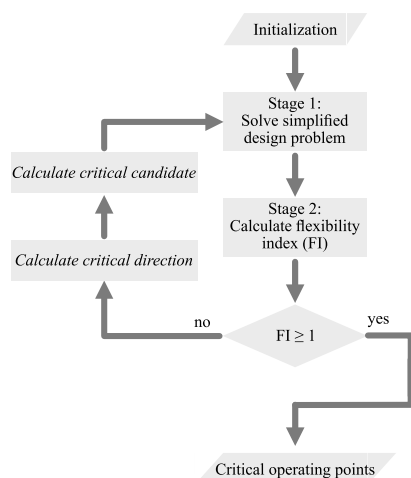
Figure 8. Visual representation of the two-stage iterative algorithm for identifying the critical operating points of a hyperrectangle representation of the uncertainty space.

- Initialization of list of candidates with the solution of the structural flexibility index problem and the nominal operating point of the given uncertainty space
- Stage 1: Solve simplified design problem for the identified candidate(s) to obtain necessary equipment size to be able to operate at all identified candidate(s)
- Stage 2: Calculate flexibility index including design constraint(s) with updated size value(s) obtained by solving the simplified design problem
- If flexibility index ≥ 1: terminate
- If flexibility index <1: update candidate list (according to the scheme presented below) and return to stage 1

In the first stage, the optimal (with respect to cost) design parameter values are identified, which allow for operation at each point from a list of critical candidates (C). At the second stage, it is checked if the obtained design parameter values allow for steady-state flexible operation within the respective uncertainty space, that is, if the list of critical candidates is complete. To avoid a full and computational expansive evaluation of the problem 1 at this stage, a simplified design problem is suggested. This simplified version of problem 1 searches for the set of optimal design parameter values with respect to the investment cost (considering the list of critical candidates, C) but without optimizing the expected operating cost of the system. The simplified design problem is given in problem 6.

$$\min_{d} G(d)$$

s. t.

$$\left.\begin{array}{l} h_i(x, z, d, \theta_c) = 0;\ i \in I \\ g_j(x, z, d, \theta_c) \leq 0;\ j \in J \\ d \geq 0 \\ x, z, d, \theta_c \in \mathbb{R} \end{array}\right\} c \in C \tag{6}$$

Two essential steps of the suggested two-stage algorithm are initialization and updating of the candidate list. We suggest to utilize the solution of the structural flexibility analysis ($\theta^*_{\text{structual}}$) as well as the nominal operating point for the initialization. Note that if the structural flexibility index is $> rbin 1$, $\theta^*_{\text{structual}}$ needs to be

projected to the expected uncertainty space. For additional information on the initialization of the list of critical candidates, the interested reader is referred to Section 3 in the Supporting Information file.

Updating the candidate list is necessary if the flexibility index considering the design specifications obtained during stage 1 is < 1 since this result indicates that the candidate list is incomplete. We suggest the following scheme for identifying a new candidate based on the solution of the flexibility index problem obtained during stage 2 (θ^*, δ^*)

$$d_{i,c} = \begin{cases} \dfrac{\theta_i^*}{\theta_{i,N} + \delta^* \cdot \Delta\theta_i^+} & \forall\ \theta_i^* > \theta_{i,N} \\[4mm] \dfrac{\theta_i^*}{\theta_{i,N} - \delta^* \cdot \Delta\theta_i^-} & \forall\ \theta_i^* < \theta_{i,N} \end{cases} \tag{7}$$

$$\theta_{i,c} = \begin{cases} \theta_{i,N} + d_{i,c} \cdot \Delta\theta_i^+ & \forall\ d_i > 0 \\[2mm] \theta_{i,N} - d_{i,c} \cdot \Delta\theta_i^- & \forall\ d_i < 0 \end{cases} \tag{8}$$

1 Calculate the critical direction for each uncertain parameter according to eq 7
2 Utilize the critical direction to calculate a new candidate according to eq 8

When calculating the critical direction (as described by eq 7) followed by the calculation of the new candidate (eq 8), we utilized the expected positive and negative deviation from the nominal operating point (θ_N) of the respective operating period ($\Delta\theta^-$ and $\Delta\theta^+$). This way, we can project the solution of the flexibility index problem (θ^*) from the feasible uncertainty space to the expected uncertainty space[i]. The new candidate is added to the list of candidates, and the algorithm returns to stage 1.

3.6. Updating of Critical Operating Points Considering Parameter Dependencies. If dependencies in the uncertain parameters are identified in step 3, we proposed to model the expected uncertainty space using boundary functions to ensure a good representation of the actual uncertainty space (see Sections 2.2.1 and 3.3). However, such modeling implies that the previously identified critical operating points may represent combinations of uncertain parameter values which are outside of the hyperpolygon model, that is, the model of the expected uncertainty space defined by the extreme values of the independent parameters and the boundary functions. Such a situation can lead to unnecessary overdesign of equipment since equipment may be designed for operating points that are not expected/observed.

In this work, we developed an iterative updating scheme that aims to determine if the previously identified critical operating points are within the hyperpolygon model of the expected uncertainty space. If the previously identified critical operating points are outside the hyperpolygon model, the critical values of (at least) some uncertain parameters must be adjusted/updated. For this adjustment, it is assumed that the actual critical operating points represent uncertain parameter values at the boundary of the hyperpolygon model. Note that in multidimensional cases, not all independent uncertain parameters, θ_{ind}, are expected to be utilized when defining boundary function models. Thus, it can be assumed that updating/adjusting of the critical parameter values is only required for those uncertain parameters (dependent and independent) that show correlating

trends. The iterative updating scheme is visualized in Figure 9 and given in the following:
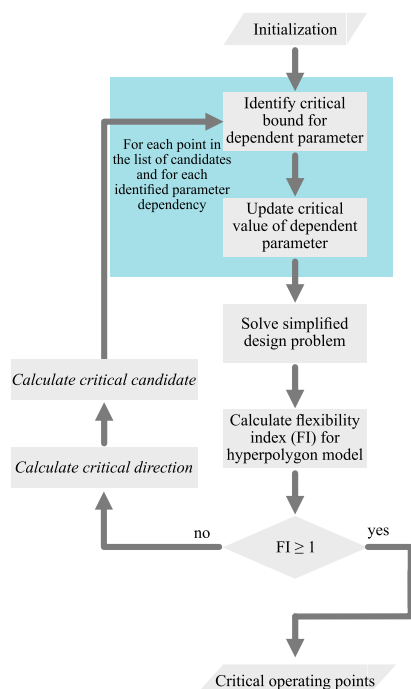

Figure 9. Visual representation of the iterative updating scheme to identify the critical operating points of a hyperpolygon uncertainty space based on the critical operating points of a hyperrectangle uncertainty space.

    1 Initialization with critical point(s) for independent variation and the nominal operating point of the uncertainty space

    2 For each point in the list of critical point(s) and for each identified parameter dependency:

(a) Identify the critical direction ($d_{i,c}$) of the dependent parameter using eq 7 setting $\delta^* = 1$:
- If $d_{i,c} < 0$, the lower bound is critical
- If $d_i > 0$, the upper bound is critical

(b) Evaluate the critical bound at the critical value of the independent parameter and update the critical value of the dependent parameter with the obtained solution

    3 Solve the simplified design problem (i.e., problem 6) for the updated candidate(s)

    4 Calculate the flexibility index considering the hyperpolygon model of the uncertainty space, and including the (convex) design constraint(s) with values for design parameters obtained by simplified design problem
If flexibility index ≥1: terminate
If flexibility index <1: update each critical point in the list according to the updating scheme in Section 3.5 and return to step 2

If certain conditions are fulfilled, we assume that the suggested scheme will terminate after 1−2 iterations. These conditions are

- convex design constraints,
- two-dimensional dependencies meaning that a parameter dependency is described between two uncertain parameters, only (one dependent and one independent parameter),
- one upper and one lower boundary function.

The reasoning behind our assumptions is discussed in Section 4 in the Supporting Information file, where it is also illustrated in Figures S2 and in S3. Note that the corresponding material presented in the Supporting Information file does not constitute a rigorous proof for the validity of the updating scheme. Further note that the suggested scheme may terminate also if the abovementioned conditions are not fulfilled. Alternative formulations may be necessary if the suggested updating scheme does not terminate.

3.7. Identification of Representative Operating Points. In this work, we suggest to decrease the number of scenarios or operating periods considered in the discretized design under uncertainty problem, (i.e., problem 1) by identifying representative operating points. As mentioned in Section 2.1, the defined representative operating points should enable a good approximation of the objective function value (compared to considering all operating conditions) while avoiding issues with the problem complexity.

Note that the identification of representative operating points is not equivalent to the identification of independent operating periods, even though the methodological approaches show similarities (see the next paragraph). There is a fundamental difference concerning the origin and the objective of independent operating periods on one side and representative operating points on the other side. More specifically, when identifying representative operating points, the probability that certain operating conditions will occur is the essential criteria: for example, the mean operating point is the most trivial interpretation of a representative operating point. This means that operating conditions with high probability are to be represented, while operating conditions with low probability are omitted. On the other hand, when defining independent operating periods, the aim is to approximate the actual (i.e., expected or observed) uncertainty space as accurately as possible (see Section 3.2). Commonly, this uncertainty space is influenced or even defined by extreme operating conditions with (usually) low occurrence over the entire operating period and thereby low probability. Therefore, for defining independent operating periods, the essential criteria is that all expected/observed combinations of uncertain parameter values are captured (independent of probability), while all other combinations are excluded. This is further elaborated in Section 5 of the Supporting Information file.

To identify representative operating points, clustering algorithms such as Lloyd's clustering algorithm[39] commonly known as K-means can be utilized to identify centroids among all the operating points (see also, e.g., Jin and Han[40]). Alternatively, mean values for certain time periods (daily, monthly, etc.) may be considered. If an uncertain singular event is expected to affect (future) operation, assumptions about how the operating conditions change are necessary. In this context, note that the exact time-point when the singular event is expected to happen may also be uncertain. Consequently, if cost for the entire lifetime is to be considered in the final design problem, for example, TAC, further assumptions are necessary. In this context, Marton et al.[33] presented a strategy based on different extreme scenarios which can be utilized to obtain a lower and upper bound for the expected objective function value.

