Flexibility analysis using boundary functions for considering dependencies in uncertain parameters

(article starts on next page)

Flexibility analysis using boundary functions for considering dependencies in uncertain parameters

Christian Langner [a],[*], Elin Svensson [a],[b], Stavros Papadokonstantakis [c], Simon Harvey [a]

[a] *Department of Space, Earth and Environment, Chalmers University of Technology, Hörsalsvägen 7A, 412 96, Gothenburg, Sweden*
[b] *CIT Renergy AB, Sven Hultins Plats 1, 412 58, Gothenburg, Sweden*
[c] *Institute of Chemical, Environmental and Bioscience Engineering, Vienna University of Technology (TU Wien), Getreidemarkt 9/166, 1060 Vienna, Austria*

ARTICLE INFO

ABSTRACT

In this work, we present a novel approach for considering dependencies (often called correlations) in the uncertain parameters when performing (deterministic) flexibility analysis. Our proposed approach utilizes (linear) boundary functions to approximate the observed or expected distribution of operating points (i.e. uncertainty space), and can easily be integrated in the flexibility index or flexibility test problem. In contrast to the hyperbox uncertainty sets commonly used in deterministic flexibility analysis, uncertainty sets based on boundary functions allow subsets of the hyperbox which limit the flexibility metric but in which no operation is observed or expected, to be excluded. We derive a generic mixed-integer formulation for the flexibility index based on uncertainty sets defined by boundary functions, and suggest an algorithm to identify boundary functions which approximate the uncertainty set with high accuracy. The approach is tested and compared in several examples including an industrial case study.

1. Introduction

Flexibility analysis denotes different concepts to evaluate the capability of a physical system to react towards uncertainty (e.g. disturbances) in order to maintain feasible operation (Grossmann et al., 2014). In this context, feasible operation is achieved if the physical system with given equipment (sizes) reaches pre-defined target values which can be formulated as mathematical constraints. Flexibility analysis originates from the aspiration to avoid unnecessary overdesign of the equipment while guaranteeing steady state flexible operation, i.e. steady state operation at numerous operating points within a limited uncertain parameters space or uncertainty set T (Grossmann et al., 2014). Unnecessary overdesign may result from conservative safety margins during the design process (e.g. rule-of-thumb estimates), which are the consequence of insufficient knowledge about how a physical system reacts towards uncertainty, e.g., when only nominal conditions were considered during the design phase (Ochoa and Grossmann, 2020). To avoid unnecessary overdesign of equipment but still allow for steady state flexible operation, it is therefore vital to consider those operating conditions which set the hardest requirements on the equipment size also known as critical operating points. Halemane and Grossmann (1983) defined critical operating points as those realizations of the uncertain parameters, θ, for which a given design, d, (i.e. structural layout, equipment size) has either the smallest degree of feasibility

(i.e. at the critical operating point feasible operation can be observed while the smallest deviation from the critical parameter values can lead to infeasibility) or the largest degree of infeasibility (i.e. at the critical operating point occurs the maximum constraint violation). In order to identify the degree of feasibility/infeasibility, they formulated the feasibility constraint (1) also known as max–min–max constraint. The authors further proposed a solution algorithm for flexible process design based on the procedure suggested by Grossmann and Sargent (1978) in which the set of potentially critical points is obtained by maximizing the inequalities with respect to the uncertain parameters assuming monotonicity (Halemane and Grossmann, 1983).

$$\max_{\theta \in T} \min_{z} \max_{j \in J} f_j(d, z, \theta) \leq 0 \qquad (1)$$

The feasibility constraint (1) involves a non-trivial max–min–max optimization problem, whose objective is to ensure feasible operation ($f_j(d, z, \theta) \leq 0$, $j \in J$) over the entire uncertainty set, T, while accounting for possible adjustment of the control variables, z, in order to adapt to the uncertainty in the parameters θ. Based on the formulation of the feasibility constraint, Swaney and Grossmann (1985) formulated two concepts: the flexibility test and the flexibility index to perform flexibility analysis for a hyperbox uncertainty set, T_{box}. (compare (2)).

$$T_{box} = \left\{ \theta_i | \theta_{i,L} \leq \theta_i \leq \theta_{i,U} \right\} \forall \ \theta_i \in \theta \qquad (2)$$

For a given design, d, and any given uncertainty set, T, the solution of the flexibility test defines a critical point, θ^*, for feasible operation in line with the definition of Halemane and Grossmann (1983) (see above). While the primary purpose of the flexibility test is to determine if the operation of a given design, d, remains feasible for all realizations of θ in T, the flexibility index is a metric (single scalar δ) which indicates the maximum feasible deviation of each uncertain parameter in θ given an expected variation range and nominal/mean value(s) (Swaney and Grossmann, 1985). Furthermore, the flexibility index implicitly indicates potentially critical points which can been utilized to develop flexible process designs (see e.g., Langner et al., 2020). The problem formulation of the flexibility index for a hyperbox uncertainty set, T_{box}, is given in (A.1) in Appendix A. Note that in the problem formulation of the flexibility index, the hyperbox uncertainty set is scaled by the scalar δ ($T_{box}(\delta)$) which thus refers to the largest scaled hyperbox which can be inscribed in the feasible region.

Based on the definition of the flexibility test and index (compare Swaney and Grossmann, 1985) as key metrics to perform flexibility analysis, much effort has been dedicated to study a variety of aspects related to the field during the last decades. For an extensive review and a historical perspective on this topic, the interested reader is referred to Grossmann et al. (2014) and Zhang et al. (2016). An overview of some selected literature will be provided here. To determine the flexibility index, Swaney and Grossmann (1985) proposed search procedures for the special case of exclusively convex constraint functions, $f_j(d, z, \theta)$, in which the solution of (A.1) corresponds to vertices of $T_{box}(\delta)$. To overcome the limitations of vertex exploration (exponential number of vertices, 2^θ, and convexity of constraint functions $f_j(d, z, \theta)$), Grossmann and Floudas (1987) reformulated the flexibility index problem (compare (A.1)) following the idea that the solution of (A.1) must be on the boundary of the feasible region. Detailed information on the study by Grossmann and Floudas (1987) is provided in Appendix A. Their reformulation yielded a bi-level optimization problem (compare (A.2)) which they could further transform to a (single-level) Mixed-Integer (Non-)Linear Program (MI(N)LP) for the flexibility index. Together with the MI(N)LP formulation, Grossmann and Floudas (1987) proposed the active constraint strategy which is based on the idea that the flexibility analysis for a given design, d, can be performed in the space of constraints that can potentially limit the flexibility, i.e. the constraints that are active at the solution of the flexibility analysis problem. Note that the active constraint strategy does not rely on the assumption that critical points correspond to vertices of $T_{box}(\delta)$.

Raspanti et al. (2000) used constraint aggregation functions (Kreisselmeier–Steinhaus a.k.a. KS functions) and smoothing functions to reformulate the MI(N)LP formulation for the flexibility index by Grossmann and Floudas (1987) to a single non-linear programming problem (NLP). The resulting NLP is non-convex which may lead to increased difficulties solving the problem for global optimality. Li et al. (2015) suggested a framework to calculate the flexibility index by means of an alternating direction matrix embedded in a Simulated Annealing algorithm. They suggest a search strategy to explore the feasible region by means of randomly created search directions and control the search via the simulated annealing of a temperature. While the recent development of personal computer capacity allows for fast analysis of search directions in the order of $\mathcal{O}(1000)$, the resulting value remains an upper bound approximation of the flexibility index. In addition to the traditional definition of uncertainty where the control variables, z, can be manipulated to counteract variation in the uncertain parameters, Ostrovsky et al. (2003) and Rooney and Biegler (2003) grouped the uncertain parameters, θ, into two types, measured θ_m and unmeasured θ_u. Variation of unmeasured uncertain parameters cannot be counteracted by means of recourse actions (i.e. manipulation of control variables) since these countermeasures require knowledge of the source of the disturbance. Recently, Ochoa and Grossmann (2020) reformulated the MI(N)LP formulation by Grossmann and Floudas

(1987) to analyze the flexibility by means of the active constraint strategy also when unmeasured uncertain parameters, θ_u, are present.

An alternative approach for flexibility analysis was suggested by Pistikopoulos and Mazzuchi (1990) and Straub and Grossmann (1990). Both works suggest determining a stochastic flexibility index which measures the probability that a given design (defined by linear constraint functions) remains feasible given the joint probability density function, p, of the uncertain parameters. The stochastic flexibility index can be obtained directly by integrating p over the feasible region projected in the space of the uncertain parameters (Grossmann et al., 2014). This integration can be done using Monte Carlo Sampling and requires a feasibility check of every sampled realization. Straub and Grossmann (1993) later extended the framework of the stochastic flexibility index also for systems described by non-linear constraint functions. Since Monte Carlo Sampling can be computationally expensive due to the large number of samples required to cover the uncertainty set, Pulsipher and Zavala (2018) suggested a mixed-integer conic program which can be used to compute a lower bound for the stochastic flexibility index.

In addition to the deterministic and the stochastic flexibility indexes, a third flexibility index has been proposed by Lai and Hui (2008): the volumetric flexibility index. The volumetric flexibility index quantifies the percentage of the expected hyperbox uncertainty set that can be feasibly handled. In a geometric sense, the volumetric flexibility index describes the volumetric fraction of the hypervolume of the feasible space within the expected hyperbox, compared to the volume of this hyperbox bounded by the expected upper and lower limits of uncertain parameters. The authors concluded that the approximation of the feasible space within the hyperbox is challenging, and, recently, Zheng et al. (2021) presented a novel approach to approximate the volumetric flexibility index based on the symbolic computation method.

Due to the nature of joint probability density functions, the stochastic flexibility index is able to capture dependencies in the uncertain parameters. On the other hand, in most studies focusing on deterministic flexibility analysis, the uncertain parameters are assumed to vary independently which is well expressed by the hyperbox representation of the uncertainty set (T_{box} and $T_{box}(\delta)$, respectively). However, for practical applications, this assumption is not always valid. Especially for industrial processes it is very probable that dependencies in (some of) the uncertain parameters are present. In case of dependencies in the uncertain parameters, several studies pointed out that the flexibility index based on the hyperbox uncertainty set (significantly) underestimates the flexibility of the underlying physical system (compare Rooney and Biegler, 1999, Pulsipher and Zavala, 2018 and Langner et al., 2021).

We identified two approaches in the literature which present alternatives to the hyperbox uncertainty set, namely the approaches presented by Grossmann and Floudas (1987) and Pulsipher and Zavala (2018) (the approach of Pulsipher and Zavala, 2018 is based on an idea presented by Rooney and Biegler, 1999). Both approaches were suggested explicitly for considering dependencies in the uncertain parameters when modeling the uncertainty set. A detailed overview on these existing approaches is provided in Section 2. However, as we show in Section 2, the approach of Grossmann and Floudas (1987) can lead to the obtained (deterministic) flexibility index overestimating the actual flexibility. Additionally, we identified that the approach of Pulsipher and Zavala (2018) is not suitable for performing deterministic flexibility analysis since it was developed as a method for approximation of the stochastic flexibility index. Although this approximation is done using deterministic operations, the obtained results are inherently different from the findings provided by the deterministic flexibility index.

To allow for considering dependencies in the uncertain parameters when performing deterministic flexibility analysis while lowering the risk for over- and underestimation, we developed a novel approach which was outlined in a recent conference contribution (Langner et al.,
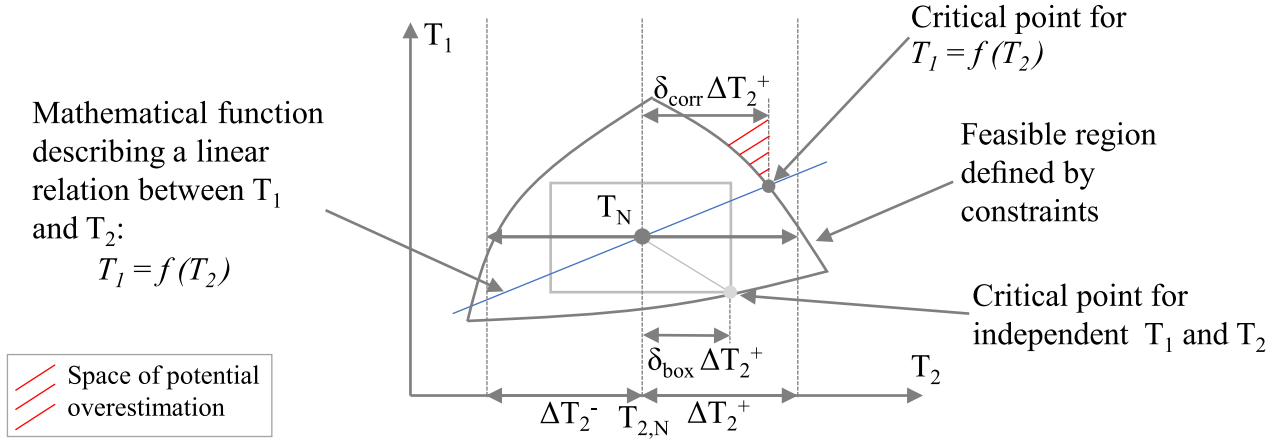
Fig. 1. Conceptual illustration of the flexibility index using a single equation (regression) model to capture dependencies in the uncertain parameters. Note, the presence of expected operating points in the red hatched area would indicate that the feasible interval indicated by the flexibility index $[\delta_{corr}\Delta T_2^-, \delta_{corr}\Delta T_2^+]$ overestimates the flexibility since these operating points would be within the feasible interval but are de facto infeasible.

2021) and applied to several small-scale theoretical examples to demonstrate its applicability. In the work described in this paper, we continue the development of our approach by deriving (and presenting for the first time) a generic problem formulation which is applicable not only to small-scale literature and theoretical examples but also to more complex and multi-dimensional examples. In addition to the generic problem formulation, we formulated supporting algorithms to extend our proposed approach. The generic problem formulation and the supporting algorithms are presented in Section 3. In Section 4.1, we use an illustrative example to demonstrate the differences between our proposed approach and the other approaches presented in Section 2. Additionally, the applicability of the proposed approach to more complex and multi-dimensional cases is illustrated by means of a literature Heat Exchanger Network (HEN) example (see Section 4.2) and an industrial case study in Section 5.

2. Analysis of (existing) approaches for flexibility analysis considering dependencies in the uncertain parameters

As mentioned in Section 1, several studies in the literature point out that the deterministic flexibility index based on the hyperbox uncertainty set potentially underestimates the flexibility of a process if dependencies in the uncertain parameters are present. In this context, the term underestimation refers to the phenomenon that the flexibility index indicates only a small feasible interval $[\delta_{box}\Delta\theta^-, \delta_{box}\Delta\theta^+]$ while a much larger share of the expected operating points or expected realizations of the uncertain parameters are within the feasible region. There is consensus in the literature that the reason for this underestimation is bad resemblance of the hyperbox uncertainty set with the real uncertainty set, i.e. the actual distribution of operating points. To overcome this problem, we identified two approaches in the literature which are suitable to consider dependencies when modeling the expected uncertainty set. Both approaches build upon knowledge regarding the distribution of operating points, i.e. historical or expected values of the uncertain parameters.