3.8. Design Problem. After the critical operating points as well as the representative operating points are identified (CP ∪ OP), the discretized design under uncertainty problem (i.e., problem 1) can be solved for the remaining structural design proposals[j]. The solution yields the optimal design parameter
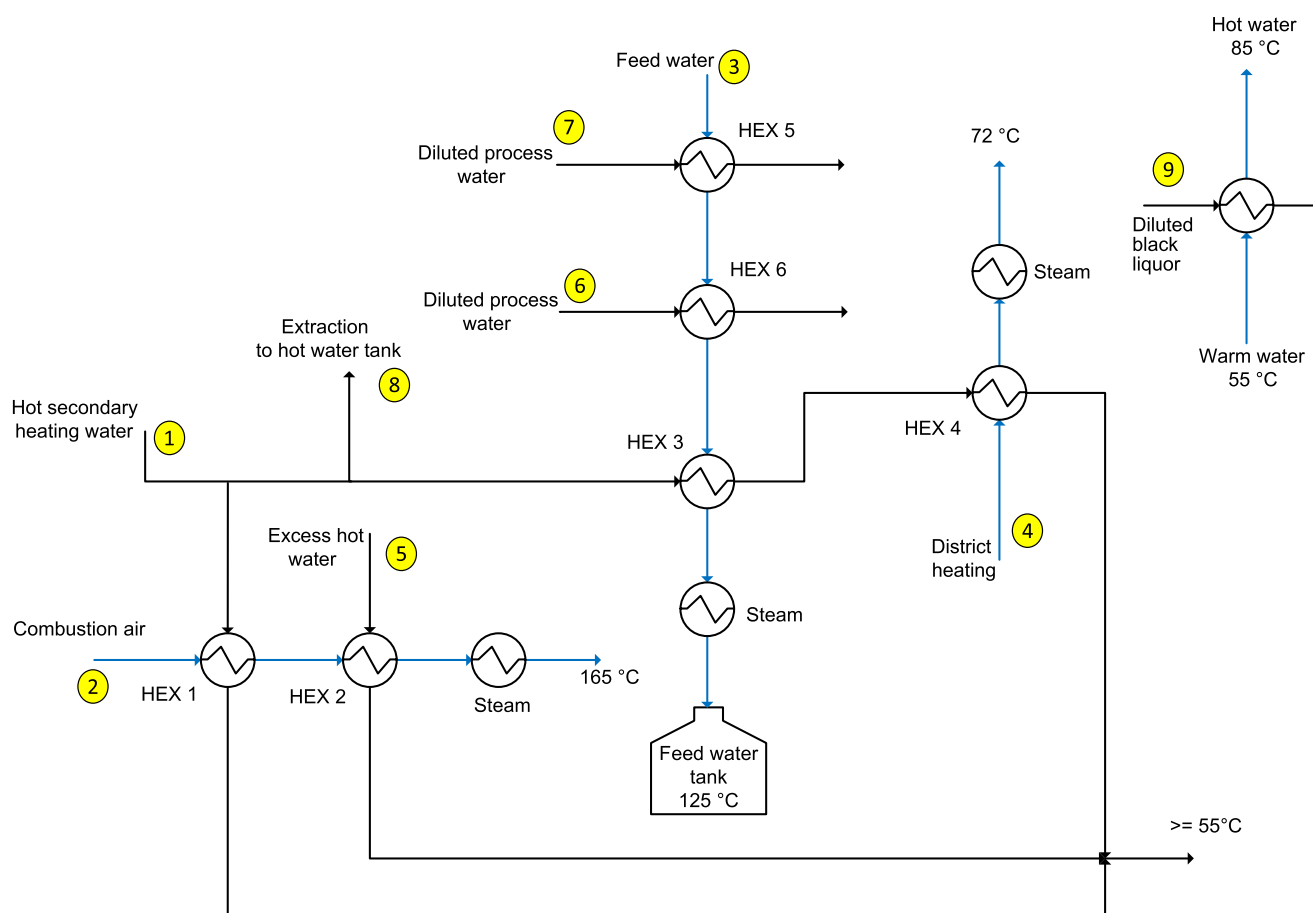
Figure 10. Structural layout (process flow diagram) of the mill subsystem before retrofitting. Locations where process conditions (heat capacity flow rates or supply temperatures) are determined outside the system's boundary are highlighted in yellow.

values, as well as an approximation of the expected objective function value, usually TAC, by optimizing the settings for the control variables. In this context, data for the investment and installation cost of the equipment as well as data for the operating costs for the process and the maintenance costs of the equipment are necessary. Additionally, possible cash flows generated by selling products or by avoiding expenses need to be considered.

3.9. Feasibility Check. The final step of the framework is a feasibility assessment to check that the obtained values for the design parameters allow for steady-state flexible operation within the expected uncertainty space, that is, a final validation of the previously identified critical operating points. In this context, the flexibility index, including design constraints, may be utilized considering the identified independent operating periods and possible parameter dependencies. Alternatively, the optimal operation problem (see problem 9) may be solved for discrete operating points (OP_{all}). These operating points may be generated by means of sampling methods. Also, historical measurement data, if available, may be utilized. Problem 9 is a single-period optimization problem to identify the optimal control variable settings for fixed design parameter values.

$$
\left.\begin{aligned}
&\min_{z_s} C_s(d, z_s, \theta_s) \\
&\text{s. t.} \\
&h_i(x_s, z_s, d, \theta_s) = 0; \; i \in I \\
&g_j(x_s, z_s, d, \theta_s) \leq 0; \; j \in J \\
&x_s, z_s, d, \theta_s \in \mathbb{R}
\end{aligned}\right\} s \in OP_{all}
$$

(9)

In general, the (N)LP given in problem 9 is significantly easier to solve compared to the discretized design under uncertainty problem (i.e., problem 1) since the number of optimization variables is lower. If a feasible solution ($C_s(d, z_s, \theta_s) \in \mathbb{R}$) can be identified for each operating point ($s \in OP_{all}$), the design parameters derived in step 8 allow for operation at all expected/observed operating points. If the final feasibility check reveals that operation is not possible at certain operating points, that is, that parts of the uncertainty space(s) are outside of the feasible region, then the previously identified set of critical operating points is incomplete. In such a case, the infeasible operating points need to be added to the previously identified set of critical operating points, and steps 8 and 9 need to be repeated until a final indication of feasibility of the design is achieved[k].

4. INDUSTRIAL CASE STUDY

In this section, the benefits of the proposed framework are illustrated by applying it to an industrial case study of a heat exchanger network retrofit. The studied system is a simplified
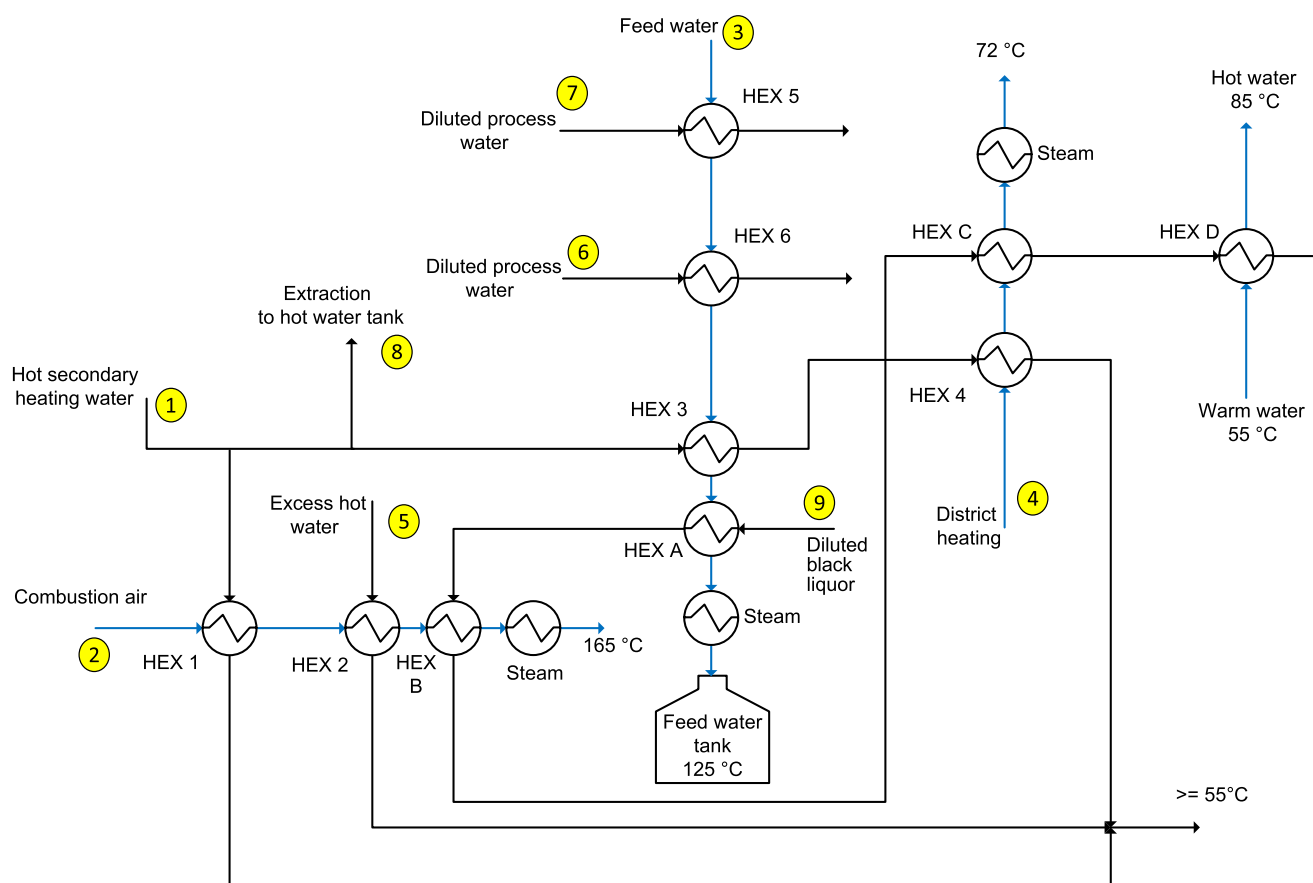
Figure 11. Structural layout (process flow diagram) of the subsystem after retrofitting. Locations where process conditions (heat capacity flow rates and/or supply temperatures) are determined outside the system's boundary are highlighted in yellow.

representation of a subsystem of the secondary heating system of a Swedish pulp mill. In comparison to the real system installed in the mill, in the simplified system, several heat exchangers (HEX) have been aggregated. The system layout is shown in Figure 10.

In the system, three cold process streams (combustion air, feedwater, and district heating water) are heated to their target temperatures by different heat sources, namely, hot secondary heating water, excess hot water, and hot diluted process water. Although a substantial amount of heat can be recovered by HEXs, utility steam is needed to ensure that the target temperatures of the cold process streams are met.

To illustrate a situation in which the proposed framework can be utilized to guide a retrofit project related to the presented system, a fictitious scenario is assumed, where the three steam heaters in the system need to be replaced. This assumed scenario creates an economic incentive to investigate opportunities for reducing steam demand (low and medium pressure steam) by increasing heat recovery. As shown in Figure 10, a diluted black liquor stream is currently utilized to upgrade warm secondary heating water to hot secondary heating water. The full thermal potential is, however, not utilized since the supply temperature of the diluted black liquor flow (133 °C) is significantly higher than the target temperature of the upgraded hot water. Estimations by mill process engineers indicate that the amount of hot water generated by using this heat source corresponds to roughly 50% of the heat capacity flow rate of the diluted black liquor stream. Consequently, the integration of heat recovery from the black liquor with the given system could potentially

lead to reduced steam demand and thereby a lower investment requirement for new steam heaters.

The steam used in the process is extracted from the mill's extraction-condensing turbine. The steam flow to the turbine is high-pressure steam generated in the recovery boiler. Since the amount of black liquor fired in the boiler is constant, the amount of high-pressure steam is also constant. There are several options for harnessing the benefits of the reduced steam demand in the investigated system. If the capacity of the condensing section of the turbine is sufficiently large, then the steam that is not required for process heating can be expanded in the condensing section to increase electricity production. Alternatively, steam extracted at low and medium pressure levels that is no longer needed by the mill may be utilized as a heat source for other processes. In this case study, only changes in electricity production are considered. Several flow arrangements are possible when integrating the black liquor cooling in the given system as well as resequencing of the existing HEXs. However, since the case study is used for illustrative purposes, only one flow arrangement is investigated, as shown in Figure 11. In this design proposal, four new HEXs (HEX A, B, C, and D) are installed. HEX A, B, and C are placed immediately upstream of the respective steam heater of each of the three cold process streams.

Data regarding the heat-transfer areas and the heat-transfer coefficients of the already installed HEX units were not available. To account for installed heat-transfer area and effective heat-transfer coefficients, individual minimum temperature differences in the exchangers were assumed. The exact values of the

assumed minimum temperature differences, as well as additional assumptions and data regarding the mathematical modeling of the existing and new HEX units are listed in Section 6 in the Supporting Information file. The mathematical model of the retrofitted system, including definitions of state and control variables as well as uncertain parameters (see Section 4.1), is provided in Section 7 of the Supporting Information file together with the system design parameters and constraints. The system model includes bilinear and other nonlinear terms, which may lead to difficulties when calculating the flexibility index based on the active set approach (see Grossmann and Floudas[28] for more information). On the other hand, all constraints considered are representative of typical constraints in regular heat exchanger network superstructures, and Floudas and Grossmann[24] concluded that the active set approach is applicable for systems described by similar constraints. Additionally, the global NLP framework BARON[41] (Version 21.1.13) was used for solving all optimization problems in this work. This framework guarantees the global solution of nonconvex optimization problems within the given variable bounds and for a manually chosen epsilon tolerance. All optimization problems presented in this work were formulated using the PYTHON package Pyomo[42] (Version 5.6.6) and translated to GAMS (Version 37.1.0) using the interface between Pyomo and GAMS.[43]

4.1. Identification of Uncertain Parameters.
In Figures 10 and 11, nine locations (marked in yellow) within the current and the retrofitted system were identified where process conditions, more specifically heat capacity flow rates and/or (supply) temperatures, are determined outside the system's boundary. For these measurement locations, 3694 hourly measurement instances were available. The available data were adjusted to reflect normal operating conditions. For this filtering step, it was decided to only consider operating points at which the mill's recovery boiler was operating at a minimum of 75% of its nominal load, and 3415 operating points were thus retained for the analysis. In the Supporting Information, the mean value as well as the maximum positive and the maximum negative deviation from this mean value for all uncertain parameters are provided.

4.2. Identification of Independent Operating Periods.
In general, different factors and strategies may be considered when analyzing the operating data for independent operating periods as outlined in Section 3.2. Since the secondary heating system of the pulp mill is affected by ambient conditions, for example, through the fresh water inlet temperature, seasonal variations over the year are likely. However, since the time period considered in the case study is relatively short, no seasonal effect on the data was expected. However, the mill produces softwood and hardwood pulp in campaigns according to a regular schedule with roughly a 3 weeks production of softwood pulp followed by a 1 week production of hardwood pulp. The considered 3415 operating points were classified according to the processed raw material, resulting in 2481 operating points for five different softwood campaigns and 934 operating points for five different hardwood campaigns. Consequently, two independent uncertainty spaces were identified, corresponding to the two independent operating periods representing softwood and hardwood campaigns. For each raw material and each uncertain parameter, the mean value as well as the maximum positive and the maximum negative deviation from this mean value is listed in the Supporting Information file.

4.3. Analysis for Parameter Dependencies.
Both identified uncertainty spaces were analyzed for parameter dependencies. As mentioned in Section 3.3, the presence of parameter dependencies reduces the volume of the actual uncertainty space compared to the hyperrectangle model based on lower and upper bound values. In Section 3.3, two approaches were presented to identify parameter dependencies, and the results of each approach can be visualized using heat maps. In Section 10 of the Supporting Information file, for each uncertainty space heat maps indicate the estimated strength as well as the sign of a dependency between each pair of uncertain parameters according to the two approaches.
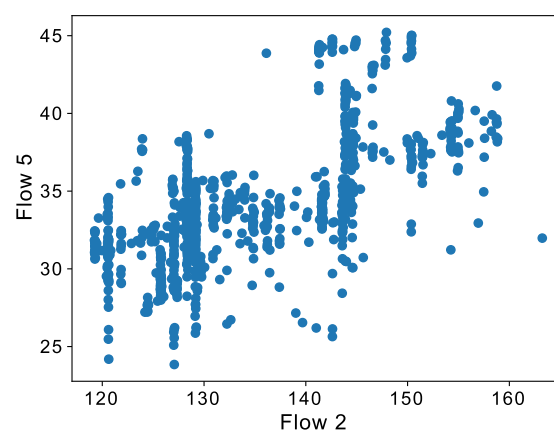
In Section 2 of the Supporting Information file, we present guidelines to identify if a (given) dependency between two uncertain parameters is likely to have an influence on the result of the flexibility index for the specific application of heat exchanger network design. These guidelines are based on the effect of variations in the temperature and heat capacity flow rate on the total availability and demand of heat from hot and cold streams. The heat maps shown in the Supporting Information file were specifically analyzed for such dependencies considering that reaching the target temperatures of the three different (cold) process streams defines the operational target of the given subsystem. Eventually, three parameter dependencies were identified, of which two are present in both softwood and hardwood campaigns while a third parameter dependency was observed during hardwood campaigns, only. These dependencies are listed in the following and shown in Figure 12 and 13:

1. The heat capacity flow rate of excess hot water correlates with the heat capacity flow rate of the combustion air (positive correlation coefficient during softwood and hardwood campaigns, see Figures 12a and 13a).
2. The heat capacity flow rate of the extraction flow to the hot water tank correlates with the heat capacity flow rate of the (main) hot secondary heating water flow (positive correlation coefficient during softwood and hardwood campaigns, see Figures 12b and 13b).
3. The supply temperature of the feedwater correlates with the heat capacity flow rate of the combustion air (positive correlation coefficient during hardwood campaigns, only, see Figure 13c).

The first and third parameter dependency can potentially have an impact on the flexibility index, as identified by following the guidelines listed in Section 2 in the Supporting Information file. For the second parameter dependency, the guidelines presented in the Supporting Information file could not be applied since the extraction flow to the hot water tank is not clearly defined as a hot or cold stream within the investigated system. More precisely, Figure 11 shows that the extraction flow to the hot water tank is a split branch of the main hot water flow of the secondary heating system. The extraction flow leaves the system and thereby reduces the flow rate of the main hot secondary heating water flow which is utilized to provide heat to the feedwater stream and district heating water. Therefore, variations in the flow rate of the extraction flow would have opposite implications as flow rate variations in hot streams; that is, an increased flow rate (of the extraction flow) would reduce the (net) heat available in the system. Thus, a positive correlation between the extraction flow and the main flow of the hot secondary heating water can influence the flexibility index since high extraction flow rates are only expected when the flow rate of the main hot secondary heat water flow is also high
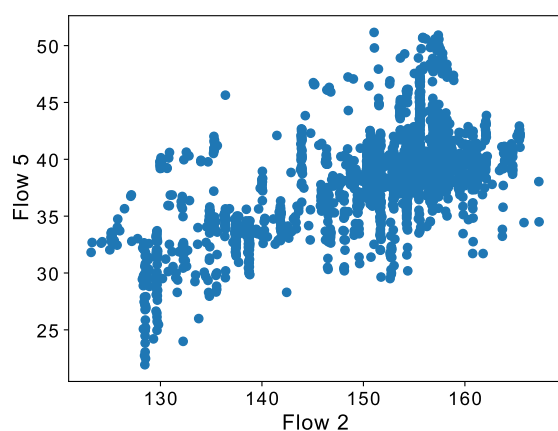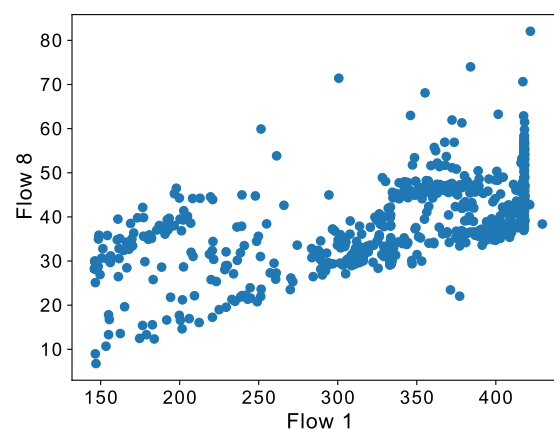
(a) Correlation between heat capacity flow rate of excess hot water and heat capacity flow rate of the combustion air.
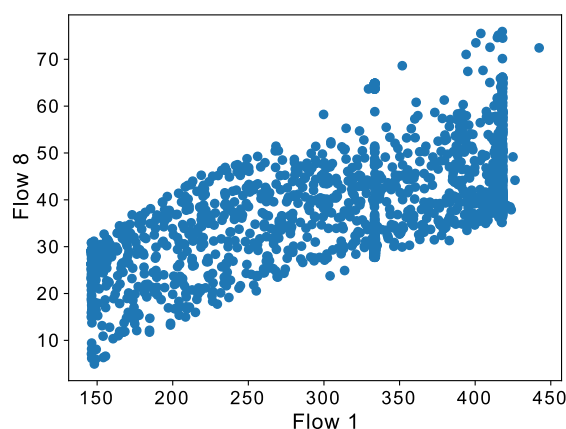
(b) Correlation between heat capacity flow rate of the extraction flow to the hot water tank and heat capacity flow rate of the (main) hot secondary heating water flow.