2.1. Approach based on regression models

Grossmann and Floudas (1987) suggested to express the dependent uncertain parameters through algebraic equations, $f(\theta) = 0$, which can be included as additional constraints in the flexibility index problem (compare (A.2)). Although not explicitly formulated by the authors, the approach can be generalized by reformulating the uncertainty set $(T_{box}(\delta))$ in the flexibility index problem (compare (A.2)). For this
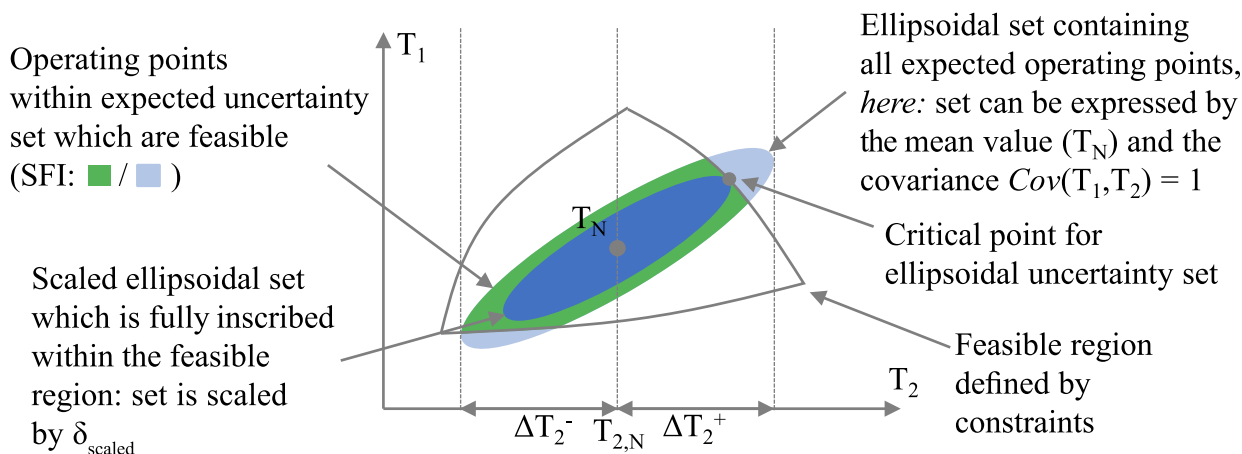
purpose, we group the uncertain parameters into independent uncertain parameters (θ_{ind}) and dependent uncertain parameters (θ_{dep}) and express the dependent uncertain parameters by a set of functions C. The uncertainty set in which the uncertain parameters vary can then be described as done in (3).

$$T_{corr,box}(\delta) = \begin{cases} \{\theta_i \mid \theta_{i,N} - \delta\Delta\theta_i^- \leq \theta_i \leq \theta_{i,N} + \delta\Delta\theta_i^+ \} \; \forall \; \theta_i \text{ in } \theta_{ind} \\ \{\theta_j \mid f_c(\theta_{ind}) \; \forall \; c \text{ in } C \} \; \forall \; \theta_j \text{ in } \theta_{dep} \end{cases} \quad (3)$$

In general, different mathematical models can be considered to express the dependencies between dependent and independent uncertain parameters. An intuitive approach, is to express dependencies via (linear) single equation models as demonstrated by Grossmann and Floudas (1987) and reported in the aforementioned review on concepts and models for flexibility analysis (Grossmann et al., 2014). A conceptual illustration of the flexibility index for a linear single equation model in comparison to the hyperbox uncertainty set is shown in Fig. 1.
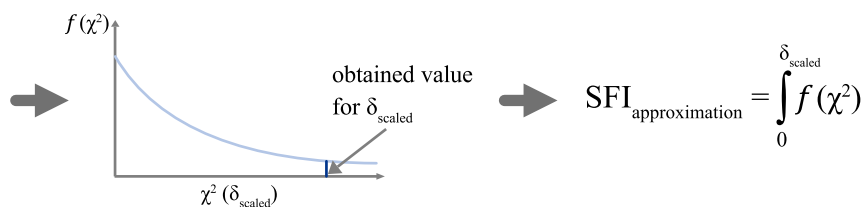
A single equation model is comparable to a regression model, i.e. a model which defines how a dependent variable changes with respect to one or several independent variable(s). In the two-dimensional case shown in Fig. 1, we see that the result of the flexibility index (scale parameter δ) is different when the uncertainty set is modeled using a single equation model to express the dependency between the two uncertain parameters T_1 and T_2 compared to the hyperbox uncertainty set. More specifically, δ_{corr} is larger than δ_{box} which implies that the underestimation of the flexibility by means of the flexibility index based on the hyperbox uncertainty set could possibly be decreased. On the other hand, it needs to be ensured that the larger value of δ_{corr} compared to δ_{box} is not the result of an overestimation of the flexibility. This is analyzed in more detail in the next paragraph. Firstly, we want to highlight that in the hyperbox case the scale parameter, δ_{box}, explicitly defines the feasible variation of all uncertain parameters, i.e. the products $\delta_{box}\Delta\theta_i^-$ and $\delta_{box}\Delta\theta_i^+$, respectively indicate the feasible range in which the uncertain parameter, θ_i, may vary. On the other hand, when expressing dependent uncertain parameters via single equation models, the scale parameter, δ_{corr}, defines the feasible variation of the independent uncertain parameters, only. The variation of the dependent parameters is expressed as functional relationship of the independent parameters, i.e. when using single equation models one assumes that for each value or realization of the independent uncertain parameters exactly one value or realization of the dependent uncertain parameters can be expected as shown in Fig. 1.

The observations described in the previous paragraph relate to the main drawback of single equation regression models, which is that they are only exact if the strongest possible agreement exists between

Operating points within expected uncertainty set which are feasible (SFI: ■ / ▨)

Scaled ellipsoidal set which is fully inscribed within the feasible region: set is scaled by δ_{scaled}

Ellipsoidal set containing all expected operating points, *here:* set can be expressed by the mean value (T_N) and the covariance $Cov(T_1, T_2) = 1$

Critical point for ellipsoidal uncertainty set

Feasible region defined by constraints

ΔT_2^- $T_{2,N}$ ΔT_2^+

Derivation of SFI approximation based on δ_{scaled} and PDF of χ^2-distribution:

PDF of χ^2-distribution, $f(\chi^2)$, defined for discrete values of n^θ, *here:* $n^\theta = 2$ (T_1, T_2)

$f(\chi^2)$

obtained value for δ_{scaled}

$\chi^2(\delta_{scaled})$

$$SFI_{approximation} = \int_0^{\delta_{scaled}} f(\chi^2)$$

SFI: stochastic flexibility index
PDF: probability density function

Fig. 2. Conceptual illustration of an ellipsoidal uncertainty set to approximate the stochastic flexibility index (SFI) according to the approach of Pulsipher and Zavala (2018). Note, the stochastic flexibility index can be interpreted as the share of the green colored shape (covering all expected operating points within the feasible region) and the light blue colored shaped (expected uncertainty space).

the correlated uncertain parameters. Commonly, correlated uncertain parameters in chemical processes agree only to some extent which means that single equation regression models are able to capture the trend between these uncertain parameters well while neglecting operating points which deviate from this trend (i.e. operating points caused by "other" sources of uncertainty). Consequently, when using single equation regression models, there is a high likelihood that realizations of the uncertain parameters or operating points which are expected are not included in the modeled uncertainty set as we have also shown previously (compare Langner et al., 2021). However, these operating points which are not captured by the chosen regression model can be vital for actual plant operation meaning that their consideration in the modeled uncertainty set is critical. Therefore, the non-consideration of such operating points leads to the possibility that the flexibility index based on single equation models overestimates the flexibility of a process. In this context, we want to define overestimation of flexibility as the situation where operating points which should be feasible according to the analysis (i.e. the values of the independent uncertain parameters are within the interval $[\delta_{corr}\Delta\theta_{ind,i}^-, \delta_{corr}\Delta\theta_{ind,i}^+]$) are indeed outside of the feasible region and thereby not feasible. This has been visualized in Fig. 1. A numerical example where the deterministic flexibility index based on a single equation model overestimates the flexibility of a given process is presented in Section 4.1.

We conclude that the aforementioned overestimation is a consequence of the dimension reduction of the uncertainty set (compared to the hyperbox uncertainty set) when utilizing single equation (regression) models. This dimensionality reduction of the uncertainty set can be seen in Fig. 1. When incorporating the single equation (regression) model, the 2-dimensional hyperbox uncertainty set reduces to a 1-dimensional uncertainty set. Also in higher dimensions, the dimensionality reduction can be observed. For example, in the case of three uncertain parameters the 3-dimensional hyperbox can be transformed into a 2-dimensional plane (single parameter dependent on

one independent parameter) or into a 1-dimensional line (single parameter dependent on both independent parameter). The consequence of this dimensionality reduction is that the modeled uncertainty set (significantly) underestimates the expected uncertainty set. Note that overestimating the flexibility of a process can have severe consequences since the infeasibility of certain operating conditions may not be identified before actual operation. Consequently, (very) costly retrofits may be required which are likely to exceed the cost of (unnecessary) overdesign which may occur as a result of underestimating the flexibility of the process.

2.2. Approach based on ellipsoidal uncertainty sets

An alternative approach to account for dependencies between uncertain parameters was suggested by Rooney and Biegler (1999) who suggested to approximate the expected uncertainty set based on the covariance matrix and the mean values of θ (V_θ and $\bar{\theta}$, see (4)). In (4), n^θ is the number of the uncertain parameters (dimension of uncertainty space), and $\chi_{n^\theta}^2(\alpha)$ is the critical value of a χ-squared distribution at the probability level α and with n^θ degrees of freedom. Assuming that the uncertain parameters of a physical system can be described as multivariate Gaussian random variables, Rooney and Biegler (1999) explained that the value of $\chi_{n^\theta}^2(\alpha)$ (obtained by solving (4)) can be interpreted as a joint confidence region of ellipsoidal shape. Such a joint confidence region attempts to enclose all joint parameter combinations up to the desired confidence level, $1 - \alpha$. The authors concluded that uncertainty sets described by (4) (hereinafter referred to as ellipsoidal uncertainty sets) are able to capture (linear) dependencies in the uncertain parameters more accurately compared to hyperbox uncertainty sets.

$$T_{ellip} = \left\{ \theta \mid (\theta - \bar{\theta})^T V_\theta^{-1} (\theta - \bar{\theta}) \leq \chi_{n^\theta}^2(\alpha) \right\} \quad (4)$$

Based on the findings of Rooney and Biegler (1999), Pulsipher and Zavala (2018) reformulated the ellipsoidal set by replacing $\chi^2_{n^\theta}(\alpha)$ with δ yielding a scalable uncertainty set. They claimed that the scalable ellipsoidal set can substitute the hyperbox uncertainty set in the approach by Grossmann and Floudas (1987). Thereby, the largest scaled ellipsoidal can be identified which can be inscribed in the feasible region. Note that the interpretation of the solution of the scale parameter (δ^*) obtained for an ellipsoidal uncertainty set is different compared to the solution obtained with a hyperbox uncertainty set. More precisely, Pulsipher and Zavala (2018) followed the idea of Rooney and Biegler (1999) and utilized δ^* to compute the confidence level (of ellipsoidal shape) based on the p-value calculation of the probability density function of the χ-squared distribution with n^θ degrees of freedom at the critical value δ^*. Furthermore, the authors provided mathematical proof that the obtained confidence level represents a lower bound to the stochastic flexibility index which can be obtained without computationally expensive Monte Carlo Sampling. A conceptual illustration of an ellipsoidal uncertainty set is shown in Fig. 2.

In Fig. 2, the expected distribution of uncertainty is visualized by an ellipsoidal set defined by a given mean value and covariance. It is assumed that this ellipsoidal covers all expected realizations of the uncertain parameters, i.e. it represents the expected uncertainty set. Utilizing the approach by Pulsipher and Zavala (2018), the ellipsoidal uncertainty set was scaled, yielding an ellipsoidal set which is fully inscribed in the feasible region. Fig. 2 shows further how the scale parameter characterizing this scaled ellipsoidal set (δ_{scaled}) can be utilized to approximate the stochastic flexibility index following the p-value calculation of the probability density function of the χ-squared distribution. As a comparison to this scaled ellipsoidal, the area covering all expected operating points which are within the feasible region is also indicated. The latter can be interpreted as a visualization of the stochastic flexibility index especially if it is seen in relation to the expected distribution of uncertainty. Furthermore, Fig. 2 allows for assessing the quality of the approximation of the stochastic flexibility index which is obtained by following the approach by Pulsipher and Zavala (2018). This can be achieved by comparing the difference in size of the in Fig. 2 purple colored shape and the scaled ellipsoidal.

Pulsipher and Zavala (2018) identified that the approach by Grossmann and Floudas (1987) is compatible with any compact set to approximate the uncertainty set as long as the size of the set can be parameterized in terms of the scale parameter, δ. In this context, the authors refer to the work by Li et al. (2011) who presented several scalable uncertainty sets based on the l_∞ norm, l_1 norm and l_2 norm. Note that, e.g., the hyperbox uncertainty set can be seen as a special interpretation of the l_∞ norm (see Li et al., 2011 for further information). However, the interpretation of the scale parameter can be different depending on the chosen uncertainty set, i.e. the results may allow a statistic or a deterministic interpretation as explained in the previous paragraphs. Additionally, both (Li et al., 2011) and Pulsipher and Zavala (2018) only refer to compact sets which can be imagined as regular geometric shapes such as a box, an ellipsoidal or a rhombus. This can cause inexact results when the distribution of operating points, i.e. the expected uncertainty set, cannot (or only partially) be captured by these regular geometric shapes. Consequently, if an ellipsoidal only partially resembles the distribution of operating points, the approach of Pulsipher and Zavala (2018) may only reveal a conservative approximation of the stochastic flexibility index. Note that this conservative approximation is partly comparable to the aforementioned bad resemblance of a hyperbox uncertainty set with the real distribution of uncertainty leading to the consequence that the deterministic flexibility index underestimates the feasible variation range. However, it is beyond the scope of this work to suggest developments of uncertainty sets which allow for a statistical analysis of the flexibility such as ellipsoidal uncertainty sets.

2.3. Similarities and differences of the existing approaches

This Subsection summarizes the essential findings of this Section. Both approaches presented in Sections 2.1 and 2.2 allow for considering parameter dependencies when modeling a scalable uncertainty set which can be integrated in a deterministic framework, namely the MI(N)LP formulations by Grossmann and Floudas (1987), to perform flexibility analysis. However, single equation (regression) models allow for considering dependencies between different uncertain parameters in deterministic flexibility analysis while ellipsoidal uncertainty sets allow for approximating the stochastic flexibility index (which inherently considers dependencies in the uncertain parameters) deterministically. Note, stochastic flexibility analysis, i.e. the stochastic flexibility index, and deterministic flexibility analysis are complementary since different information is obtained. As mentioned in Section 1, the stochastic flexibility index returns the (maximum) probability for feasible operation given the joint probability density function of the uncertain parameters. Knowledge regarding the probability for feasible operation can be advantageous, when comparing different system designs. On the other hand, the stochastic flexibility index does not provide information on the feasibility or infeasibility of specific operating conditions, i.e. to identify if a randomly drawn sample from the expected distribution of uncertainty is feasible, additional analysis is necessary. Such knowledge is provided by deterministic flexibility analysis since the deterministic flexibility index returns the maximum feasible disturbance for each uncertain parameter from a nominal/mean value. Consequently, the deterministic flexibility index can be important for operative decisions and/or problems, e.g. in case an operator needs to identify the feasibility of an expected operating point.

In contrast to the (traditional) Monte Carlo sampling for calculating the stochastic flexibility index, the approach of Pulsipher and Zavala (2018) can explicitly reveal constraint functions which limit the uncertainty set and thereby the flexibility. Traditionally, such knowledge could only be revealed by deterministic flexibility analysis since it is based on identifying the feasible scaling of a compact uncertainty set. The disclosure of such information may favor a certain approach over another since it can be advantageous for identifying and eliminating bottlenecks of processes. On the other hand, the inherent difference between stochastic and deterministic flexibility analysis does not allow for a direct comparison of the obtained results. To illustrate these differences, we calculate the deterministic flexibility index and we also approximate the stochastic flexibility index utilizing the approach of Pulsipher and Zavala (2018) for a numerical example in Section 4.1.