Figure 12. Operating points and parameter dependencies for the softwood campaigns.

(and vice versa). Consequently, worst-case operating conditions where the extraction flow is high while the main flow of the hot secondary heating water is low can be excluded from the analysis.

4.4. Structural Flexibility Analysis. In the structural layout of the retrofit proposal shown in Figure 11, there is no structural feasibility issue as long as sufficient amounts of steam are available to always be able to provide the heat required by the three cold process streams (combustion air, feedwater, and district heating). Sufficient availability of steam can be assumed since there were no indications that the steam availability would be affected by the retrofit[1]. Also the heating of the secondary heating water through heat recovery from the diluted black liquor stream is not critical (based on the information that the heat capacity flow rate of the water stream corresponds to only 50% of the liquor rate). The given retrofit proposal thereby fulfills the requirement to be structurally feasible, and it is possible to proceed with step 5 of the framework.

To illustrate the application of structural flexibility analysis, we utilized it to investigate the impact of a (small) modification

(a) Correlation between heat capacity flow rate of excess hot water and heat capacity flow rate of the combustion air.
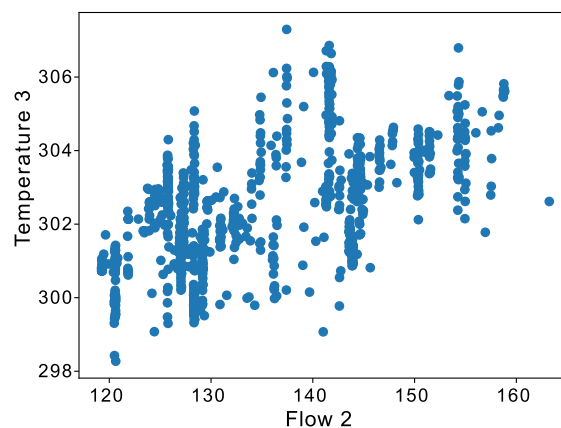

(b) Correlation between heat capacity flow rate of the extraction flow to the hot water tank and heat capacity flow rate of the (main) hot secondary heating water flow.

(c) Correlation between supply temperature of the feed water and heat capacity flow rate of the combustion air.

Figure 13. Operating points and parameter dependencies for the hardwood campaigns.

of the suggested structural retrofit proposal. An interesting modification would be to investigate if the steam heater for the district heating water flow can be removed since the target temperature of the district heating system is lower than the supply temperature of the hot secondary heating water. By means of structural flexibility analysis, we can identify if the heat available in the hot secondary heating water and the diluted black liquor is sufficient to provide the necessary heating to the district heating system (at all times). The results of this analysis are summarized in Table 1. When dependencies were

Table 1. Results of Structural Flexibility Analysis for the Modified (Structural) Retrofit Proposal[a]

raw material	dependencies considered	structural flexibility index
softwood		0.96
	1 and 2	1.05
hardwood		1.26
	1 and 2 and 3	1.29

[a]In contrast to the initial (structural) retrofit proposal, no new steam heater is considered for controlling the target temperature of the district heating water flow.

considered, one upper and one lower boundary function were used to improve the representation of the uncertainty space. The active set approach[28] in combination with optimization framework BARON[41] was chosen to calculate the flexibility index.

The results illustrate the impact when parameter dependencies are ignored during the analysis. More precisely, according to the analysis the modified structural retrofit proposal would be structurally infeasible for the operating data expected during softwood campaigns. However, when considering the identified parameter dependencies when modeling the uncertainty space (by means of boundary functions), the structural flexibility index is ≥ 1 for both softwood and hardwood campaigns. Consequently, operating conditions that fall within the hyperrectangle uncertainty space, but not within the more accurate space defined by boundary functions, are infeasible. We can conclude that removing the steam heater of the district heating system can be viable option with respect to structural flexibility analysis. Note that when considering the identified parameter dependencies when modeling the uncertainty space (by means of boundary functions) for hardwood campaigns, the structural flexibility index also increases. However, since the structural flexibility index for hardwood campaigns is also ≥ 1 when ignoring parameter dependencies, the effect of this observation is irrelevant for the analysis.

In addition to the analysis of (small) design modifications, we utilized structural flexibility analysis to quantify the minimum total steam savings potential of the retrofit design proposal. Such information can be important if the "saved" steam is intended to be used to provide the base-load of another (new) process. To identify the minimum total steam savings potential, we solved the (structural) flexibility index problem, including an upper bound for the total steam available. The upper bound of available steam was changed iteratively to identify the minimum steam demand for which the value of the flexibility index is exactly 1. The results of this analysis are summarized in Table 2. When dependencies were considered, one upper and one lower boundary function were used to improve the representation of the uncertainty space. The active set approach[28] in combination

Table 2. Minimum Total Steam Demand of the Installed and Retrofitted Systems for the Different Raw Materials[a]

raw material	system design (original/retrofit)	dependencies considered	minimum total steam demand for reaching FI = 1
softwood	original		56 MW
		1 and 2	54 MW
	retrofit		52 MW
		1 and 2	50 MW
hardwood	original		40 MW
		1 and 2 and 3	39 MW
	retrofit		36 MW
		1 and 2 and 3	35 MW

[a]Note that only structural constraints were considered.

with the optimization framework BARON[41] was chosen to calculate the flexibility index.

The values for the minimum total steam demand in Table 2 represent lower bounds for the steam demand in all three steam heaters (compare Figures 10 and 11) when assuming that an unlimited heat-transfer area is available. The results illustrate the impact when parameter dependencies are ignored during the analysis. More precisely, according to the analysis both the original and the retrofit layout are estimated to require a higher steam demand when dependencies are ignored. Consequently, operating conditions that fall within the hyperrectangle uncertainty space, but not within the more accurate space defined by boundary functions, are responsible for this overestimated steam demand. In other words, the difference is caused by operating conditions representing parameter values that can be expected individually, but never in combination.

4.5. Identification of Critical Operating Points Assuming No Parameter Dependencies. The suggested two-stage algorithm for identifying critical operating points presented in Section 3.5 was utilized to identify critical operating points for both hardwood and softwood campaigns. The optimization framework BARON[41] was utilized to solve the simplified design problem in stage 1 of the algorithm. To calculate the flexibility index during the second stage, the active set approach[28] in combination with BARON[41] was chosen. In both cases, more than one iteration was necessary.

For softwood campaigns, the algorithm terminated during the second iteration after it utilized the solution of the flexibility index obtained during the first iteration to add a second critical point to the list. For hardwood campaigns, a total of three iterations was necessary, resulting in a list of three critical points.

Note that for both lists of critical points (softwood and hardwood campaigns) it was not possible to obtain a (general) flexibility index ≥ 1 (design constraints were included) when the simplified design problem (i.e., problem 6) was solved as a single-period problem, that is, when design parameter values were utilized which were obtained for individual points in the lists of critical points. This indicates that either the identified critical points in a list need to be combined using a set covering algorithm as used by for example Pintarič and Kravanja[16] or that only the consideration of all identified critical points yield values for the design parameters which allow for steady-state flexible operation (within the hyperrectangle uncertainty space). The list of critical points for the two different uncertainty spaces (softwood and hardwood campaigns) is provided in the Supporting Information file.

4.6. Updating of Critical Operating Points Considering Dependencies in the Uncertain Parameters. The scheme for updating the list of critical points identified for a hyperrectangle model when considering parameter dependencies presented in Section 3.6 was utilized. Again, the optimization framework BARON[41] was utilized to solve the simplified design problem while the active set approach[28] in combination with BARON[41] was chosen to calculate the flexibility index. Both lists, that is, the list considering operating conditions during softwood campaigns and the list considering operating conditions during hardwood campaigns were successfully updated after one iteration. This underlines the expected efficiency of the proposed scheme.

Note that before utilizing the updating scheme, the conditions listed in Section 3.6 were checked partially. With respect to the considered dependencies, only 2-D dependencies were identified (compare Section 4.3) which were modeled using a single upper and a single lower bound. The convexity of the design constraints, which depends on an approximation of the logarithmic mean temperature difference (see the Supporting Information file), was not rigorously checked. Nevertheless, the two lists of critical points obtained for softwood campaigns and hardwood campaigns could be updated (i.e., the updating scheme terminated successfully). The two lists are presented in the Supporting Information file.

4.7. Identification of Representative Operating Points. As previously stated, 2481 and 934 operating points were available for softwood and hardwood campaigns, respectively, representing typical operating conditions for the two campaigns. Including all available operating points in the discretized design under uncertainty problem (i.e., problem 1) would increase the problem size dramatically and the computational burden for solving it may be unacceptable. To reduce the number of operating points while not (or at least not significantly) reducing the accuracy of the obtained result, representative operating points were identified using Lloyd's clustering algorithm[39] commonly known as K-means.