3. Methodology

As an alternative to single equation models, we suggest to formulate uncertainty sets by means of upper and lower boundary functions to capture dependencies in the uncertain parameters. Based on the grouping of the uncertain parameters into independent uncertain parameters (θ_{ind}) and dependent uncertain parameters (θ_{dep}) (see Section 2.1), we reformulated the hyperbox uncertainty set to (5). A conceptual illustration of the deterministic flexibility index based on upper and lower boundary functions in comparison to the hyperbox uncertainty set is shown in Fig. 3.

$$T_{boundary}(\delta) = \begin{cases} \{\theta_i \mid \bar{\theta}_i - \delta\Delta\theta_i^- \leq \theta_i \leq \bar{\theta}_i + \delta\Delta\theta_i^+\} \; \forall \; \theta_i \text{ in } \theta_{ind} \\ \{\theta_j \mid f_l(\theta_{ind}) \leq \theta_j \leq f_u(\theta_{ind})\} \; \forall \; \theta_j \text{ in } \theta_{dep} \end{cases} \quad (5)$$

The advantage of boundary functions in comparison to single equation (regression) models is that the dimensionality of the uncertainty set is not reduced. Boundary functions allow for considering uncertainty even in the dependent parameters since this uncertainty is expressed as space between the upper and lower boundary function(s) (see colored parallelogram in Fig. 3). Due to the additional uncertainty in the dependent parameters, the overestimation of the flexibility can be avoided (compare Figs. 1 and 3). More specifically, the boundary
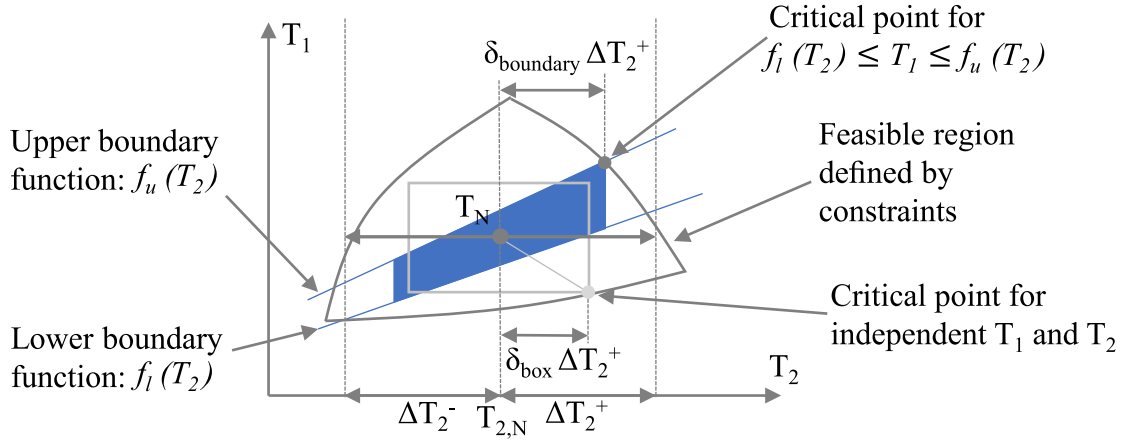
Fig. 3. Conceptual illustration of the flexibility index using boundary functions to capture dependencies in the uncertain parameters.

functions can be chosen is such a way that all expected realizations of the uncertain parameters are enclosed by the modeled uncertainty set. In comparison to the hyperbox uncertainty set, incorporating boundary functions transforms the uncertainty set into a hyperpolygon set. This transformation allows for excluding irrelevant subsets of the hyperbox set, i.e. subsets in which no operation is observed or expected while maintaining the dimensionality of the uncertainty set. When assuming that the boundary functions in Fig. 3 enclose all expected realizations of the uncertain parameters T_1 and T_2, we can identify such subsets in Fig. 3 (not colored parts of the rectangle). Consequently, the incorporation of boundary functions can allow for a better resemblance between the modeled uncertainty set and the real uncertainty set which helps decreasing the risk for underestimating the feasible variation range when calculating the deterministic flexibility index.

In this section, we derive a mixed-integer formulation for the flexibility index based on uncertainty sets defined by boundary functions (see Section 3.1). In this context, we utilize the findings of Pulsipher and Zavala (2018), i.e. that the bi-level formulation of Grossmann and Floudas (1987) (compare (A.2)) is valid for any compact uncertainty set. The derived formulation can be solved by means of the active constraint strategy developed by Grossmann and Floudas (1987) and allows for the generic application of boundary functions when performing deterministic flexibility analysis. In a second step, we focus on the question how to define boundary functions in such a way that the expected or observed uncertainty set is represented as accurately as possible. Such a task can be solved manually (see Langner et al., 2021) but especially for multi-dimensional dependencies, a manual definition may be burdensome and error-prone. We therefore developed an algorithm to automate the definition of boundary functions. In this way, multiple upper and lower boundary functions (in contrast to single upper and lower boundary functions) can be identified, effectively. Consequently, the number of degrees of freedom increases which means that the observed or expected uncertainty set can be approximated with a high accuracy. The proposed algorithm is based on the polygon convex hull and is presented in Section 3.2.

3.1. Mixed-integer formulation for the flexibility index based on uncertainty sets described by boundary functions

As shown in Appendix A, Grossmann and Floudas (1987) developed a MI(N)LP ((A.2) in combination with (A.3a) to (A.3g)) for determining the deterministic flexibility index also for non-vertex solutions of a hyperbox uncertainty set. Based on this work, Pulsipher and Zavala (2018) observed that for any compact set $T(\delta)$, such as $T_{box}(\delta)$ or $T_{ellip}(\delta)$, the derived MI(N)LP is valid (based on mathematical proof provided by Swaney and Grossmann (1985)). Pulsipher and Zavala

(2018) further provided proof that for a scalable ellipsoidal uncertainty set, the solution (θ^*) of (A.2) lies on the boundary of both the feasible region ($\psi(d, \theta) = 0$) and the ellipsoidal uncertainty set, $T_{ellip}(\delta)$. A generalized version of this proof for any compact set is provided in Appendix B.

Since $T_{boundary}(\delta)$ (compare (5)) defines a set around the nominal/mean value(s), the resulting hyperpolygon set is a compact set (similar to $T_{box}(\delta)$ or $T_{ellip}(\delta)$). Following the observation of Pulsipher and Zavala (2018), the MI(N)LP ((A.2) in combination with (A.3a) to (A.3g)) developed by Grossmann and Floudas (1987) can be adapted to incorporate an uncertainty set based on boundary functions. The MI(N)LP formulation to determine the deterministic flexibility index based on uncertainty sets described by boundary functions is given in (6). Note that the complementarity conditions of the inner minimization problem's Karush–Kuhn–Tucker (KKT) conditions have already been replaced in line with the active constraint strategy by Grossmann and Floudas (1987). Further note that in (6), equality constraints and state variables, x, have been considered explicitly, while in (A.2), the state variables x were eliminated by means of the equality constraints.

$$F = \min \delta$$

$$
\begin{aligned}
s.t.\ & g_j(d, x, z, \theta) + s_j = 0, \ j \in J \\
& h_i(d, x, z, \theta) = 0, \ i \in I \\
& \sum_{j \in J} \lambda_j = 1 \\
& \sum_{i \in I} \mu_i \frac{\partial h_i}{\partial z} + \sum_{j \in J} \lambda_j \frac{\partial g_j}{\partial z} = 0 \\
& \sum_{i \in I} \mu_i \frac{\partial h_i}{\partial x} + \sum_{j \in J} \lambda_j \frac{\partial g_j}{\partial x} = 0 \\
& s_j - M(1 - y_j) \leq 0, \ j \in J \\
& \lambda_j - y_j \leq 0, \ j \in J \\
& \sum_{j \in J} y_j \leq n_z + 1 \\
& T_{boundary}(\delta) = \\
& \quad \begin{cases} \{\theta_i \mid \bar{\theta}_i - \delta \Delta \theta_i^- \leq \theta_i \leq \bar{\theta}_i + \delta \Delta \theta_i^+\} \ \forall\ \theta_i \text{ in } \theta_{ind} \\ \{\theta_j \mid f_l(\theta_{ind}) \leq \theta_j \leq f_u(\theta_{ind})\} \ \forall\ \theta_j \text{ in } \theta_{dep} \end{cases} \\
& \delta,\ \lambda_j,\ s_j \geq 0;\ y_j \in 0, 1,\ j \in J
\end{aligned}
\tag{6}
$$

In (6), s_j refers to the slack variable and λ_j refers to the Lagrangian multiplier of the inequality constraint $g_j \leq 0$. Additionally, a binary variable, y_j, indicates if the corresponding inequality constraint, $g_j \leq 0$, is active, i.e. $s_j = 0$ at the found solution. For more detailed information on the derivation of the MI(N)LP formulation, the reader is referred to

Appendix A. Note, that uncertainty sets based on boundary functions can also be integrated in the MI(N)LP formulation for the flexibility test problem.

3.2. Algorithm for identification of boundary functions to approximate the expected uncertainty set

The main idea of defining boundary functions is to establish a tight representation of an expected or observed uncertainty set. An uncertainty set can be visualized by plotting all operating points observed/expected. The distribution of operating points allows for identifying dependencies in the uncertain parameters. On the other hand, manual analysis of operating point distributions may be tedious for problems with more than two uncertain parameters especially if there exist multi-dimensional dependencies. To overcome this difficulty, we suggest to calculate a numerical metric to determine the statistical relationship among the present data such as correlation coefficients. A commonly used correlation coefficient is the coefficient established by Pearson (7) which determines the strength and the direction of the linear relationship between two random variables (Boslaugh and Watters, 2008).

$$\rho(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \qquad (7)$$

In (7), X and Y describe two random variables forming a population (or 2-dimensional data set) and σ denotes the standard deviation of each random variable. Furthermore, the covariance between the two random variables is expressed by $cov(X, Y)$. The calculation of such correlation coefficients can be automated and large data sets can rapidly be evaluated and analyzed for dependencies. Once certain indication exists for a dependency between two or more uncertain parameters, the uncertain parameters need to be classified as dependent or independent. This classification can be enhanced by background knowledge about the process of interest to identify the origin of the dependencies. If this background knowledge is not available, an arbitrary classification may be considered. The identified dependencies can be formulated in the following general form:

$$\theta_j = f(\theta_{i,1}, \theta_{i,2}, \ldots), \ \theta_i \in \theta_{ind}, \ \theta_j \in \theta_{dep}. \qquad (8)$$

Having identified dependencies and classified the uncertain parameters into dependent and independent, we can proceed with the formulation of upper and lower boundary functions. These upper and lower boundary functions can be chosen freely (as long as the compactness of the uncertainty set is preserved, see Section 3.1) but we assume that linear functions present a good compromise between level of accuracy and computational expenses. To automate the definition of boundary functions, we suggest to reduce multi-dimensional dependencies to their 2-dimensional projections and apply the Quick Hull algorithm (see e.g., Barber et al., 1996). For a 2-dimensional data set, the Quick Hull algorithm yields the vertices of the polygon convex hull (in clockwise or counter-clockwise order, see Bykat (1978)). We suggest to use the identified vertices of the polygon convex hull to identify potential boundary functions. In this context, we define a potential boundary function by the linear function which spans between two neighboring vertices. Consequently, for each 2-dimensional projection, the maximum number of boundary functions is given by the number of vertices: $N_{bound,max} = N_{vertices}$. For illustrative purposes, Fig. 4 shows the distribution of a 2-dimensional data set, the respective vertices identified with the Quick Hull algorithm and the potential boundary functions as the edges of the polygon convex hull of the data set.

In general, including all available boundary functions in (6) leads to the best possible approximation of the expected/observed uncertainty set and thereby also to the most accurate value of the flexibility index (based on the polygon convex hull). However, including all available boundary functions may lead to increasing problem complexity for high numbers of dependent uncertain parameters. Additionally,
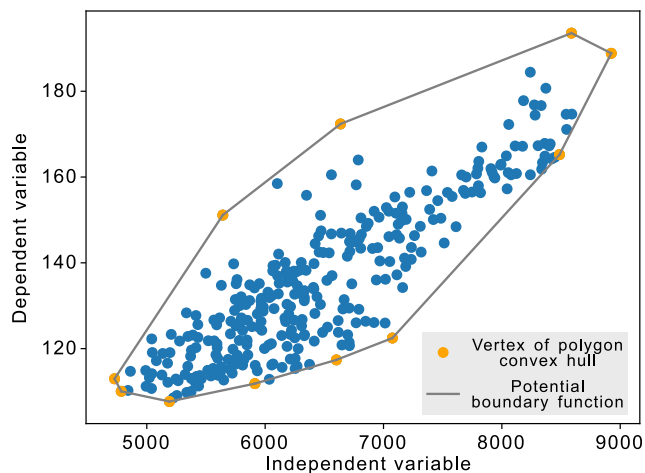

Fig. 4. Illustration of the polygon convex hull for a 2-dimensional data set.

we assume that some bounds of the polygon convex hull (potential boundary functions) have a stronger influence on the solution of (6) when included than other potential boundary functions. Therefore, we formulated an algorithm to identify the most influencing boundary functions, i.e. those boundary functions which capture the individual characteristics of the expected/observed uncertainty set, for a given maximum number of boundary functions. In this context, the influential character of each potential boundary functions is assessed by means of a selection criterion. The flowchart of this algorithm is given in Fig. 5 and further information also regarding the selection criterion is provided in the following paragraphs.

The boundary function selection algorithm needs two input arguments, the 2-dimensional data set and the number of upper and lower boundary functions, respectively, which the algorithm should return. To be able to classify the different boundary functions as lower and upper bounds, it is essential that the abscissa (x-coordinate) of the data set corresponds to the independent parameter while the ordinate (y-coordinate) describes the dependent parameter. In the first step, the vertices of the polygon convex hull are determined using the Quick Hull algorithm on the 2-dimensional data set. It is vital, that the Quick Hull algorithm is initialized in such a way that the order (clockwise or counter-clockwise) of the polygon convex hull vertices is known. Next we identify if the linear function spanned by two neighboring vertices forms an upper or a lower bound of the convex hull. Here we utilize the order in which the vertices were returned by the Quick Hull algorithm. More precisely, we check if the value of the independent parameter (abscissa of data set) of a vertex is smaller or larger than the value of the independent parameter of the following (neighboring) vertex. If the vertices were returned in counter-clockwise order and the value of the independent parameter of a vertex is smaller than the independent parameter value of the following (neighboring) vertex, the two vertices must form a lower bound. If the independent parameter value of a vertex is larger than the independent parameter value of the following vertex, they must form an upper bound. In the case that the independent parameter values of both vertices are the same, we discard the bound defined by these two vertices since it cannot be classified as upper or lower bound.[1] This step yields two lists of neighboring vertices, the first comprising all upper and the second all lower bound vertices. Note that, consequently, there is a maximum number of lower and a maximum number of upper boundary functions which is defined

---

[1] Note that such "steps" are only possible at left or right end of a distribution since otherwise the enclosing shape would not be convex.

Fig. 5. Algorithm to identify a predefined number of upper and lower boundary functions for 2-dimensional data. Note that counter-clockwise order of the vertices of the polygon convex hull was assumed.

by the number of vertex pairs in each list. This implies that the number of upper and/or lower boundary functions returned by the algorithm may be different from the requested number if the number of requested boundary functions exceeds the maximum number of lower and/or the maximum number of upper boundary functions. In the final step, the upper and lower boundary functions are ranked with respect to a predefined selection criterion. Different selection criteria are possible. In this paper, we investigated two selection criteria which are listed below:

- the Euclidean distance (l_2 norm) between neighboring (or following) vertices given that all data values had been normalized between 0 and 1 (hereinafter referred to as the selection criterion *Euclidean distance*),
- the number of data points which are above/below a bound (hereinafter referred to as the selection criterion *Counting of points*).

In this context, the number of data points which are above/below a certain bound can be identified by counting the number of data points for which the value of the independent parameter is between the values (of the independent parameter) of the vertices forming the lower/upper bound. If the *Euclidean distance* is chosen as selection criterion, a bound is ranked higher if the distance between the neighboring vertices is larger compared to another bound. If *Counting of points* is chosen as selection criterion, a bound is ranked higher if the number of data points which are found above or below the respective bound is larger compared to another bound. Finally, the highest ranked upper and lower boundary functions are returned by the algorithm with respect to the requested number of boundary functions (if more candidate functions are available than requested).

With respect to our previous assumption regarding the influence of different boundary functions on the solution of (6), we can state that for any suitable selection criterion the flexibility index should
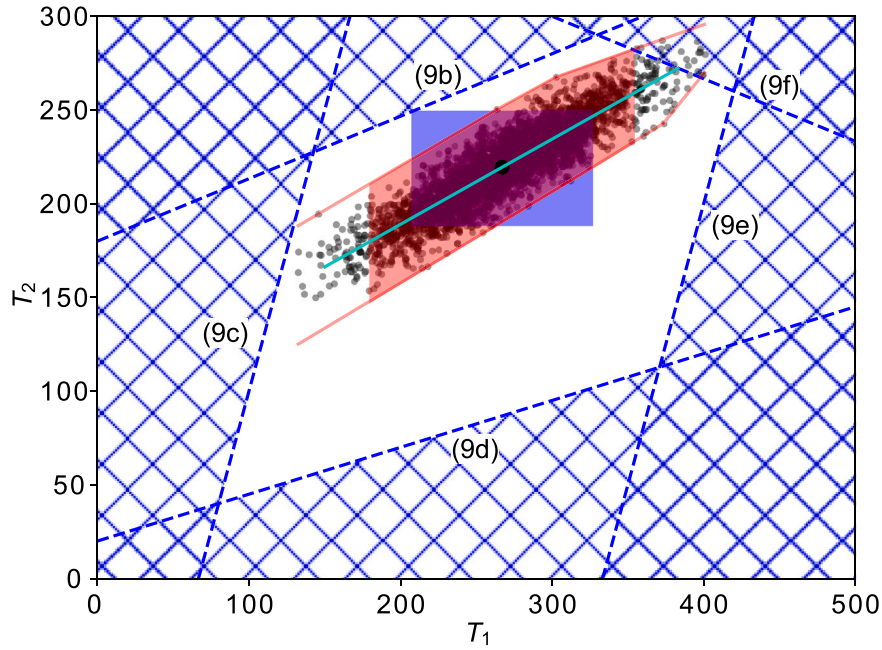
Fig. 6. Visualization of the feasible region of the illustrative example, the data points of the assumed distribution and the results of the deterministic flexibility index calculation for the (hyper) box uncertainty set (blue), the uncertainty set based on a linear regression function (cyan) and the uncertainty set based on boundary functions (red). Note that the labeling of the constraint functions refers to the equation numbering in the paper.

converge to the same value for $N_{bound,requested} < N_{bound,max}$. Consequently, when increasing the number of boundary functions (which should be returned) the choice of the selection criterion should become less important since the probability increases that the same boundary functions are selected (although the order of the selection may differ). In Section 4.2, the flexibility index is calculated for different numbers of boundary functions and both selection criteria.

4. Numerical examples

In this Section, we present two numerical examples to illustrate the calculation of the flexibility index based on boundary functions. All calculations were performed on a Intel Core i7 with 32 GB installed RAM using BARON 21.1.13 (Sahinidis, 2017).

4.1. Illustrative example — comparison of different modeled uncertainty sets

To compare the differences in the results obtained with the traditional hyperbox uncertainty set to the results obtained for the uncertainty set based on single equation models as well as the uncertainty set based on boundary functions, an illustrative example was developed. Additionally, the ellipsoidal uncertainty set is utilized to approximate the stochastic flexibility index to illustrate the essential differences between stochastic and deterministic flexibility analysis. Consider the following example with two uncertain parameters (T_1 and T_2), one state variable (T_3), one equality and five inequality constraints. The uncertain parameters T_1 and T_2 are expected to vary around a given mean vector $(267, 219)$:

$$T_2 - 2 * T_3 = 0 \tag{9a}$$

$$2 * T_3 - (\frac{1}{3} * T_1 + 180) \leq 0 \tag{9b}$$

$$T_2 - (3 * T_1 - 200) \leq 0 \tag{9c}$$

$$\frac{1}{4} * T_1 + 20 - T_2 \leq 0 \tag{9d}$$

$$3 * T_1 - 1000 - 2 * T_3 \leq 0 \tag{9e}$$

$$T_2 - (-\frac{1}{3} * T_1 + 400) \leq 0 \tag{9f}$$

As operating data, artificially generated data representing a multivariate Gaussian distribution for the given mean vector, an absolute positive and negative disturbance of $(\pm 134, \pm 68)$ and a Pearson correlation coefficient (compare (7)) of 0.9 was assumed. The active constraint strategy proposed by Grossmann and Floudas (1987) was utilized to solve the flexibility index problem for the various uncertainty sets. In all problem formulations in Section 4.1, the Big-M parameter value was set to 3000.

For the given set of observed operating points, the deterministic flexibility index based on the hyperbox approach (compare (A.2)) yields a value of 0.44 with constraint function (9b) being active at the solution as can be seen in Fig. 6. When expressing the dependency between the two uncertain parameters T_1 and T_2 by means of a linear regression model, a value of 0.87 is obtained for the flexibility index which is significantly higher. In line with Section 2.1, the obtained result indicates that feasible operation is possible within $[0.87 \Delta T_1^-, 0.87 \Delta T_1^+]$. The deterministic flexibility index was further calculated for an uncertainty set based on boundary functions following our proposed approach presented in Section 3. The number of upper and lower boundary functions was set to three and the *Euclidean distance* was chosen as selection criterion. Again, it was assumed that T_1 is independent while the uncertainty of T_2 could be expressed by means of (upper and lower) boundary functions depending on T_1. For three upper and lower boundary functions, the flexibility index was calculated using (6) and a value of 0.65 was received which means that, based on all observed data points, feasibility can be guaranteed if the independent parameter T_1 varies between $\pm(0.65 * 134)$. The obtained feasible region for the flexibility analysis based on boundary functions is shown in red in Fig. 6. Note that a different number of boundary functions may yield a different value for the flexibility index. This is further investigated in Section 4.2.

Fig. 6 shows three different results for the deterministic flexibility analysis which originate from different modeling strategies for the
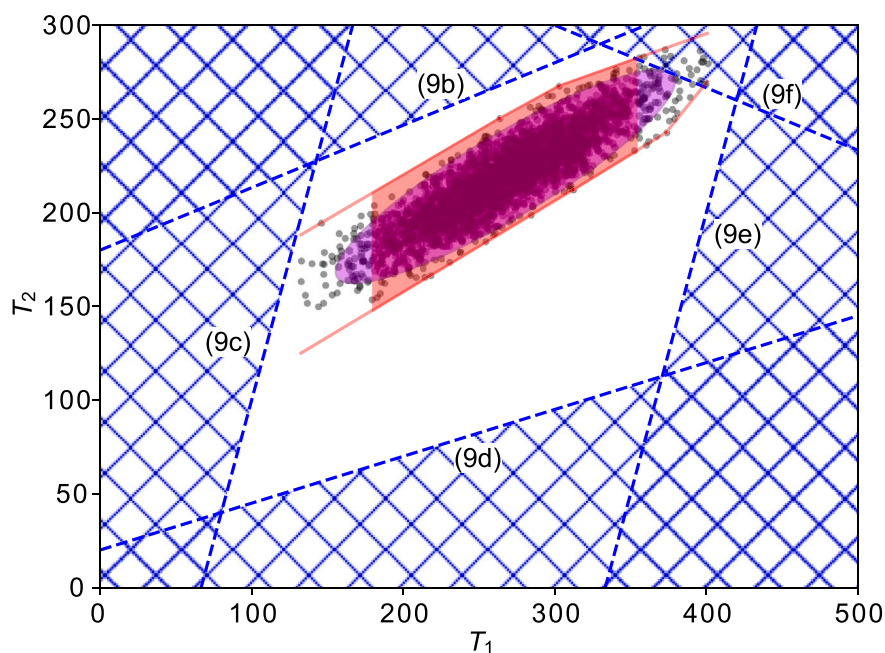
Fig. 7. Visualization of the feasible region of the illustrative example, the data points of the assumed distribution and the results of the deterministic flexibility index calculation for the uncertainty set based on boundary functions (red) as well as the approximation of the stochastic flexibility index by means of an ellipsoidal uncertainty set (purple). Note that the labeling of the constraint functions refers to the equation numbering in the paper.

expected uncertainty set. In this context, Fig. 6 illustrates that modeling the expected uncertainty set using boundary functions derived from the convex hull allows for the best resemblance of the set of observed operating points compared to the box uncertainty set and the uncertainty set based on the linear regression model. When comparing the boundary function uncertainty set with the box uncertainty set, we see that two triangular-shaped subsets of the box uncertainty set were excluded from the analysis since no operating points were observed in these subsets. Additionally, Fig. 6 shows that the scaling of the box uncertainty set was limited by an operating point which is not expected (T_1 at its lowest expected value while T_2 is at its highest expected value). Consequently, it cannot be expected that constraint function (9b) limits the flexibility of the process described by constraint functions (9a)–(9f). More precisely, constraint function (9b) is not limiting the given distribution, but rather the mathematical model of a (hyper)box (see Fig. 6). On the other hand, the feasible uncertainty set based on boundary functions was limited by constraint function (9f), and Fig. 6 shows that this constraint function de facto limits the distribution and thereby also the flexibility of the process. Consequently, a good representation of the observed data is vital to identify the real bottleneck(s) in a system to be able to initiate the best possible countermeasures (if necessary).

Lastly, the highest value for the deterministic flexibility index was obtained for the uncertainty set based on the linear regression function. However, Fig. 6 shows operating points outside of the feasible region even for values of T_1 that lie within the interval $\pm(0.87 * 134)$ which was indicated to be feasible according to the analysis. More precisely, a rigorous analysis of all operating points visualized in Fig. 6 reveals that for 7 samples the value of T_1 is within the interval $\pm(0.87 * 134)$ while those samples are outside the feasible region. Consequently, the obtained result of the deterministic flexibility index overestimates the flexibility of the process, since the uncertainty set based on the linear regression function significantly underestimates the set of observed operating points. Note that the identified number of samples which were indicated to be feasible according to the analysis, but which are de facto infeasible, was provided for illustrative purposes only. Further note that the corresponding fraction of the total number of operating points was intentionally not presented to avoid confusion between stochastic and deterministic flexibility analysis.

Hereafter, we summarize the findings presented in Fig. 6. We obtained three different numerical values for the deterministic flexibility index depending on the modeling approach of the expected uncertainty set:

- 0.44 for the box uncertainty set,
- 0.87 for the uncertainty set based on the linear regression function,
- 0.65 for the uncertainty set based on boundary functions

However, as shown in Fig. 6, a value of 0.87 overestimates the feasible variation range of the (independent) uncertain parameter T_1 since operating points within this interval are found to be infeasible. Due to the aforementioned potentially severe consequences of overestimating the feasible variation range, linear regression functions are discarded from further analysis. Since overestimating the feasible variation range cannot occur when using the box uncertainty set (see Swaney and Grossmann, 1985 for extensive mathematical proof) or the uncertainty set based on boundary functions (see Section 3), both approaches are suitable for calculating the deterministic flexibility index. In this context, the larger value of the flexibility index obtained for the uncertainty set based on boundary functions can be seen as the most exact quantification of the feasible interval in which the (independent) uncertain parameter T_1 may vary while achieving feasible operation. This conclusion is also illustrated in Fig. 6 which shows that the box uncertainty set approximates the actual distribution of uncertainty with less accuracy than the uncertainty set based on boundary functions.

To illustrate the different interpretation of the use of the ellipsoidal uncertainty set compared to the deterministic methods, the covariance matrix of the data set was calculated (see (10)).

$$\begin{bmatrix} 2045.6 & 929.9 \\ 929.9 & 527.1 \end{bmatrix} \tag{10}$$

Following the approach of Pulsipher and Zavala (2018), a solution of 6.1 was found for δ_{ellip}. Calculating the left-tail p-value for the probability density function of the χ-squared distribution with $n^\theta = 2$ (degrees of freedom) returns a confidence level of 95.3%. As mentioned in Section 2.2, this result approximates the stochastic flexibility index,

meaning that a random sample (i.e. a random operating point) drawn from a multivariate Gaussian distribution which is characterized by the given mean vector and covariance matrix is feasible with a probability of 95.3%. The scaled ellipsoidal uncertainty set is visualized in Fig. 7. To assess the quality of the approximation of the stochastic flexibility index, we determined the share of the expected operating points within the feasible region. Note that in general rigorous Monte Carlo Sampling is necessary to determine the stochastic flexibility index, i.e. evaluate the feasibility of randomly drawn realizations of the uncertain parameters from the expected distribution of uncertainty. However, it was decided that the feasibility evaluation of the already drawn samples is sufficient for these illustrative purposes. The evaluation revealed that the probability for feasible operation is 99.3%. Although the difference between the approximation and the (more rigorous) determined stochastic flexibility index is not negligible, the quality of the obtained approximation may be considered satisfactory.

Although the obtained feasible uncertainty set based on boundary functions as well as the feasible ellipsoidal set can be graphically compared as shown in Fig. 7, the information obtained by the different analyses is inherently different (see above). As mentioned in Section 2.3, the obtained information is complementary. On the other hand, the approach of Pulsipher and Zavala (2018) is also based on an uncertainty set described by a geometric shape, and given Fig. 7 it seems that both uncertainty sets, namely the purple colored ellipsoidal as well as the red colored polygon, are of similar geometric size, and, furthermore, that the number of operating points covered by these shapes is of similar magnitude. This latter observation can be mathematically confirmed by calculating the share of operating points which are covered by the respective shape from the total number of feasible operating points. This share is 95.7% for the purple colored ellipsoidal uncertainty set and 95.0% for the red colored uncertainty set based on boundary functions. Consequently, we can conclude that for the given example the ellipsoidal uncertainty set is a good choice to model the expected distribution of uncertainty. This was expected since the (artificial) data originates from a multivariate Gaussian. The situation may be different if the expected distribution of uncertainty deviates from a Gaussian distribution. For the uncertainty set based on boundary functions, this share provides additional information, i.e. that 95.0% of the feasible operating points are found within the interval in which T_1 deviates with $\pm(0.65 * 134)$ from its given mean value. This information is complementary to the results of the deterministic flexibility analysis since probability analysis is not part of deterministic flexibility analysis. It should be noted that such information was straightforward to obtain for the investigated example, but may be more burdensome to obtain for more complex examples characterized by more than two uncertain parameters.

Finally, note that the limiting constraint identified using the ellipsoidal uncertainty set (constraint function (9f)) is identical to the one identified using the uncertainty set based on boundary functions. We previously identified that constraint function (9f) indeed limits the flexibility of the process described by constraint functions (9a)–(9f). Consequently, this observation is again an indication that both uncertainty sets show a good resemblance with the expected distribution of uncertainty/uncertainty set.

4.2. Heat exchanger network example

By means of a second numerical example, we aim to illustrate the applicability of the suggested approach (see Section 3) in cases of multidimensional dependencies in the uncertain parameters. Additionally, we investigate the influence of the two selection criteria for identifying boundary functions (presented in Section 3.2) on the deterministic flexibility index. The example of interest is a HEN and it is shown in Fig. 8. This HEN-example is well-known in the flexibility analysis literature and slightly different versions can be found, e.g. in Saboo
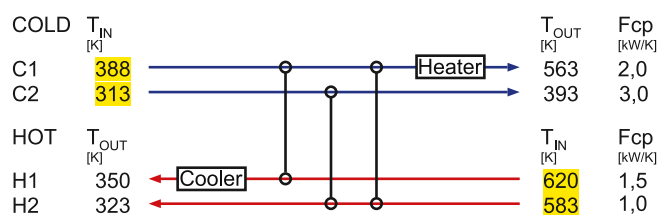

Fig. 8. Grid diagram representation of the heat exchanger network example.