When utilizing the K-means algorithm, the number of clusters must be specified in advance. Consequently, there may be an influence of the number of clusters on the result of the problem 1. To investigate this influence, problem 1 was solved for different numbers of clusters (Section 4.8). Note that clustering was performed on the entire data set and not on the separated operating data for the individual raw materials.

4.8. Design Problem. Problem 1 was solved, including the previously identified critical operating points in the constraints and different numbers of representative operating points in both constraints and the objective function. The optimization framework BARON[41] was utilized to solve the resulting multiperiod NLP. To illustrate the influence of the critical operating points as well as the effect of considering the parameter dependencies for updating the critical operating points, three cases were considered for each set of representative operating points:

1 No critical operating points and only representative operating points.

2 Critical operating points of uncertainty space modeled as hyperrectangle as well as representative operating points.

3 Critical operating points of uncertainty space modeled as hyperpolygon using boundary functions as well as representative operating points.

The results of these calculations are presented in Section 4.10. Note that the modification discussed in Section 4.4 (i.e., the removal of the steam heater for the district heating water flow) is implicitly considered since the heat-transfer area of each (new) HEX unit can be set to 0. The cost functions used to quantify the investment cost of the new HEX units are provided in Section 8 of the Supporting Information file together with the assumed economic conditions that were used to calculate the capital recovery factor.

Since exact data for the current steam demand of the investigated subsystem were not available, it was not possible to quantify the steam savings resulting from increased heat recovery to calculate the potential revenue from increased electricity production. Instead, to compare the different cases (see the list above) with each other, we calculated the loss of revenue which occurs when using steam for heating in the investigated subsystem instead of producing and selling electricity. More precisely, it was assumed that utilizing steam in the steam heaters increases the TAC (as a loss of revenue) since this steam cannot be utilized to produce electricity. This loss of revenue was calculated by approximating the lost electricity production when using steam for heating multiplied with the representative electricity price for the respective representative period (obtained by utilizing the K-means algorithm, see Section 4.7). To approximate the lost electricity production, effective electrical efficiencies were assumed to estimate the electricity generation by means of expanding medium and low pressure steam to the condensing level, as shown in the Supporting Information file. For the electricity price, scenario data generated by Göransson et al.,[44] which estimates the electricity price with an hourly resolution in the year 2030 in the Nordic countries, was utilized[m].

4.9. Final Feasibility Check. Due to the previously mentioned nonlinear terms in the mathematical model of the subsystem as well as the potentially nonconvex design constraints, the final feasibility check is important to ensure that the identified critical points (for hyperrectangle and hyperpolygon modeling approach) indeed ensure operability for all expected/observed operating points. Consequently, the calculation of the flexibility index is not chosen for evaluating the final feasibility of this case study (since it was already utilized in steps 4, 5 and 6). As an alternative approach, the single-period operational optimization model (i.e., problem 9) was evaluated for all 3415 operating points. The evaluation of each operating point took on average less than 0.3 s on an Intel Core i7-11800H with 32 GB of installed RAM. The final feasibility check revealed that for the investigated cases where critical operating points were considered when solving problem 1 (cases 2 and 3, see Section 4.8), operation at each of the 3415 operating points was possible. However, if only representative operating points were considered, operation was only possible for a fraction of these operating points. The fraction increased with an increase in the number of representative operating points, as outlined in Section 4.10.

4.10. Numerical Results of Case Study. The discretized design under uncertainty problem (i.e., problem 1) was solved considering different numbers of representative operating points, namely, 1 (mean value), 5, 25, 50, and 75. For each set of representative operating points, the three different cases that differ with respect to the critical operating points included in the problem formulation were considered (case 1: only representative operating points; case 2: representative and critical operating points for hyperrectangle approach; case 3:

representative and critical operating points for hyperpolygon approach; compare Section 4.8). As mentioned in Section 4.9, feasible operation for all (3415) discrete operating points was achieved only when critical operating points were included (cases 2 and 3).

As mentioned previously, increasing the number of representative operating points allows for a better approximation of the expected objective value: here: the expected TAC. On the other hand, we assume that the influence on the obtained solution of adding more representative operating points is reduced for higher numbers of representative operating points. Figure 14 shows how the expected TAC is changed by a stepwise
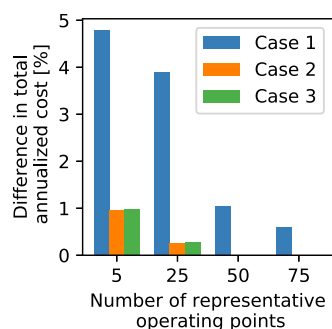

Figure 14. Difference in the expected TAC (in percentage) obtained for a certain set of representative operating points compared to the next lower set with fewer representative operating points, e.g., the bars for the set of 25 representative operating points show how much the expected TAC increase compared to the expected TAC obtained when considering (only) 5 representative operating points.

increase in the number of representative operating points (from 1 to 5, from 5 to 25, etc.) for the three cases. For example, the bars for the set of 25 representative operating points show how much the expected TAC changed compared to the expected TAC obtained when considering (only) 5 representative operating points.

For all three cases, the largest change in the expected TAC can be observed when comparing the results obtained for the mean value and 5 representative operating points. When increasing the number of representative operating points in the set (here: 25, 50, and 75), the difference in the expected TAC decreases. Note that for cases 2 and 3, only a marginal difference in the expected TAC can be observed when the number of representative operating points in the set is increased beyond 5. This observation indicates the strong influence of the critical operating points which are included in cases 2 and 3. Moreover, the change in the expected TAC becomes negligible when comparing the solutions obtained for 25, 50, and 75 representative operating points. Consequently, we can conclude that for cases 2 and 3, a set of 25 representative operating points allows for a good approximation of the expected TAC. For case 1, Figure 14 indicates that a significantly higher number of representative operating points must be considered to obtain a good approximation of the expected TAC. Note that for case 1 even the largest set of representative operating points considered in this work (75 points) may not be representative; that is, the expected TAC may change remarkably when considering sets with more than 75 representative operating points. This observation illustrates that due to the absence of critical operating points, operating conditions with lower probability (which are increasingly represented in large(r) sets of representative operating) have a measurable effect on the

solution. Additionally, we can conclude that the additional effort for identifying critical operating points can be justified since higher numbers of representative operating points lead to increasing problem complexity.

Moreover, the expected TAC as well as the optimized design specifications which were obtained for cases 2 and 3 were analyzed and compared. Differences between cases 2 and 3 are of special interest since such differences originate from the different approaches to model the uncertainty space. Note that the final feasibility check revealed that steady-state flexible operation is possible for both cases independent of the chosen set of representative operating points. Table 3 presents the

Table 3. Comparison of Optimal Design Specifications and Expected TAC when Modeling the Expected Uncertainty Space as a Hyperpolygon (i.e., when Considering Parameter Dependencies: Case 3) or as Hyperrectangle (i.e., when Assuming that No Parameter Dependencies are Present: Case 2)[a]

heat-transfer area $[m^2]$	case 2	case 3	abs (difference)
HEX A	1098.47	1194.86	96.39
HEX B	602.88	0	602.88
HEX C	0	0	0
HEX D	360.47	318.3	42.17
feed water heater	1107.75	1034.16	73.59
district heating water heater	135.8	135.8	0
combustion air heater	11,973.32	11,846.11	127.21
total heat-transfer area	15,278.69	14,529.23	749.46
expected total annualized cost $[€]$	6,484,349	6,313,190	171,159

[a]Results obtained with 25 representative operating points.

results obtained for cases 2 and 3 for the set of 25 representative operating points[n]. Significant differences between the cases can be identified for HEX B. For the other HEXs, no difference or only a small difference ($\leq 10\%$) was identified. Nevertheless, the numbers show that the total optimized heat-transfer area is lower when the uncertainty space is modeled as hyperpolygon (i.e., when considering parameter dependencies: case 3) compared to the hyperrectangle approach (i.e., when assuming that no parameter dependencies are present: case 2). Additionally, the total heat-transfer area is distributed differently among the different HEX units for cases 2 and 3.

Although the differences are rather low for this case study, Table 3 shows a 5% lower requirement of total heat-transfer area as well as a 3% decrease in the expected TAC when boundary functions are used to model the uncertainty space(s) compared to results obtained for the hyperrectangle approach. These findings support the made assumption that a more accurate model of the uncertainty space allows to guarantee steady-state flexible operation with less overdesign and reduced cost.