Table 1

Flexibility analysis results of the heat exchanger network example with respect to the number of upper/lower boundary functions and the selection criterion for the boundary functions.

Number of boundary functions	Selection criteria	Flexibility index
1	Euclidean distance	0.66
	Counting of points	0.66
2	Euclidean distance	0.73
	Counting of points	0.69
3	Euclidean distance	0.73
	Counting of points	0.73
4	Euclidean distance	0.74
	Counting of points	0.73
5	Euclidean distance	0.74
	Counting of points	0.74

Big-M value: 200; solver BARON, solver tolerance: 1E-6.

et al. (1985), Grossmann and Floudas (1987) and Ochoa and Grossmann (2020). In contrast to the version used by Grossmann and Floudas (1987) a heater was added to control the target temperature of stream $C1$ while the target temperature of stream $H2$ was fixed. For the flexibility analysis, it was assumed that the start temperatures of the four streams vary with ±10 K around the given mean values (see highlighted values in Fig. 8). The flexibility index for the hyperbox uncertainty set was calculated using the active constraint strategy (Grossmann and Floudas, 1987), and a value of 0.5 was obtained (Big-M parameter value: 200).

To illustrate the influence of parameter dependencies, it was assumed that the uncertain start temperatures of streams $H2$, $C1$ and $C2$ show correlating trends. Fig. 9 shows an artificially generated distribution of operating points. Based on Fig. 9, the start temperature of stream $H2$ was chosen to be dependent since it shows strong dependencies with the start temperatures of streams $C1$ and $C2$ (the start temperatures of streams $C1$ and $C2$ show also a dependency with each other, but it is somewhat weaker).

To consider the identified dependencies in the start temperatures of streams $H2$, $C1$ and $C2$ when computing the flexibility index, boundary functions were derived following the algorithm presented in Section 3.2. Both selection criteria for identifying the most influencing boundary functions (i.e. those boundary functions which capture the individual characteristics of the observed uncertainty set) mentioned in Section 3.2 were investigated and selected results of the analysis are shown in Table 1. The values for the flexibility index shown in Table 1 were obtained by applying the active constraint strategy (Grossmann and Floudas, 1987) for solving (6). Note that an uncertainty set based on single equation regression models was not considered due to the previously identified risk that the obtained deterministic flexibility index overestimates the flexibility.

The results in Table 1 show that considering the dependencies in the start temperatures of streams $H2$, $C1$ and $C2$ has a significant influence on the result of the flexibility analysis (flexibility index for the hyperbox uncertainty set: 0.5). As mentioned in Section 3.2, considering more boundary functions increases the accuracy of the approximation of
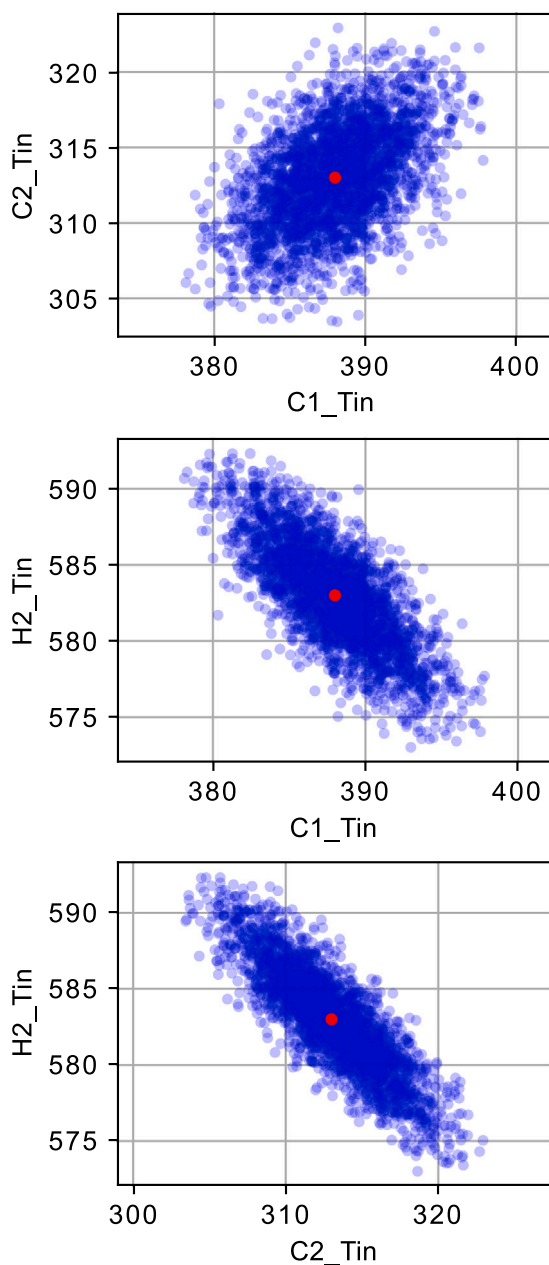
	Max number of lower boundary functions $N_{lower,max}$	Max number of upper boundary functions $N_{upper,max}$
$H2\_Tin = f(C1\_Tin)$	6	8
$H2\_Tin = f(C2\_Tin)$	8	11

$H2\_Tin = f(C1\_Tin)$ and $H2\_Tin = f(C2\_Tin)$ was reached while no influence on the flexibility index was identified. The maximum number of lower and the maximum number of upper boundary functions for the two dependencies $H2\_Tin = f(C1\_Tin)$ and $H2\_Tin = f(C2\_Tin)$ is given in Table 2. This observation supports our assumption that some bounds of the polygon convex hull do not influence the result of (6). Additionally, since for both selection criteria the flexibility index converges for $N_{bound,requested} < N_{bound,max}$ (for both dependencies), we can conclude that both selection criteria are suitable for the given problem (i.e. with both selection criteria the most influencing boundary functions could be identified before reaching the maximum number of boundary functions).

5. Industrial case study

As mentioned in Section 1, dependencies between uncertain parameters can be expected, especially in industrial case studies. Since problem size and complexity often go hand in hand, it is rather likely that when defining the system boundaries of an industrial case study (and thereby also the input parameters to the system of interest which often are subject to some uncertainty) upstream dependencies between these input parameters can be missed. However, ignoring dependencies in input parameters (i.e. modeling the expected uncertainty set as a hyperbox) can lead to a significant underestimation of the flexibility, as previously discussed and as illustrated by the numerical example in Section 4.1. Consequently, equipment may be overdesigned or design proposals may even be discarded since they seem to be inherently inflexible.

In the following, we will analyze the flexibility of a design proposal for a HEN used to recover process heat through a warm and hot water system in a Swedish kraft pulp mill. Measurement data representing real mill operating conditions was available from a previous study of the mill in question (Persson and Berntsson, 2009). The design proposal for the HEN is shown in Fig. 10. Fig. 10 shows a system of 14 process streams (five cold streams and nine hot streams) which are interconnected by heat exchangers. Five of the hot streams (H02, H03, H04, H05 & H06) are describing energy release during a phase-change (condensation). The operational target of this HEN is to heat the five cold streams to their defined target temperatures. The cooling of the respective hot streams is not critical for operation, i.e. the target temperatures of the hot streams shown in Fig. 10 are soft target temperatures with the presented value being the lowest feasible value.

In addition to the target temperatures of the streams, Fig. 10 also provides information on the heat capacity flow rates and the start temperatures of the 14 process streams (for condensing streams, a temperature drop of 1 K was assumed and the corresponding heat capacity flow rate in Fig. 10 correspond to the heat load released during condensation). While the operational target, i.e. the target temperatures of the cold streams, is fixed, variation is observed for some of the heat capacity flow rates (condensation heat loads) and start temperatures. These uncertain parameters are highlighted in yellow and the values shown in Fig. 10 correspond to the mean values. Operating data was collected and for each uncertain parameter 313 data points were available to identify the uncertainty set.

To reduce the complexity of the flexibility index problem formulation, the system of heat exchangers was divided into its four

Fig. 9. Distribution of operating points of the heat exchanger network example; the mean values are highlighted in red.

the observed uncertainty set. Table 1 shows that when increasing the number of boundary functions, the flexibility index increases up to 0.74 meaning that variations (in the independent parameters) of ± 7.4 K are feasible (compared to variation of ± 5 K calculated with the hyperbox uncertainty set). Furthermore, for lower numbers of boundary functions some influence of the selection criterion on the flexibility index can be observed. However, it must be noted that this is a highly case-specific observation and no conclusion can be drawn regarding which selection criterion should be preferred. Additionally, note that the flexibility index converges to the same value (0.74), for higher numbers of boundary functions. The number of boundary functions was further increased until the maximum number of lower and the maximum number of upper boundary functions for the two dependencies
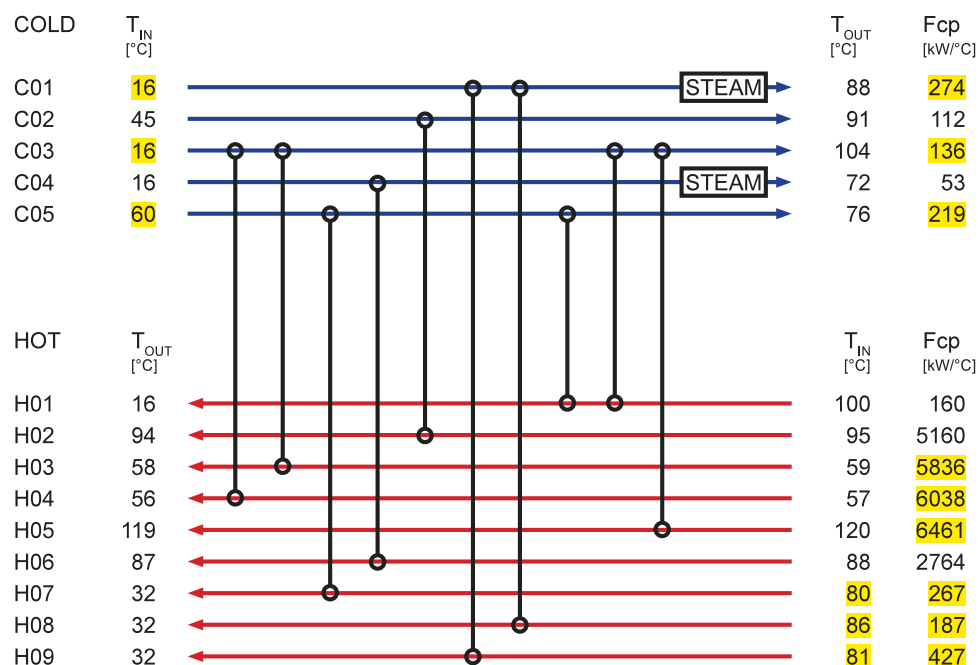
Fig. 10. Grid diagram representation of the design proposal for the heat integration project.

Table 3
Overview of independent subsystems and respective process streams present in the studied system.

Subsystem	Cold process streams	Hot process streams
1	C01	H08, H09
2	C02	H02
3	C03, C05	H01, H03, H04, H05, H07
4	C04	H06

Table 4
Mean values and maximum observed positive/negative deviation from the mean value of the uncertain parameters of the studied system.

θ	$\bar{\theta}$	$\Delta\theta_i^+$	$\Delta\theta_i^-$
C01_Tin [K]	289,62	7,73	8,37
C01_Fcp [kW/K]	274,23	83,80	112,83
C03_Tin [K]	289,62	7,73	8,37
C03_Fcp [kW/K]	135,75	57,80	28,08
C05_Tin [K]	333,49	4,46	9,84
C05_Fcp [kW/K]	219,36	44,04	175,94
H03_Load [kW/K]	5836,01	9273,89	2116,65
H04_Load [kW/K]	6038,76	3492,10	4179,08
H05_Load [kW/K]	6461,99	2462,55	1734,44
H07_Tin [K]	352,79	1,36	8,64
H07_Fcp [kW/K]	267,28	141,54	202,96
H08_Tin [K]	359,15	6,00	8,00
H08_Fcp [kW/K]	186,83	126,96	140,95
H09_Tin [K]	353,82	5,33	6,67
H09_Fcp [kW/K]	427,02	134,01	217,25

independent subsystem(s). The streams of each independent subsystem are listed in Table 3. As we can see by comparing Fig. 10 and Table 3, uncertain parameters are only present in subsystems 1 and 3 (uncertain parameters are highlighted yellow in Fig. 10). We can further conclude that only subsystem 3 is critical for the flexibility of the overall system since the operational target in subsystem 1 (target temperature of stream C01) is controlled by means of hot utility steam (which is assumed to be available in necessary amount). Summarizing, the flexibility analysis of the system of interest can be performed by reducing the problem to subsystem 3.

In a first step of the flexibility analysis, the flexibility index was calculated using the hyperbox set of the uncertain parameters. To define the hyperbox uncertainty set, the largest positive and negative deviation from the mean value of each data series was identified and these values are presented in Table 4.

Using the active constraint strategy and the hyperbox uncertainty set, a value of 0.64 was obtained for the flexibility index (Solver: BARON; Big-M parameter value: 200). This flexibility index value indicates that the design proposal will be able to handle 64% of the observed variation, e.g., the heat capacity flow rate of stream C01 may be allowed to vary by +53.63 kW/K and −72.21 kW/K from the mean value given in Fig. 10. Consequently, the design proposal would need to be reworked or other measures would be necessary to control the variation of the uncertain parameters (in case the operational targets of the system of interest must be met).

In the next step, the data series were analyzed to identify possible dependencies in the uncertain parameters. In Fig. 11(a), a heat map visualizes Pearson's correlation coefficient (compare (7)) calculated for all pairs of uncertain parameters. Darker colors show a stronger dependency in the parameters (with positive or negative sign of the

correlation coefficient) while lighter colors indicate a weaker dependency. The diagonal represents pairs of the same data series and, thus, here the strongest possible agreement can be observed. In this context, we can conclude that the start temperatures of streams C01 and C03 are identical which can be explained by the fact that both streams are water streams originating from the same source. Consequently, for the final flexibility analysis the start temperatures of streams C01 and C03 were merged into one uncertain parameter. To allow a better overview of the data series pairs which are characterized by a strong dependency, in Fig. 11(b), only values of the Pearson's correlation coefficient which are greater or equal to 0.6 are highlighted in color. Using Fig. 11(b), we can identify four dependencies between different uncertain parameters. The observed dependencies are listed in Table 5.

Based on these findings, we can summarize that dependencies in the uncertain parameters are present and that the flexibility index should be re-calculated considering these dependencies. This way we could analyze if ignoring the observed dependencies in the uncertain parameters (flexibility analysis based on hyperbox uncertainty set) has led to an underestimation of the flexibility. The proposed approach (presented in Section 3) was applied, aiming to transform the hyperbox uncertainty set to a multi-dimensional polygon by means of boundary functions.
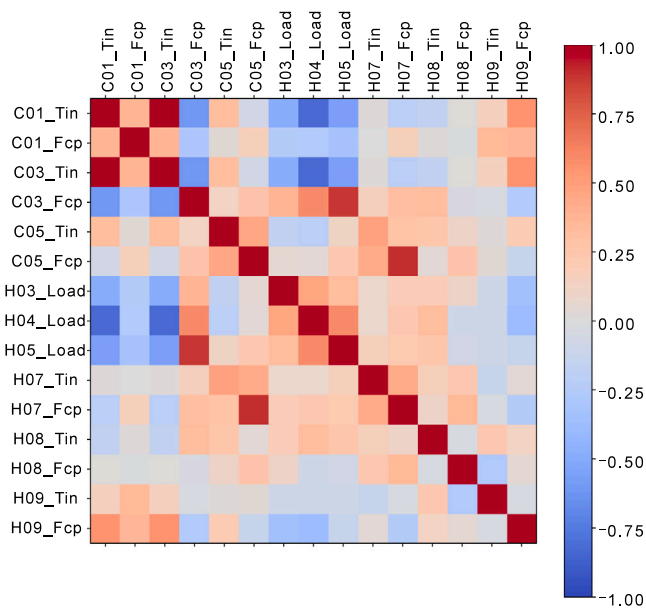
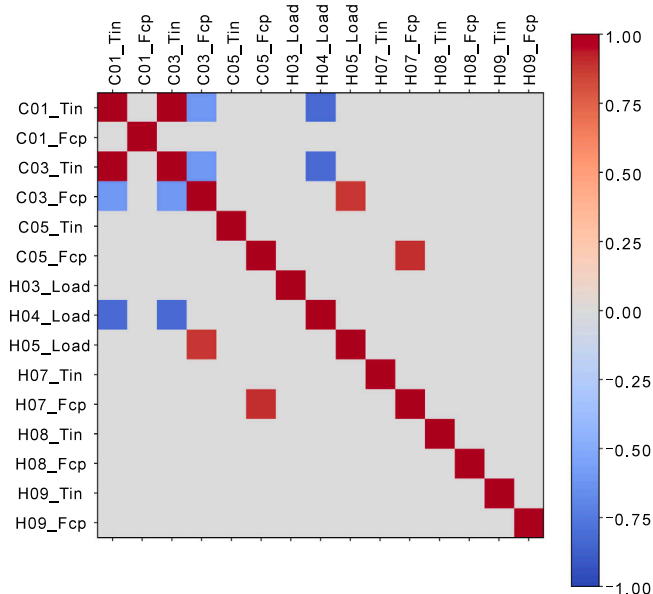(a) Pearson's correlation coefficient for all pairs of uncertain parameters.

(b) Pearson's correlation coefficient for values ≥ 0.6.

Fig. 11. Visualization of the Pearson's correlation coefficient.

Table 5
Observed dependencies between uncertain parameters.

Dependency	Uncertain parameters
1	C03_Fcp, C03_Tin
2	C03_Fcp, H05_Load
3	C05_Fcp, H07_Fcp
4	H04_Load, C03_Tin

Table 5 shows that formerly independent uncertain parameters can be expressed as dependent parameters. The classification into dependent and independent parameters is shown in Table 6. We assessed that it is unreasonable to assume that the temperature of the water source (start temperature of stream C03 and thereby C01) is dependent on the other uncertain parameters listed in Table 5. Consequently, $C03\_Tin$ was classified as an independent parameter. Further, we merged dependencies 1 and 2 into dependency 1* by also classifying the heat

Table 6
Impact of identified dependencies and the number of boundary functions on the flexibility index.

Dependency	Dependent parameter	Independent parameter(s)	Number of boundary functions	Flexibility Index
1*	C03_Fcp	C03_Tin, H05_Load	1	0.68
			2	0.76
			3	0.76
3	C05_Fcp	H07_Fcp	1	0.62
			2	0.63
			3	0.68
			4	0.68
4	H04_Load	C03_Tin	1	0.50
			2	0.55
			3	0.55

Big-M value: 200; solver BARON, solver tolerance: 1E-6.

load of stream H05 as independent. Finally, we arbitrarily classified the heat capacity flow rate of stream H07 as independent. To analyze the impact of the identified dependencies on the flexibility analysis, each dependency was considered individually and the flexibility index was obtained by solving (6) for increasing numbers of boundary functions. The boundary functions have been identified following the algorithm described in Section 3.2 and as selection criterion *Counting of points* was utilized. Note, that the results in Section 4.2 indicate that any suitable selection criterion would be valid since the flexibility index is expected to converge to the same value as the number of boundary functions is increased. The results are given in Table 6.

In line with previous results (see Section 4), we see that increasing the number of boundary functions leads to a higher flexibility index until the value converges. This trend can be explained by the increased accuracy of the approximation of the observed uncertainty set when the number of boundary functions is increased. Additionally, we learn that including dependency 4 returns a smaller flexibility index compared to the value obtained for the hyperbox uncertainty set even when the number of boundary functions was increased. Note that this observation was only made for dependency 4. For dependencies 1* and 3, the flexibility index converges to a value beyond the value found for the hyperbox uncertainty set. To understand this phenomena, the results obtained with dependency 4 were analyzed in more detail.

For the case of dependency 4, we can identify the total feasible uncertainty span in which the dependent parameter $H04\_Load$ varies by means of Fig. 12. Fig. 12 shows the 2-D projection of the observed data points in the space of $H04\_Load$ and $C03\_Tin$. Furthermore, the three upper and lower boundary functions are shown which were identified using the algorithm explained in Section 3.2. For the case of three upper and lower boundary functions, the results in Table 6 indicate that the independent parameters can vary by ±55% of the expected/observed deviation around the mean value(s). For the independent parameter $C03\_Tin$, this feasible variation range is indicated by two dashed lines in Fig. 12. The feasible variation range of the dependent parameter $H04\_Load$ is given by the space between the boundary functions and the combined feasible variation range is visualized as hatched area in Fig. 12. For comparative reason, the 2-D projection of the feasible hyperbox which was obtained for independent variation of the uncertain parameters is also shown in Fig. 12. We could identify that the feasible variation of the dependent parameter $H04\_Load$ comprises the entire expected uncertainty span meaning that, depending on the value of $C03\_Tin$, $H04\_Load$ can deviate with up to +3492.10 kW/K or −4179.08 kW/K from the mean value of 6038,76 kW/K. Concluding, the flexibility index considering dependency 4 by means of boundary functions decreased compared to the flexibility index based on the hyperbox uncertainty set since the feasible variation of the independent parameters was limited to allow for a bigger uncertainty range of the dependent parameter $H04\_Load$. Note, the here described phenomena
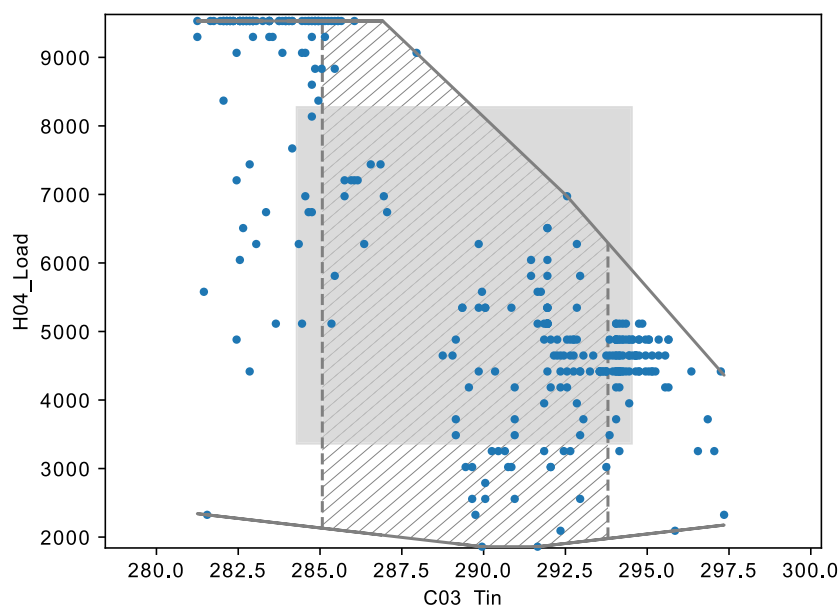
Fig. 12. 2-dimensional projection of the operating points in the space of $H04\_Load$ and $C03\_Tin$ to visualize the dependency between these two parameters. Furthermore, the feasible uncertainty set when including the aforementioned dependency in the flexibility analysis by means of three lower/upper boundary functions is shown as hatched polygon. For comparative reason, the 2-D projection of the feasible hyperbox which was obtained for independent variation of the uncertain parameters is shown as filled rectangle.

Table 7
Flexibility analysis results of the industrial case study when considering the identified parameter dependencies.

Dependency	Dependent parameter (s)	Independent parameter(s)	Number of boundary functions	Flexibility index
1* & 3	C03_Fcp, C05_Fcp	C03_Tin, H05_Load, H07_Fcp	1 2 3 4	0.66 0.77 0.87 0.87
1* & 3 & 4	C03_Fcp, C05_Fcp, H04_Load	C03_Tin, H05_Load, H07_Fcp	1 2 3 4	0.41 0.65 0.84 0.84

Big-M value: 200; solver BARON, solver tolerance: 1E-6.

is highly individual and depends on the distribution of the operating points and the resulting boundary functions of the investigated dependency. A more detailed discussion around this phenomena is provided in Section 6.

As mentioned previously, for dependencies 1* and 3, we can identify that with an increasing number of boundary functions, the flexibility index converges to a value beyond the value found for the hyperbox uncertainty set. This observation indicates that by including dependencies 1* and 3 in the analysis, regions of the hyperbox which limit the flexibility index (but in which no operating points were observed) could successfully be excluded from the analysis. In the next step, the flexibility analysis was repeated including all three dependencies and only dependencies 1* and 3. The results are presented in Table 7.

Considering dependencies 1* and 3 yields 0.87 for the flexibility index for three or more lower/upper boundary functions. Compared to the value of 0.64 calculated for independent uncertain parameters (hyperbox uncertainty set), the result indicates, that the design proposal is more robust towards the observed variation than previously assumed. When including also dependency 4, the flexibility index for three or more lower/upper boundary functions was calculated to 0.84 (see Table 7). Consequently, including all three dependencies reduces the influence of dependency 4. On the other hand, we concluded that

including dependency 4 by means of boundary functions returns a biased result for the flexibility index and therefore excluded dependency 4 from the following analysis (see Section 6 for a detailed discussion).

In the next step, we analyzed the impact of possible outliers on the solution of the flexibility analysis. By means of the unsupervised machine learning algorithm *Isolation Forest*, the three data sets describing the dependencies (compare Table 5: $C03\_Fcp$ and $C03\_Tin$; $C03\_Fcp$ and $H05\_Load$; $C05\_Fcp$ and $H07\_Fcp$) were filtered to identify possible outliers. Using a low contamination ratio of 1% for each data set, 10 potential outliers were identified (3.2% of the total amount of data points) which were removed from the data series. Background information on the algorithm *Isolation Forest* can be found in Liu et al. (2008). Fig. 13 shows the distribution of the three data sets before (blue) and after (orange) filtering.

Excluding the identified outliers has an impact on the hyperbox uncertainty set. In Table 8 the updated values for the mean and for the largest positive and negative deviation from the mean of each data series are presented. Additionally, values are highlighted if they deviate significantly (deviation $\geq abs(1\%)$). Calculating the flexibility index for a hyperbox uncertainty set with the data after filtering yields a value of 0.69 which is marginally higher compared to the value of 0.64 received with the initial data set. On the other hand, when including dependencies 1* and 3 in the flexibility analysis and using the data set after filtering, for three (or more) upper/lower boundary functions a value of 1.01 was calculated for the flexibility index (for both calculations, the Big-M parameter value was set to 200). Consequently, the design proposal can handle all observed operating points assuming that the excluded operating points are outliers resulting of, e.g. faulty measurements.

6. Analysis and discussion

In this Section, we want to analyze and discuss two observations made during the calculations of the deterministic flexibility index using an uncertainty set based on boundary functions which were presented in this work. Finally, also an overview on the manual user input is given which is required when automating the approach suggested in this paper.
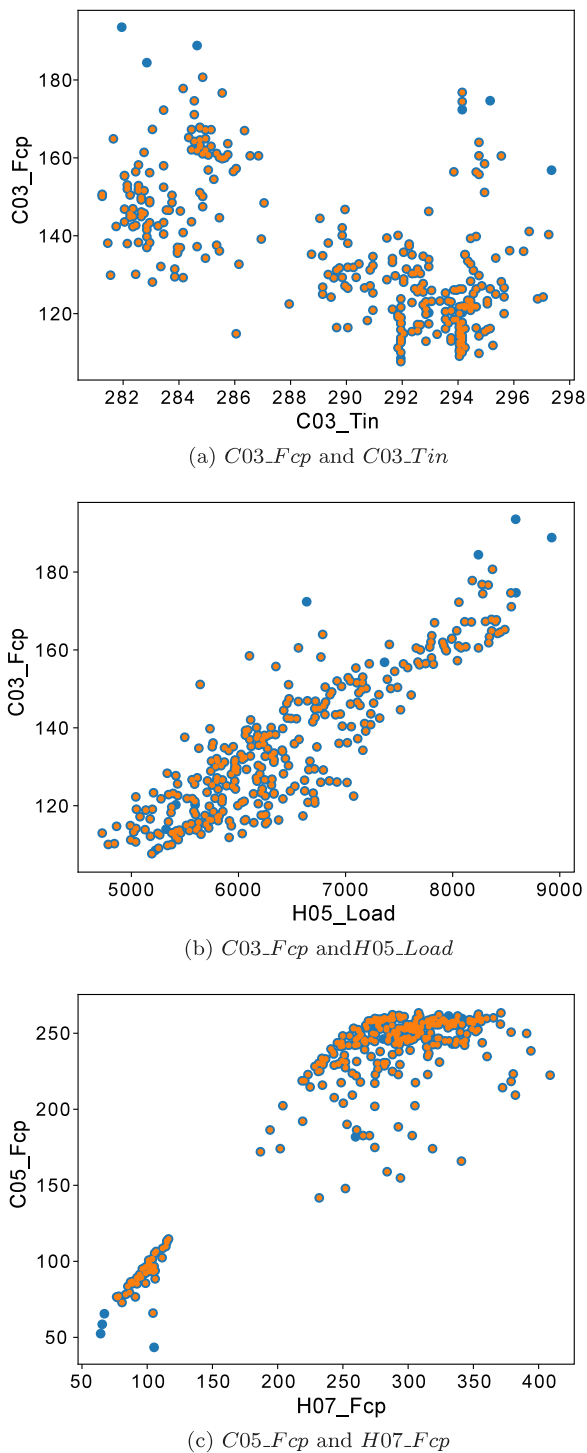
(a) $C03\_Fcp$ and $C03\_Tin$

(b) $C03\_Fcp$ and $H05\_Load$

(c) $C05\_Fcp$ and $H07\_Fcp$

Fig. 13. Distribution of the three data sets before (blue) and after (orange) filtering.

6.1. Analysis of identified critical points

It should be noted that for all but one calculation of the deterministic flexibility index (using an uncertainty set based on boundary functions) in Sections 4 and 5, the obtained solution (θ^*) was a vertex solution of the hyperbox formed by the independent uncertain parameters (θ_{ind}). Only for the case of the filtered data set for the industrial case study in Section 5, the returned solution was a non-vertex solution of the hyperbox formed by the independent uncertain parameters, meaning that some of the independent uncertain parameters were at

Table 8
Mean values and maximum observed positive/negative deviation from the mean value of the uncertain parameters of the studied system after filtering; values are highlighted if they deviate with $\geq abs(1\%)$.

θ	$\bar{\theta}$	$\Delta\theta_i^+$	$\Delta\theta_i^-$
C01_Tin [K]	289,58	7,67	8,33
C01_Fcp [kW/K]	273,96	84,07	112,56
C03_Tin [K]	289,58	7,67	8,33
C03_Fcp [kW/K]	135,17	**45,53**	**27,50**
C05_Tin [K]	333,51	4,44	9,86
C05_Fcp [kW/K]	221,10	**42,30**	**155,15**
H03_Load [kW/K]	5817,64	9292,26	2098,28
H04_Load [kW/K]	6052,40	3478,46	4192,72
H05_Load [kW/K]	6444,55	**2102,60**	**1717,00**
H07_Tin [K]	352,82	**1,33**	**5,67**
H07_Fcp [kW/K]	269,10	**139,72**	**192,50**
H08_Tin [K]	359,13	6,02	7,98
H08_Fcp [kW/K]	187,24	126,55	141,36
H09_Tin [K]	353,82	5,33	6,67
H09_Fcp [kW/K]	426,40	134,63	216,63

a value between $\bar{\theta}_i - 1.01\Delta\theta_i^-$ and $\bar{\theta}_i + 1.01\Delta\theta_i^+$. More precisely, the identified critical value of the independent parameter $H05\_Load$ was 6072.03 kW/K and thereby a non-vertex solution. A further analysis of the results revealed that the critical value of $H05\_Load$ and the identified critical value of the dependent parameter $C03\_Fcp$ form a corner point of the upper boundary functions of dependency 1* (see Fig. 14). Fig. 14 shows that the slope ($m_{bf,3}$) of the boundary function for values > 6072.03 kW/K is smaller than the slope ($m_{bf,2}$) of the boundary function for values < 6072.03 kW/K. Knowing that the intersection point of these two boundary functions was identified to be the critical point for the flexibility, we can assume that the identified point, θ^*, intersects with the boundary of the feasible region (i.e. $\psi(d,\theta^*) = 0$). We can further postulate that the slope of the boundary of the feasible region at this point is between $m_{bf,3}$ and $m_{bf,2}$. These two assumptions were verified by means of a Monte Carlo feasibility test which can be found in Appendix C. Note, that although the returned solution was a non-vertex solution of the hyperbox formed by the independent uncertain parameters, the solution is still a corner point of the hyperpolygon describing the uncertainty set.