Furthermore, we analyzed and compared the expected TAC when solving problem 1 for case 3 (critical points for hyperpolygon approach and representative operating points) as well as case 1 (only representative operating points). With this comparison we aimed to quantify the additional cost for flexibility, that is, the additional cost to guarantee steady-state flexible operation. Figure 15 presents the results of this analysis. Note that Figure 15 shows the differences in the expected TAC for both cases compared to a reference value which is defined as the expected TAC obtained for case 1 with one representative
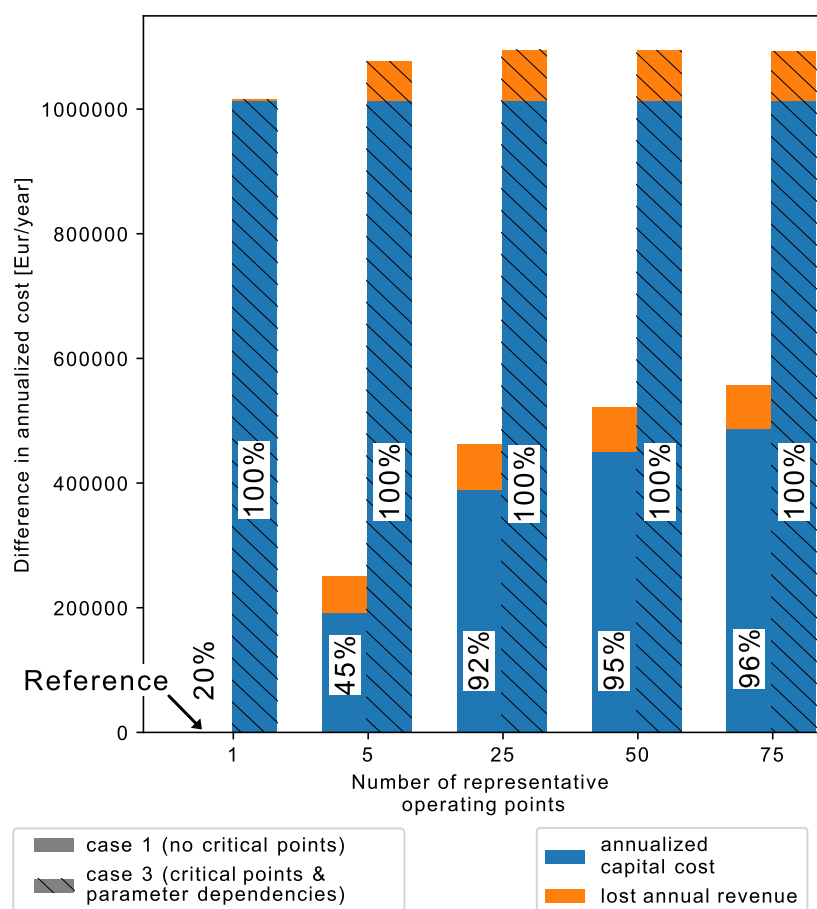
Figure 15. Differences in expected TAC compared to the reference value. Results obtained by solving the discretized design under uncertainty problem (problem 1) considering different numbers of representative operating points and either no critical operating points (case 1) or critical operating points for the hyperpolygon approach (case 3). Reference value obtained by solving problem 1 as a single-period problem for the mean operating point (with no critical operating points). The percentage given in the respective column illustrates the share of the 3415 operating points at which the respective design is able to operate.

operating point. Further note that this reference value corresponds to a design for which feasible operation is possible for (only) 20% of the 3415 operating points (result obtained during step 9 of the framework; see Section 4.9).

Figure 15 shows a substantial increase in the expected TAC for the different optimization runs of case 3 compared to that of the reference. For example, for case 3, the expected TAC considering (only) one representative operating point (mean value) is 23% higher than the reference value. Additionally, we see that the majority of the expected additional TAC for case 3 compared to the reference is caused by increased capital costs since more investment in heat-transfer area is necessary. Note, however, that for case 3 almost no changes in the annualized capital cost were observed when increasing the number of representative operating points. On the other hand, Figure 15 shows an increased steam demand (compared to the reference case) when 5 or more representative operating points were considered (which is shown as an increased loss in revenues from steam turbine electricity generation). Consequently, the steam demand was underestimated when only the mean operating point was considered.

Figure 15 also shows the expected TAC for the different optimization runs of case 1 compared to the reference, that is, the optimization runs for the four additional sets of representative operating points (5, 25, 50, and 75). We see that when more representative operating points are considered,

the expected TAC for case 1 increases. In comparison to the expected TAC for case 3, this increase in the expected TAC for case 1 compared to that for the reference is significantly more pronounced. However, even with a large number of representative operating points, the TAC for case 1 remains significantly lower compared to the results achieved for case 3.

The percentage given in the respective column shown in Figure 15 illustrates the share of the 3415 operating points at which the respective design is able to operate (results obtained during step 9, see Section 4.9). For case 1, this percentage increased when more representative operating points were included. However, even for the largest investigated set of 75 representative operating points, a steady-state flexible operation was not achieved.

A closer look at the results obtained for the set of 75 representative operating points reveals that the additional expected TAC obtained when solving problem 1 considering only representative operating points (case 1) is more than 50% lower compared to the additional expected TAC for case 3 (additional expected TAC compared to the reference case, see Figure 15). In absolute numbers, the expected TAC values for case 3 are 10% higher than the expected TAC for case 1 (75 representative operating points). We can conclude that the significantly higher cost for case 3 is needed for ensuring steady-state flexible operation, that is, feasible operation at the

(remaining) 4% of the operating points at which the design obtained for case 1 cannot reach the operational targets.

It is worth mentioning that the expected TAC would be even higher when parameter dependencies are not considered during the identification of critical operating points (comparison of cases 2 and 3, see Table 3). However, this increase in the estimated expected TAC would not correspond to actual improvements in process flexibility but rather be a result of a poorer representation of the uncertainty space.

Finally, as mentioned in Section 1, the traditional approach for dealing with uncertainty in (chemical) process design is to approximate the design specifications for a single operating point, such as the mean operating point, and apply empirical overdesign factors to these values. For the purposes of illustration, we investigated this option. The design obtained by solving problem 1 considering (only) the mean operating point was used as a basis. The design specifications obtained suggested installation of HEX A and HEX D as well as steam heaters for the feedwater and the combustion air (compare the process flow sheet of the suggested retrofit in Figure 11). Based on these design specifications, different overdesign factors were investigated. The resulting share of operating points at which feasible operation can be obtained is presented in Table 4. The

Table 4. Impact of Different Overdesign Factors on the Feasibility Share (Share of the Total Operating Points at which Feasible Operation is Possible) as well as the TAC[a]

overdesign factor	20%	30%	40%	50%
feasibility share [%]	59	67	76	83
Δ TAC (to case 1) [%]	15	23	30	38
Δ TAC (to case 3) [%]	−4	3	9	15

[a]The design obtained by solving problem 1 considering (only) the mean operating point was used as a basis. Note that only the initially suggested heat exchanger units were considered. Additional investment in other (initially not planned) heat exchanger units was not considered.

results clearly show that even when the heat-transfer areas of the suggested HEX units are increased by 50%, steady-state operation at all 3415 operating points is not feasible.

Table 4 also reports the difference in the expected TAC compared to the expected TAC for case 1 as well as case 3 (for one representative operating point). We see that considering overdesign factors leads to a significant increase in the expected TAC which can be higher than the expected TAC for case 3 (overdesign factor ≥30%) while steady-state flexible operation still cannot be achieved. This illustrates the potential ineffectiveness of overdesign factors for design under uncertainty problems. However, there is reason to believe that additional HEX units would have been considered for ensuring that all process streams reach their target temperatures. Nevertheless, such experience-based modifications of a given process design would not provide any guidance as to how much heat-transfer area would be appropriate to install in the complementary units.

5. DISCUSSION

5.1. General Discussion Regarding the Suggested Framework. The proposed framework does not include a design synthesis step but relies on the application of existing methodologies for generating high-quality structural design proposals. The framework can therefore be used for both retrofit

(redesign) as well as greenfield design problems. The structural design proposals may be collected in a superstructure. Note that structural design proposals in retrofit projects may include resequencing of existing equipment, while the costs associated with such measures can be hard to quantify. It should also be noted that when following the different steps of the framework, the optimal solution with respect to a defined objective can be found (only) among the design proposals considered in the superstructure. However, especially for large-scale industrial problems, too many structural design options in the superstructure can be inefficient, since the computational burden increases.

The proposed framework allows uncertainties to be considered when identifying the optimal design specifications for structural design proposals. These uncertainties may result from uncertain parameters, which are defined outside the system boundary and can thereby not be affected by the respective system, while recourse action is possible to counteract changes in the uncertain parameters. The proposed framework does not provide a strategy for dealing with uncertain parameters that are not measurable since recourse action is not possible in such cases.

The advanced modeling of the uncertainty space suggested in this work relies upon availability of good-quality data, especially data related to parameter dependencies. The availability of such data is commonly limited, especially for the greenfield design of chemical plants. However, it is usually possible to assume that certain independent operating periods occur, such as differing operating conditions during summer and winter times (typical example of seasonality). Nevertheless, in certain projects, a rough estimation of the uncertainty space such as upper and lower bound values for the uncertain parameters (hyperrectangle uncertainty space) for the entire expected operating period may be the best approximation possible in the early design planning stage. Note that if a single hyperrectangle representation is the only available approximation of the uncertainty space, steps 2, 3, and 6 of the proposed framework (compare Figure 6) may be omitted accepting a potentially lower accuracy of the obtained results.

In this work, we propose a new approach for identifying critical operating points. The reason is that in previous work,[26] we identified issues when utilizing the algorithms suggested by Pintarič and Kravanja.[15,16] More specifically, in our previous work, we investigated different structural design proposals for a heat exchanger network retrofit, and for some proposals the full set of critical operating points could not be identified using the suggested algorithms. Based on these results, we identified that the complexity of a structural design proposal can be an essential barrier for the successful application of the previously suggested algorithms. Since we assumed the design proposals in our previous work to be less complex compared to the case study addressed in this work (e.g., due to the presence of stream splitting), we suggested a new approach to avoid the aforementioned issues. Note that we did not perform a rigorous testing of our suggested approach for identification of critical operating points but simply concluded that it allowed the critical operating points for the investigated industrial case study to be identified. In future work, the performance of the different approaches for the identification of critical points should be further investigated for case studies of different complexity.

To allow the identification of critical operating points also when the expected uncertainty space is modeled using the hyperpolygon approach, that is, considering parameter depend-

encies, a novel updating scheme was developed (compare Section 3.6). We could confirm that the updating scheme identifies the critical operating points for a hyperpolygon representation of the uncertainty space if three conditions (compare with Section 3.6) are satisfied. However, these conditions limit the applicability of the developed updating scheme since the successful termination of the updating scheme cannot be guaranteed if multivariate dependencies as well as multiple upper and/or lower boundary functions are present. In future work, these limitations should be addressed.

In our framework, we suggest to identify representative operating points to reduce the number of scenarios considered in the discretized design under uncertainty problem and thereby the computational complexity. However, it should also be noted that available operating data (especially with high resolution) usually represent historic measurements. In this context, there is uncertainty regarding the extent to which such historic data also represent future operating conditions. Consequently, aggregating available operating data to (a lower number of) representative operating points can be a better approximation of the future operating conditions compared to high-resolution data, which may include operating conditions which will not occur again.