6.2. Parameter dependencies — higher or lower deterministic flexibility index?

As illustrated by Langner et al. (2021) and the examples presented in Sections 4 and 5 of this paper, dependencies between (some) of the uncertain parameters can have a significant influence on the deterministic flexibility index if the respective dependency is captured well by means of a mathematical model.

Grossmann and Floudas (1987) also assumed that dependencies between (some of) the uncertain parameters, θ, should have a significant influence on the flexibility index calculated for a process. More precisely, they assumed that considering dependencies in the uncertain parameters results in a higher flexibility index, compared to the case of independent uncertain parameters. As previously mentioned (see Section 2), Grossmann and Floudas (1987) accounted for dependencies by means of single equation models, including them as constraints in the flexibility index problem (compare (A.2)). To illustrate this, they presented a HEN example where one uncertain temperature (*here: for simplicity T_B*) was expressed as a linear function of another uncertain temperature (*here: for simplicity T_A*) (compare Grossmann and Floudas (1987)). When including this linear equation, the authors reported that compared to the independent case (i.e. T_A and T_B vary independently by ±10 K), the flexibility index increased by 0.08824. However, it can easily be identified that due to the assumed linear equation, the (absolute) expected disturbance range of the dependent parameter (T_B) decreased from ±10 K to ±8 K, i.e. for the case that T_A is 10 K
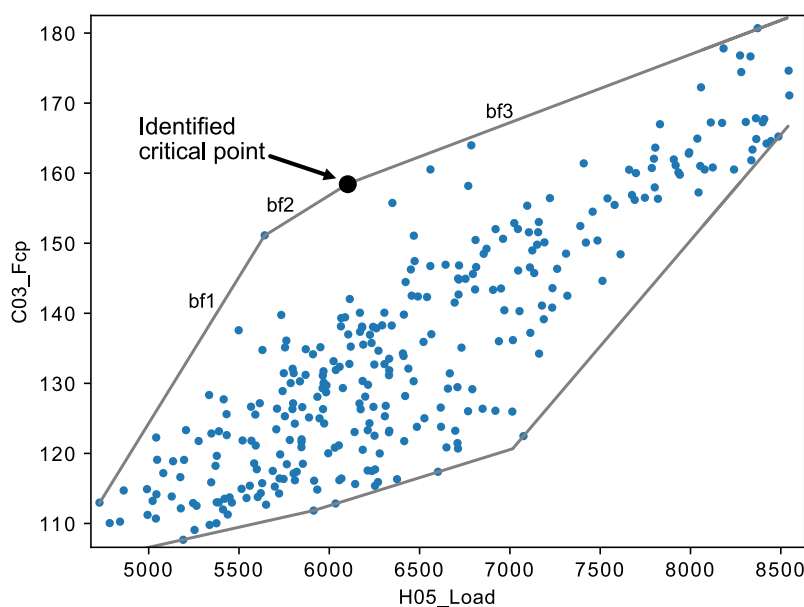
Fig. 14. Dependency between $C03\_Fcp$ and $H05\_Load$ captured by means of three lower and upper bounds and identified critical point.

higher than the nominal value, the linear equation returns a value for T_B which is not 10 K but only 8 K higher than the nominal value. Consequently, to allow a fair comparison, we decided to compare the result obtained for the linear equation to the case where T_A and T_B are assumed to vary independently but while for T_A variations of ± 10 K are assumed for T_B variations of ± 8 K are assumed, i.e. a modified hyperbox uncertainty set. We then identified that for the modified hyperbox uncertainty set, the flexibility index is also 0.08824 higher compared to the initial hyperbox uncertainty set, i.e. the case where both parameters vary independently by ± 10 K. Therefore, we can conclude that the presented example of Grossmann and Floudas (1987) illustrates, how changes in the maximum and minimum extreme values influence the result of the flexibility analysis. On the other hand, the same value for the flexibility index was obtained when:

1. including the linear equation to model a dependency between the uncertain parameters T_A and T_B,
2. assuming independent variation of T_A and T_B while adjusting the expected maximum and minimum extreme values of T_B (modified hyperbox uncertainty set).

Consequently, the linear equation did not help to exclude subsets of the hyperbox uncertainty set which limit the flexibility metric but in which no operating points are expected/observed. In fact, the consequence of including the linear equation is reduced expected extreme values of the uncertain parameters. However, the absolute extreme values of the uncertain parameters are independent of possible parameter dependencies since a parameter dependency (only) describes the relation between the extreme values, e.g., if they occur at the same time point(s). We can therefore also conclude that the example of Grossmann and Floudas (1987) does not explicitly show the impact on the results of the flexibility analysis when subsets of the hyperbox uncertainty set which limit the flexibility metric (e.g. flexibility index) are excluded since in these subsets no operating points are expected/observed. In this paper, we presented examples where we successfully excluded subsets of the hyperbox uncertainty set which are irrelevant for the actual operation but which limit the flexibility analysis (see Sections 4 and 5). The assurance for the successful exclusion of aforementioned subsets is given by the fact that the result of the flexibility analysis differs when considering dependencies between uncertain parameters compared to the hyperbox uncertainty set while considering the same absolute extreme values of the uncertain parameters.

As mentioned in the previous paragraph, Grossmann and Floudas (1987) assumed that considering dependencies in the uncertain parameters results in a higher flexibility index, compared to the case of independent uncertain parameters. This assumption is intuitive and was shared by other authors in the field, such as Rooney and Biegler (1999) and Pulsipher and Zavala (2018). However, in this paper, and in our previous conference contribution (Langner et al., 2021), we showed that when considering dependencies (by means of boundary functions) the flexibility index can be lower compared to the flexibility index calculated using a hyperbox uncertainty set:

- Langner et al. (2021): a case was identified for which $\delta(T_{boundary}) = 0.27$ is smaller than $\delta(T_{box}) = 0.36$,
- Considering only dependency 4 for the industrial case study (compare Section 5) yields a flexibility index of $\delta(T_{boundary}) = 0.55$ which is smaller than $\delta(T_{box}) = 0.64$

(Note that, in our conference contribution we also used a single equation model to express the above-mentioned dependency which resulted in a slightly higher flexibility index ($\delta(T_{corr,box}) = 0.385$) compared to the hyperbox uncertainty set ($\delta(T_{box}) = 0.36$) (Langner et al., 2021). However, it can be shown that changes in the maximum and minimum extreme values as a consequence of the used regression model are the reason for the difference in values, similar to the example of Grossmann and Floudas (1987).)

These observations contradict the intuitive assumption that considering dependencies in the uncertain parameters results in a higher flexibility index, compared to flexibility index based on the hyperbox uncertainty set. For the case study presented in Section 5, we illustrated (see Fig. 12) that the reduced flexibility index (compared to the hyperbox uncertainty set) is a consequence of the (additional) uncertainty span of the dependent parameter(s) implied by the boundary functions. This additional uncertainty span originates from the definition of the boundary functions which are not scaled by the scale parameter δ (which corresponds to the flexibility index) but are defined as functions of the independent parameters. This means, that for each parameter value of the independent parameters (e.g. also the mean value), a specific uncertainty span of the dependent parameters is enforced. Therefore, the obtained value of δ after solving a specific flexibility index problem expresses the limits of the feasible uncertainty span of the independent parameters, only. Note that a similar observation was

Table 9
Manual user input required during different steps of the presented methodology.

Step in methodology	Manual input
Identification of parameter dependencies	Metric to quantify the dependency between several uncertain parameters
	Threshold for dependency metric to be included in analysis
Classification of uncertain parameters	Selection of dependent and independent parameters (based on background knowledge or arbitrary)
Identification of boundary functions	Selection criterion to identify most influencing boundary functions
	number of lower boundary functions
	number of upper boundary functions
Data pre-/post- processing	filtering criteria to identify possible outlier(s)
Active Constraint Strategy (Calculation of flexibility metric)	Big-M parameter value (upper bound of inequality constraints)

also made for an uncertainty set based on single equation models (see Section 2.1).

For the dependent parameters applies that all variation is feasible which has been observed/is expected for this feasible uncertainty span of the independent parameters. In Section 5, we observed that due to the integration of the boundary functions of one identified dependency (dependency 4) a large feasible variation range of one (dependent) parameter was achieved by reducing the feasible variation range of the remaining uncertain parameters including the independent uncertain parameters and thereby reducing value of δ, i.e. the flexibility index. Consequently, we can conclude that in certain cases, the proposed usage of boundary functions to capture a dependency between two or more uncertain parameters can bias the result of the flexibility analysis to put more emphasis on the (previously) selected dependent parameters. This also means that in case of an arbitrary classification of the uncertain parameters into dependent and independent parameters, one may want to reconsider the classification. In this paper for the dependency in question (dependency 4 in Section 5), the classification was not performed arbitrary since we assessed it unreasonable to assume that the temperature of the external water source ($C03\_Tin$) is dependent on the heat load of stream $H4$. However, to illustrate the possible impact of a re-classification, we performed the theoretic experiment of regrouping $C03\_Tin$ and $H4\_Load$. When considering only dependency 4 but with reversed parameter dependency, we observed, for one or more upper and lower boundary function, a flexibility index of 0.61. This value is higher than the 0.55 which is the maximum flexibility index value achieved for the original parameter classification of dependency 4. On the other side, the aforementioned bias of the result of the flexibility analysis could not be fully avoided since the flexibility index for independent variation was calculated to 0.64. Note, that in certain situations this bias may be desirable, e.g., when the feasibility towards the uncertainty of selected parameters is prioritized. In case the bias should be avoided, the identified dependencies can be excluded (as done in Section 5) or need to be modeled using a different modeling approach. In this context, single equation models should, however, be avoided due to the risk of overestimating the flexibility.

6.3. Overview on manual user input

The methodology presented in this paper is well suited for automation. However, manual user input such as setting of parameter values is required during certain steps. In Table 9, the manual user input is grouped by the different steps in the presented methodology.

7. Conclusions and outlook

In this work, we presented the concept of boundary functions to allow for more accurate approximations of the uncertainty sets for deterministic flexibility analysis. We identified that the shape of an uncertainty set defined by boundary functions can be interpreted as a hyperpolygon which can be classified as a compact set similar to $T_{box}(\delta)$ or $T_{ellip}(\delta)$. Based on this observation, we derived a generic MI(N)LP formulation for the deterministic flexibility index based on uncertainty sets described by boundary functions which can be solved by means of available solution strategies.

We further proposed an algorithm based on the polygon convex hull to identify the most influencing boundary functions for a given maximum number of boundary functions, i.e. those boundary functions which capture the individual characteristics of the observed/expected uncertainty set. The assessment of each bound of the polygon convex hull (potential boundary functions) is performed by means of a selection criterion and we suggested two selection criteria in this paper.

The proposed algorithm was applied in several numerical examples, and we demonstrated its applicability also to multi-dimensional cases, i.e. when dependencies between more than two uncertain parameters are identified. Two numerical examples and an industrial case study were presented.

By means of a theoretical example, we compared how the results of the deterministic flexibility index can differ in the presence of parameter dependencies, when choosing a hyperbox uncertainty set compared to an uncertainty set based on a linear regression model or boundary functions. We were able to confirm previously reported observations that the flexibility index based on a hyperbox uncertainty set can underestimate feasible parameter variation if these parameters show correlating trends. Additionally, we illustrated that the incorporation of linear regression models significantly underestimates the set of expected operating points. The consequence of this observation can be that the obtained flexibility index overestimates the interval within which exclusively feasible operating points are observed. Such an overestimation can have severe consequences for the practical application of flexibility analysis, as outlined in Section 2.1. In comparison to the hyperbox uncertainty set and the uncertainty set based on a linear regression model, the uncertainty set defined by following our proposed approach showed the best resemblance with the set of observed operating points (compare Fig. 6). We can thus conclude that the obtained flexibility index is a more accurate quantification of a system's flexibility when following our suggested approach compared to the results obtained with a hyperbox uncertainty set or an uncertainty set based on linear regression functions.

To avoid confusion regarding the comparability of our proposed approach with the incorporation of ellipsoidal uncertainty sets in the deterministic framework of the flexibility index (as suggested by Pulsipher and Zavala (2018)), we utilized our theoretical example to further illustrate the differences of these approaches. We showed that ellipsoidal uncertainty sets can be effective for approximating the stochastic flexibility index which provides information on the probability for feasible operation. On the other hand, the deterministic flexibility index returns the maximum feasible disturbance for each uncertain parameter from a nominal/mean value. Consequently, the two approaches cannot be substituted but provide complementary information.

To illustrate the applicability of the proposed concepts also in more complex situations characterized by multi-dimensional dependencies, among others, we investigated a HEN example and an industrial case study. We showed that dependencies in (some) of the uncertain parameters can have a significant influence on the deterministic flexibility index if the respective dependency is accurately captured by a mathematical model. However, we also concluded that it is not possible to make general statements regarding the effect of dependencies in the uncertain parameters on the flexibility index and that such observations are dependent on the chosen modeling approach.

Finally, we concluded that, compared to previously published approaches to capture dependencies in the uncertain parameters when conducting deterministic flexibility analysis, the risk for over- or underestimation of the feasible variation span is lower when following the approach presented in this paper. More precisely, overestimation of the feasible variation range can be avoided using the proposed approach since all expected operating conditions are included in the analysis, i.e. the boundary functions are defined based on the convex hull. At the same time, it is worth mentioning that underestimation is inherent in any approach where geometrical shapes are used to approximate the actual uncertainty space. However, boundary functions allow for a more precise modeling of the uncertainty set compared to regular geometric shapes. More specifically, boundary functions allow for excluding subsets which limit the flexibility metric but contain no operating points (observed or expected). Consequently, with the proposed approach, the level of underestimation is likely to be lower since a better approximation of the actual uncertainty space is possible due to the increased degrees of freedom. This assumption is reflected in the numerical results presented in the paper since in all investigated examples (compare Sections 5, 4.1 and 4.2), a higher deterministic flexibility index was obtained when modeling the uncertainty set following our proposed approach, compared to the hyperbox approach. In this context, it can be assumed that the resemblance of the modeled uncertainty set with the expected uncertainty set should be satisfactory as long as a tight convex hull representation can be found.

The approach suggested in this paper for approximating the expected uncertainty space by means of boundary functions which are derived using the convex hull may also be utilized in other concepts related to flexibility analysis. The previously mentioned volumetric flexibility index has been defined as the percentage of the expected hyperbox uncertainty set that can be feasibly handled. However, if the distribution of uncertainty shows correlating trends, modeling the expected uncertainty space by means of a hyperbox may lead to small values of the volumetric flexibility index. According to Lai and Hui (2008), this implies that the investigated system cannot operate feasibly at all expected realizations of the uncertain parameters. On the other hand, if the model of the expected uncertainty space significantly overestimates the space in which operating conditions are expected (which can be a consequence of the hyperbox approach, as shown in this work), the result of the volumetric flexibility index for a hyperbox uncertainty set may be inaccurate. Additionally, a model of the expected uncertainty space is needed for the identification of critical parameter values which, according to Halemane and Grossmann (1983), can be utilized in a design under uncertainty problem to ensure that the obtained design allows for feasible operation at all realizations of the uncertain parameters. In this context, a more exact representation of the expected uncertainty space should avoid unnecessary overdesign of equipment which can result from designing process equipment for combinations of uncertain parameter values which are not expected.