In Section 3.9, we suggested solving the optimal operation problem (see problem 9) for a discrete number of operating points to investigate whether the obtained design specifications allow for steady-state flexible operation. However, such a feasibility check cannot guarantee that the derived values for the design parameters are feasible for all possible operating points within the uncertainty space(s). Such a proof can only be obtained by solving the flexibility index problem, including design constraints, to global optimality. However, according to our experience, this problem (i.e., flexibility index problem) can be significantly more computationally burdensome compared to the flexibility index problem in the two-stage algorithm for identifying critical operating points. The reason for this is the influence of the representative operating points which can lead to a certain overdesign of specific equipment compared to equipment which is designed to exclusively operate at a list of (potentially) critical operating points. In this context, other work[18] in the field which is based on the identification of critical operating points suggests to conduct the final feasibility check based on Monte Carlo process simulation for discrete operating points. Note that the validity of the suggested feasibility check based on the evaluation of discrete operating points increases with the number of discrete operating points considered. Consequently, more discrete operating points may be generated by means of sampling methods.

5.2. Discussion Regarding the Industrial Case Study. In Section 4.2, we identified two independent operating periods for the industrial case study. To identify the influence of these independent operating periods on the numerical results (i.e., the expected TAC and the suggested system design), steps 3 to 9 could be performed for the overall uncertainty space and the results could be compared to the results presented in Section 4.10. If the expected TAC is higher when considering only one overall uncertainty space, it can be concluded that the additional cost relates to unnecessary overdesign resulting from heat-transfer areas being designed for combinations of uncertain parameter values, which occur individually but never in combination. However, if no (significant) difference can be identified, it can be concluded that the separation into two

different uncertainty spaces and the connected additional work was unnecessary.

Table 3 shows that the expected TAC is lower when problem 1 is solved considering parameter dependencies (case 3) compared to ignoring parameter dependencies (case 2). However, as previously mentioned, the difference in the expected TAC is relatively small in this specific case. The difference may be significantly more pronounced for other cases studies.

Finally, Figure 15 illustrates that the expected TAC increases significantly if feasible operation is required for all operating conditions within the expected uncertainty space. For the set of 75 representative operating points, Figure 15 shows that more than 50% of this additional expense is incurred to enable operation for the final 4% of the operating points, which can be assumed to be extreme values within this uncertainty space. This observation underlines the necessity of careful assumptions and data preprocessing such as outlier detection before utilizing the suggested framework.

6. CONCLUSIONS AND OUTLOOK

In this work, we presented a framework that aims to support designers in (early) design stage screening processes when dealing with chemical process design under uncertainty. The framework is a stepwise approach which identifies:

1. if structural design proposals are structurally feasible,

2. the most cost-efficient overdesign of equipment size required to guarantee steady-state flexible operation,

3. a basis to compare different structural design proposals with respect to a given objective.

As the first step of the proposed framework, structural flexibility analysis is performed for (different) given structural design proposals to exclude those proposals which are structurally infeasible, that is, whose structural layout does not allow for steady-state flexible operation. The design specifications of the remaining (structurally feasible) proposals can thereafter be optimized by solving the discretized design under uncertainty problem. This implies that the expected objective function value is calculated for a set of representative operating periods, while steady-state flexible operation is ensured by including critical operating points in the constraints.

In contrast to previous work, our proposed framework includes advanced modeling of the expected uncertainty space to allow dependencies in the uncertain parameters to be considered as well as variation between independent operating periods. In this context, we also discussed different origins and types of independent operating periods as well as strategies for identifying such independent operating periods. Several theoretical examples were introduced to illustrate how our approach offers more degrees of freedom for modeling the expected uncertainty space compared to the traditional approach based on expected upper and lower bound values (hyperrectangle representation). If dependencies in the uncertain parameters and/or independent operating periods are expected, these degrees of freedom can be exploited to exclude subsets of the hyperrectangle uncertainty space that contain combinations of the uncertain parameter values that are not expected to occur. Consequently, a more accurate representation of the actual uncertainty space is achieved, and flexibility assessments based on this uncertainty model thereby better represent the actual flexibility of the process.

Furthermore, by means of different theoretical examples, we showed that the modeling of the expected uncertainty space can affect the identification of critical operating points. If the critical operating points are identified assuming a hyperrectangle representation of the uncertainty space, the identified critical points may represent combinations of uncertain parameter values that are not expected to occur simultaneously. If the design is then optimized to be able to handle these critical, but nonexpected, conditions, this is likely to result in unnecessary overdesign of the equipment. We therefore proposed a novel algorithm to consider parameter dependencies and independent operating periods when identifying critical operating points.

The applicability and benefits of the proposed framework were illustrated through an industrial case study of a heat exchanger network retrofit. A structural design proposal that was defined prior to the application of the suggested framework was evaluated. We identified correlations between different uncertain parameters as well as two independent operating periods. Furthermore, critical operating points were identified utilizing (i) a hyperrectangle model of the expected uncertainty space and (ii) our advanced modeling approach. In addition to the critical operating points, different sets of representative operating points were investigated.

The results confirm that considering (only) representative operating points in the problem 1 is problematic when aiming for steady-state flexible operation. The results indicate that steady-state flexible operation can only be guaranteed when considering the identified critical operating points in the constraints of the design problem. Note that steady-state flexible operation was possible when considering critical operating point identified (i) for the hyperrectangle model of the expected uncertainty space as well as (ii) using our advanced modeling approach. When considering only representative operating points, feasible steady-state operation at all expected operating points was not possible for the investigated sets of representative operating points. Note that the absolute number of operating points at which feasible steady-state operation is not possible is bigger than the number of representative operating points, also for the largest set investigated (75 points)[o].

Furthermore, the results show that more total heat-transfer area is required as well as a higher TAC when the design problem is solved considering the critical operating points identified for a hyperrectangle uncertainty space, compared to the critical operating points based on our modeling approach. It is worth mentioning that for the investigated case study, our modeling approach of the uncertainty space resulted in a 5% lower requirement of total heat-transfer area while also the expected TAC decreased by 3%. This observation supports our assumption that an overly simplified model of the uncertainty space may lead to overdesign and unnecessary costs.

The numerical results also indicate that including critical operating points in the design problem significantly reduces the number of representative operating points that need to be considered to obtain a fair approximation of the expected objective function value. More precisely, it was shown that for the investigated case study, a set of 25 representative operating points is sufficient to achieve a good approximation of the expected TAC when considering critical operating points in the constraints. When critical operating points are not included in the design problem, the results indicate that the expected TAC obtained for the largest set of representative operating points considered in this work (75 points) may not be representative, that is, the expected TAC may change remarkably when

considering sets with more than 75 representative operating points. Since higher numbers of representative operating points lead to increasing problem complexity, the additional effort for identifying critical operating points can be justified.

For the investigated case study, a lower value for the expected TAC was obtained when considering only representative operating points but the obtained design specifications could not handle all expected occurrences of the uncertain parameters. For example, for the set with 75 representative operating points, the expected TAC to guarantee steady-state flexible operation was identified to be 10% higher compared to the expected TAC when considering only representative operating points. Finally, we showed that using overdesign factors is rather ineffective for handling uncertainty in the investigated case study since the additional expenses incurred may be significant while steady-state flexible operation still cannot be guaranteed.

As mentioned in Section 5, rigorous testing and further development of the novel strategy to identify the critical operating points of both hyperrectangle and hyperpolygon representations of the uncertainty set remain for future work. Especially, the limitations of the proposed algorithm to identify the critical operating points of hyperpolygon representations should be addressed. Additionally, the proposed framework could be extended by incorporating operability aspects. For example, the calculation and evaluation of an operability metric, such as the operability index initially proposed by Vinson and Georgakis,[45,46] after step 4 of the proposed framework could reveal interesting insights with respect to the operability of certain process design proposals.

■ ASSOCIATED CONTENT

ⓢ **Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.iecr.3c01584.

> Further information on the developed framework as well as mathematical modeling of the investigated retrofit proposal and miscellaneous numeric results obtained during the application of the suggested framework on the presented case study (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Christian Langner − *Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg 41296, Sweden;* ⓞ orcid.org/0000-0003-0953-8504; Email: christian.langner@chalmers.se

Authors

Elin Svensson − *Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg 41296, Sweden; CIT Renergy AB, Gothenburg 41258, Sweden;* ⓞ orcid.org/0000-0002-3122-6139

Stavros Papadokonstantakis − *Institute of Chemical, Environmental and Bioscience Engineering, Vienna University of Technology (TU Wien), Vienna 1060, Austria;* ⓞ orcid.org/0000-0002-7774-0192

Simon Harvey − *Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg 41296, Sweden*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.iecr.3c01584

Notes

The authors declare no competing financial interest.

■ ADDITIONAL NOTES

[a]The term expected uncertainty space refers to the space in which the uncertain parameters are expected to vary, that is, the expected uncertainty space is defined by the expected disturbance range of each uncertain parameter.

[b]A structural design proposal describes the structural layout of the process including the placement of equipment while the design specifications, such as the size of the equipment, remain unspecified.

[c]An example can be found in our recent work on flexibility analysis considering parameter dependencies[31] where we analyzed the flexibility of a design proposal for an industrial heat exchanger network (HEN). When ignoring parameter dependencies, the resulting flexibility metric indicates that the design proposal would not allow steady-state flexible operation while the opposite is found to be true when considering dependencies between uncertain parameters during the analysis.

[d]Note that this division is an illustration and neither a recommendation nor a suggestion to (always) divide operating data according to calendar seasons.

[e]This observation is in line with the definition of critical points by Halemane and Grossmann[14] who defined critical points as those realizations of the uncertain parameters for which the feasible region of operation is the smallest.

[f]Note that this idea is based on the work by Halemane and Grossmann[14] who observed that if feasible operation at the critical operating points can be guaranteed, the same applies for any other point within the expected uncertainty space.

[g]As previously mentioned, this step can be performed relying on historic operating data or by anticipating the uncertainty space.

[h]Note that considering a parameter dependency requires (more) advanced modeling which means that (simply) considering all identified parameter dependencies may increase the problem complexity of the subsequent flexibility index problem and the identification of critical operating points.

[i]Note that this procedure is universal for projecting a solution of the flexibility index problem (θ^*, δ^*) from the solution space (feasible uncertainty space) to the expected uncertainty space. Further note that the same procedure was utilized to identify the critical operating point for the example shown in Figure [5].

[j]Note that the critical operating points which were identified for different independent operating periods must be combined to one set to be included in problem [1].

[k]Note that a feasibility check based on solving problem [9] for a discrete number of operating points cannot guarantee that the derived values for the design parameters are feasible for all possible operating points within the uncertainty space(s) of the independent operating period(s). This is further discussed in Section [5].

[l]Note that in current operation (before retrofit), sufficient steam is available to meet the steam demand at all times, and the steam demand is expected to be reduced in the future as a result of the suggested retrofit.

[m]Note that estimations for the electricity prices for the year 2030 were chosen since this year can be seen as an average year for the expected lifetime of the investment.

[n]The set of 25 representative operating points was chosen since it allows for a fair approximation of the expected TAC (conclusion drawn based on the analysis shown in Figure [14]).

[o]Therefore, the strategy outlined in Section [3.9] would be ineffective in such a scenario since it proposes to include all operating points at which feasible operation was not possible as critical operating points in problem [1].

■ REFERENCES

(1) Grossmann, I.; Halemane, K.; Swaney, R. Optimization strategies for flexible chemical processes. *Comput. Chem. Eng.* **1983**, *7*, 439−462.

(2) Grossmann, I. E.; Calfa, B. A.; Garcia-Herreros, P. Evolution of concepts and models for quantifying resiliency and flexibility of chemical processes. *Comput. Chem. Eng.* **2014**, *70*, 22−34.

(3) Ochoa, M. P.; Grossmann, I. E. Novel MINLP formulations for flexibility analysis for measured and unmeasured uncertain parameters. *Comput. Chem. Eng.* **2020**, *135*, 106727.

(4) Grossmann, I. E.; Sargent, R. W. Optimum design of chemical plants with uncertain parameters. *AIChE J.* **1978**, *24*, 1021−1028.

(5) Steimel, J.; Engell, S. Conceptual design and optimization of chemical processes under uncertainty by two-stage programming. *Comput. Chem. Eng.* **2015**, *81*, 200−217.

(6) Pistikopoulos, E. N.; Ierapetritou, M. G. Novel approach for optimal process design under uncertainty. *Comput. Chem. Eng.* **1995**, *19*, 1089−1110.

(7) Acevedo, J. N.; Pistikopoulos, E. N. A Parametric MINLP Algorithm for Process Synthesis Problems under Uncertainty. *Ind. Eng. Chem. Res.* **1996**, *35*, 147−158.

(8) Rooney, W. C.; Biegler, L. T. Optimal process design with model parameter uncertainty and process variability. *AIChE J.* **2003**, *49*, 438−449.

(9) Sahinidis, N. V. Optimization under uncertainty: state-of-the-art and opportunities. *Comput. Chem. Eng.* **2004**, *28*, 971−983.

(10) Birge, J. R.; Louveaux, F. *Introduction to Stochastic Programming*; Springer: New York, 2011; pp I−XXV, 1−485

(11) Wei, J.; Realff, M. J. Sample average approximation methods for stochastic MINLPs. *Comput. Chem. Eng.* **2004**, *28*, 333−346.

(12) Pintarič, Z. N.; Kasaš, M.; Kravanja, Z. Sensitivity Analyses for Scenario Reduction in Flexible Flow Sheet Design with a Large Number of Uncertain Parameters. *AIChE J.* **2013**, *59*, 2862−2871.

(13) Mozafari, Y.; Rosehart, W. Representative operating scenario selection with algebraic multi-grid clustering for integrated energy systems planning. *Int. J. Electr. Power Energy Syst.* **2023**, *146*, 108767.

(14) Halemane, K. P.; Grossmann, I. E. Optimal process design under uncertainty. *AIChE J.* **1983**, *29*, 425−433.

(15) Pintarič, Z. N.; Kravanja, Z. A-priori identification of critical points for the design and synthesis of flexible process schemes. *Comput.-Aided Chem. Eng.* **2006**, *21*, 503−508.

(16) Pintarič, Z. N.; Kravanja, Z. Identification of critical points for the design and synthesis of flexible processes. *Comput. Chem. Eng.* **2008**, *32*, 1603−1624.

(17) Pintarič, Z. N.; Kravanja, Z. A multilevel MINLP methodology for the approximate stochastic synthesis of flexible chemical processes. *Rev. Chem.* **2007**, *58*, 387−391.

(18) Novak Pintarič, Z.; Kravanja, Z. A methodology for the synthesis of heat exchanger networks having large numbers of uncertain parameters. *Energy* **2015**, *92*, 373−382.

(19) Swaney, R. E.; Grossmann, I. E. An index for operational flexibility in chemical process design. Part I: Formulation and Theory. *AIChE J.* **1985**, *31*, 621−630.

(20) Di Pretoro, A.; Negny, S.; Montastruc, L. The traveling deliveryman problem under uncertainty: Fundamentals for flexible supply chains. *Comput. Chem. Eng.* **2022**, *160*, 107730.

(21) da Silva, P. R.; Aragão, M. E.; Trierweiler, J. O.; Trierweiler, L. F. Integration of hydrogen network design to the production planning in refineries based on multi-scenarios optimization and flexibility analysis. *Chem. Eng. Res. Des.* **2022**, *187*, 434−450.

(22) Luo, J.; Moncada, J.; Ramirez, A. Development of a Conceptual Framework for Evaluating the Flexibility of Future Chemical Processes. *Ind. Eng. Chem. Res.* **2022**, *61*, 3219−3232.

(23) Zhang, Q.; Grossmann, I. E.; Lima, R. M. On the relation between flexibility analysis and robust optimization for linear systems. *AIChE J.* **2016**, *62*, 3109−3123.

(24) Floudas, C. A.; Grossmann, I. E. Synthesis of flexible heat exchanger networks with uncertain flowrates and temperatures. *Comput. Chem. Eng.* **1987**, *11*, 319−336.

(25) Papalexandri, K. P.; Pistikopoulos, E. N. Synthesis and Retrofit Design of Operable Heat Exchanger Networks. 1. Flexibility and Structural Controllability Aspects. *Ind. Eng. Chem. Res.* **1994**, *33*, 1718−1737.

(26) Langner, C.; Svensson, E.; Harvey, S. A framework for flexible and cost-efficient retrofit measures of heat exchanger networks. *Energies* **2020**, *13*, 1472−1524.

(27) Lima, F. V.; Jia, Z.; Ierapetritou, M.; Georgakis, C. Similarities and differences between the concepts of operability and flexibility: The steady-state case. *AIChE J.* **2010**, *56*, 702−716.

(28) Grossmann, I. E.; Floudas, C. A. Active constraint strategy for flexibility analysis in chemical processes. *Comput. Chem. Eng.* **1987**, *11*, 675−693.

(29) Pulsipher, J. L.; Zavala, V. M. A mixed-integer conic programming formulation for computing the flexibility index under multivariate gaussian uncertainty. *Comput. Chem. Eng.* **2018**, *119*, 302−308.

(30) Langner, C.; Svensson, E.; Harvey, S. Flexibility analysis of chemical processes considering dependencies between uncertain parameters. *Comput.-Aided Chem. Eng.* **2021**, *50*, 1105−1110.

(31) Langner, C.; Svensson, E.; Papadokonstantakis, S.; Harvey, S. Flexibility analysis using boundary functions for considering dependencies in uncertain parameters. *Comput. Chem. Eng.* **2023**, *174*, 108231.

(32) Rooney, W. C.; Biegler, L. T. Incorporating joint confidence regions into design under uncertainty. *Comput. Chem. Eng.* **1999**, *23*, 1563−1575.

(33) Marton, S.; Langner, C.; Svensson, E.; Harvey, S. Costs vs. flexibility of process heat recovery solutions considering short-term process variability and uncertain long-term development. *Front. Chem. Eng.* **2021**, *3*, 679454.

(34) Langner, C.; Svensson, E.; Papadokonstantakis, S.; Harvey, S. Flexibility analysis of chemical processes considering overlaying uncertainty sources. *Comput.-Aided Chem. Eng.* **2022**, *49*, 769−774.

(35) Floudas, C. A.; Gümüş, Z. H.; Ierapetritou, M. G. Global Optimization in Design under Uncertainty: Feasibility Test and Flexibility Index Problems. *Ind. Eng. Chem. Res.* **2001**, *40*, 4267−4282.

(36) Hyndman, R. J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Melbourne, Australia, 2018. https://otexts.com/fpp2/ (accessed Nov 11, 2022).

(37) Boslaugh, S.; Watters, P. A. *Statistics in a Nutshell-a Desktop Quick Reference*; O'Reilly, 2008; pp I−XXI, 1−452

(38) Barber, C. B.; Dobkin, D. P.; Huhdanpaa, H. The Quickhull Algorithm for Convex Hulls. *ACM Trans. Math Software* **1996**, *22*, 469−483.

(39) Lloyd, S. P. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theor.* **1982**, *28*, 129−137.

(40) Jin, X.; Han, J. *Encyclopedia of Machine Learning*; Sammut, C., Webb, G. I., Eds.; Springer US: Boston, MA, 2010; pp 563−564.

(41) Sahinidis, N. V. *BARON 21.1.13: Global Optimization of Mixed-Integer Nonlinear Programs, User's Manual*, 2017.

(42) Hart, W. E.; Watson, J.-P.; Woodruff, D. L. Pyomo: modeling and solving mathematical programs in Python. *Math. Program. Comput.* **2011**, *3*, 219−260.

(43) Mann, S.; Kuhlmann, R. New and improved GAMS links for Pyomo and JuMP, GAMS Blog, 2020. https://www.gams.com/blog/2020/06/new-and-improved-gams-links-for-pyomo-and-jump/ (accessed Aug 25, 2023).

(44) Göransson, L.; Lehtveer, M.; Nyholm, E.; Taljegard, M.; Walter, V. The Benefit of Collaboration in the North European Electricity System Transition-System and Sector Perspectives. *Energies* **2019**, *12*, 4648.

(45) Vinson, D. R.; Georgakis, C. A New Measure of Process Output Controllability. *IFAC Proc. Vol.* **1998**, *31*, 663−672.

(46) Vinson, D. R.; Georgakis, C. A new measure of process output controllability. *J. Process Control* **2000**, *10*, 185−194.