CRediT authorship contribution statement

Christian Langner: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Elin Svensson:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Stavros Papadokonstantakis:** Methodology, Writing – review & editing, Supervision. **Simon Harvey:** Resources, Writing – review & editing, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request

Appendix A. Theory on the flexibility index

The problem formulation of the flexibility index for a hyperbox uncertainty set is given in (A.1) as derived by Swaney and Grossmann (1985). Note that in (A.1), the uncertainty set is scaled with the scalar δ ($T_{box}(\delta)$). In (A.1), Swaney and Grossmann (1985) included the feasibility constraint (compare (1)) to find the largest value for the scale parameter of the uncertainty set (δ) so that all constraint functions are feasible ($f_j(d, z, \theta) \leq 0$) while accounting for recursive actions. Since the uncertainty set is scaled, $T_{box}(\delta)$ can be interpreted as the largest scaled hyperbox which can be inscribed in the feasible region.

$$F = \max \delta$$
$$s.t. \ \max_{\theta \in T} \min_{z} \max_{j \in J} f_j(d, z, \theta) \leq 0$$
$$T_{box}(\delta) = \left\{ \theta_i \mid \theta_{i,N} - \delta \Delta \theta_i^- \leq \theta_i \leq \theta_{i,N} + \delta \Delta \theta_i^+ \right\} \forall \ \theta_i \in \theta \tag{A.1}$$
$$\delta \geq 0$$

Grossmann and Floudas (1987) proposed a reformulation of (A.1) to solve the feasibility constraint, (1), explicitly without relying on the assumption of critical points corresponding to vertices of the hyperbox uncertainty set. They reformulated the feasibility constraint in (A.1) to explicitly search for the solution on the boundary of the feasible region ($\psi(d, \theta) = 0$) yielding the bi-level optimization problem given in (A.2).

$$F = \min \delta$$
$$s.t. \ \psi(d, \theta) = 0$$
$$\psi(d, \theta) = \min_{z,u} u$$
$$s.t. \ f_j(d, z, \theta) \leq u, \ j \in J \tag{A.2}$$
$$T_{box}(\delta) = \left\{ \theta_i \mid \theta_{i,N} - \delta \Delta \theta_i^- \leq \theta_i \leq \theta_{i,N} + \delta \Delta \theta_i^+ \right\} \forall \ \theta_i \in \theta$$
$$\delta \geq 0$$

The advantage of (A.2) is that the feasibility constraint, (1), can be solved explicitly by replacing the lower level optimization problem ($\psi(d, \theta) = \min u$) by its Karush–Kuhn–Tucker (KKT) optimality conditions. Grossmann and Floudas (1987) further identified that if the gradients $\partial f_j / \partial z$ are linearly independent, there will be $n_z + 1$ active inequality constraints at the solution. Therefore, the authors simplified the complementarity conditions with mixed-integer constraints. Later, Biegler et al. (1997) found out that if linear independence of the gradients $\partial f_j / \partial z$ cannot be guaranteed, the number of active inequality constraints can be relaxed. The mixed-integer representation of the KKT-conditions and the relaxed constraint on the number of active inequality constraints is given from (A.3a) to (A.3g). Note that the $Big - M$-parameter in (A.3d) represents an upper bound to the slack variables of the inequality constraints. Further, note that the resulting mixed-integer linear/non-linear program (MI(N)LP) does not require strict convexity of the constraint functions $f_j(x, z, \theta) \leq 0, \ j \in J$ and thereby does not rely on the assumption of critical points corresponding to vertices. To also be able to solve non-convex system formulations in general, the active constraint strategy was later extended by Floudas et al. (2001) to a global solution algorithm.

$$\sum_{j \in J} \lambda_j = 1 \tag{A.3a}$$
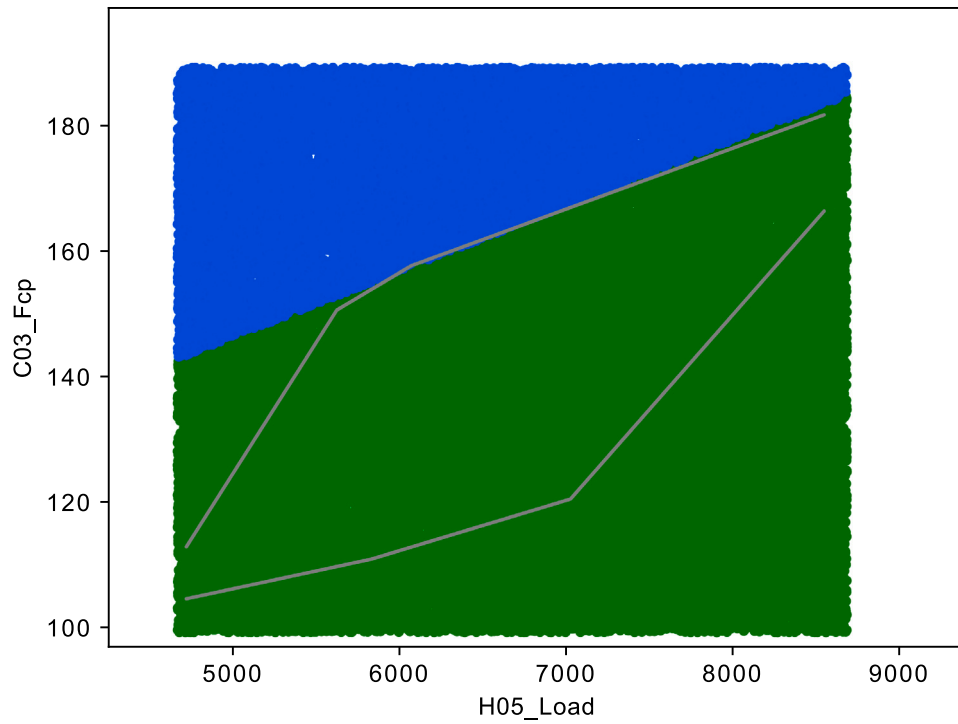
(a) Overview of all randomly chosen operating points.
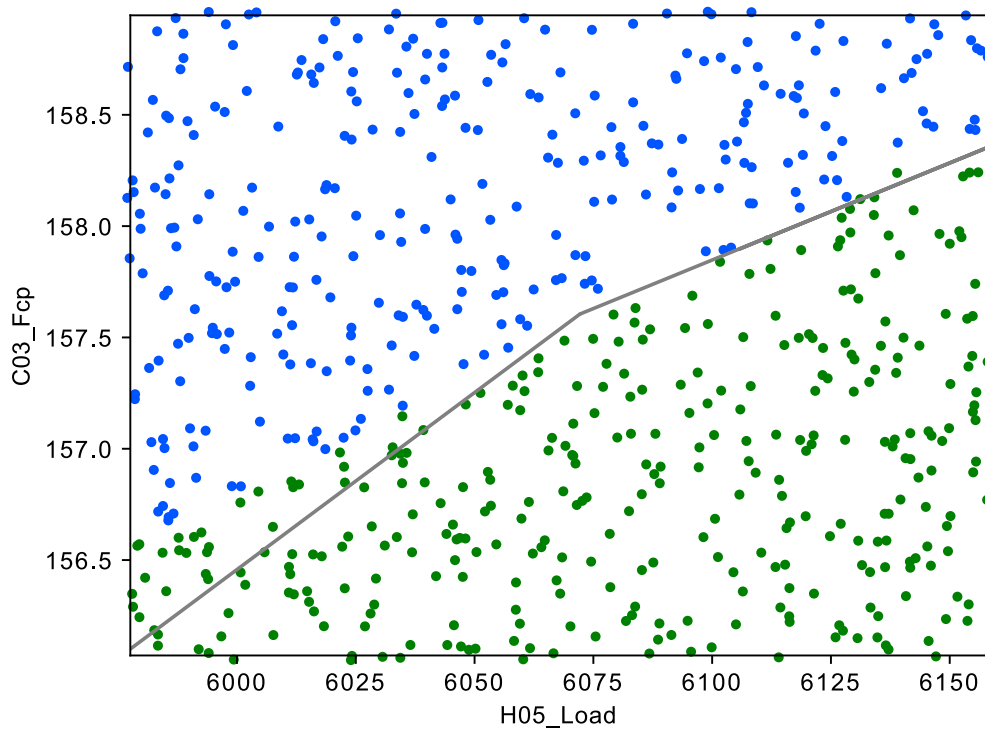
(b) Detailed zoom of operating points around the identified critical point.

Fig. C.1. Results of the Monte Carlo Feasibility Test color-coded with respect to the feasibility (blue: infeasible, green: feasible).

$$f_j(d, z, \theta) + s_j = 0, \ j \in J \tag{A.3b}$$

$$\sum_{j \in J} \lambda_j \frac{\partial f_j}{\partial z} = 0 \tag{A.3c}$$

$$s_j - M(1 - y_j) \le 0, \ j \in J \tag{A.3d}$$

$$\lambda_j - y_j \le 0, \ j \in J \tag{A.3e}$$

$$\sum_{j \in J} y_j \le n_z + 1 \tag{A.3f}$$

$$s_j \ge 0, \lambda_j \ge 0, y_j \in 0, 1 \ j \in J \tag{A.3g}$$

Appendix B. Proof of theorem

In the following, a generalized version of Theorem 3 by Pulsipher and Zavala (2018) and the corresponding proof are provided. Defining the feasible region of a physical system as $\Theta = \{\theta : \psi(d, \theta) \le 0\}$ and the boundary of this feasible region as $\partial \Theta = \{\theta : \psi(d, \theta) = 0\}$.

Theorem 1. *Given a solution pair (δ^*, θ^*) for a scalable compact uncertainty set $T(\delta)$, this solution pair satisfies the following properties: (i) The uncertainty set is contained in the feasible set ($T(\delta^*) \subseteq \Theta$), (ii) the critical parameter θ^* lies on the boundary of the uncertainty set ($\theta^* \in \partial T(\delta^*)$), and (iii) the critical parameter θ^* lies on the intersection of the boundaries of the feasible and uncertainty sets ($\theta^* \in \partial \Theta \cup \partial T(\delta^*)$).*

Proof of Theorem 1. To prove (i) assume there exists $\tilde{\theta} \in T(\delta^*)$ but $\tilde{\theta} \notin \Theta$. Since $\tilde{\theta}$ is infeasible, we have that $\psi(d, \tilde{\theta}) > 0$ which implies that $\psi(d, \tilde{\theta}) > \psi(d, \theta^*)$ because $\psi(d, \theta^*) = 0$. This is a contradiction because $\theta^* \in \underset{\theta \in T(\delta^*)}{\arg\max} \ \psi(d, \theta^*)$ and thus $\psi(d, \theta^*) = 0$ is the maximum possible value of $\psi(d, \theta)$ in $T(\delta^*)$. To prove (ii) we note that $\theta^* \in \partial \Theta$ holds and, therefore, if $\theta^* \notin \partial T(\theta^*)$ then there exists $\tilde{\theta} \in T(\delta^*)$ with $\tilde{\theta} \notin \Theta$. The result then follows from the argument used to prove (i). The proof of (iii) follows trivially from the observation that δ^* lies on the boundary of both Θ and $T(\delta^*)$.

Appendix C. Monte Carlo feasibility test

A Monte Carlo feasibility test was performed to verify that at the identified critical point, θ^*, the boundary functions (describing dependency 1^*) intersect with the boundary of the feasible region (i.e. $\psi(d, \theta^*) = 0$). In this context, we investigated the feasibility of 50000 randomly chosen operating points around the identified critical point. More precisely, all uncertain parameters were fixed at the identified critical values except for $C03\_Fcp$ and $H05\_Load$ which were varied randomly. The results are shown in Fig. C.1 (feasible points are colored green, infeasible points are marked blue). We could further verify that the slope of the boundary of the feasible region at identified critical point is between $m_{bf,3}$ and $m_{bf,2}$.

References

Barber, C.B., Dobkin, D.P., Huhdanpaa, H., 1996. The quickhull algorithm for convex hulls. ACM Trans. Math. Software 22 (4), 469–483. http://dx.doi.org/10.1145/235815.235821.

Biegler, L., Grossmann, I.E., Westerberg, A.W., 1997. Systematic Methods of Chemical Process Design, first ed. Pearson Education.

Boslaugh, S., Watters, P.A., 2008. Statistics in a Nutshell - A Desktop Quick Reference. O'Reilly, pp. I–XXI, 1–452.

Bykat, A., 1978. Convex hull of a finite set of points in two dimensions. Inform. Process. Lett. 7 (6), 296–298. http://dx.doi.org/10.1016/0020-0190(78)90021-2.

Floudas, C.A., Gümüş, Z.H., Ierapetritou, M.G., 2001. Global optimization in design under uncertainty: Feasibility test and flexibility index problems. Ind. Eng. Chem. Res. 40 (20), 4267–4282. http://dx.doi.org/10.1021/ie001014g.

Grossmann, I.E., Calfa, B.A., Garcia-Herreros, P., 2014. Evolution of concepts and models for quantifying resiliency and flexibility of chemical processes. Comput. Chem. Eng. 70, 22–34. http://dx.doi.org/10.1016/j.compchemeng.2013.12.013.

Grossmann, I.E., Floudas, C.A., 1987. Active constraint strategy for flexibility analysis in chemical processes. Comput. Chem. Eng. 11 (6), 675–693. http://dx.doi.org/10.1016/0098-1354(87)87011-4.

Grossmann, I.E., Sargent, R.W., 1978. Optimum design of chemical plants with uncertain parameters. AIChE J. 24 (6), 1021–1028. http://dx.doi.org/10.1002/aic.690240612.

Halemane, K.P., Grossmann, I.E., 1983. Optimal process design under uncertainty. AIChE J. 29 (3), 425–433. http://dx.doi.org/10.1002/aic.690290312.

Lai, S.M., Hui, C.W., 2008. Process flexibility for multivariable systems. Ind. Eng. Chem. Res. 47 (12), 4170–4183. http://dx.doi.org/10.1021/ie070183z.

Langner, C., Svensson, E., Harvey, S., 2020. A framework for flexible and cost-efficient retrofit measures of heat exchanger networks. Energies 13 (6), 1–24. http://dx.doi.org/10.3390/en13061472.

Langner, C., Svensson, E., Harvey, S., 2021. Flexibility analysis of chemical processes considering dependencies between uncertain parameters. Comput. Aided Chem. Eng. 50, 1105–1110. http://dx.doi.org/10.1016/B978-0-323-88506-5.50170-4.

Li, Z., Ding, R., Floudas, C.A., 2011. A comparative theoretical and computational study on robust counterpart optimization: I. Robust linear optimization and robust mixed integer linear optimization. Ind. Eng. Chem. Res. 50 (18), 10567–10603. http://dx.doi.org/10.1021/ie200150p.

Li, J., Du, J., Zhao, Z., Yao, P., 2015. Efficient method for flexibility analysis of large-scale nonconvex heat exchanger networks. Ind. Eng. Chem. Res. 54 (43), 10757–10767. http://dx.doi.org/10.1021/acs.iecr.5b00237.

Liu, F.T., Ting, K.M., Zhou, Z.H., 2008. Isolation forest. In: Proceedings - IEEE International Conference on Data Mining. ICDM, IEEE, pp. 413–422. http://dx.doi.org/10.1109/ICDM.2008.17.

Ochoa, M.P., Grossmann, I.E., 2020. Novel MINLP formulations for flexibility analysis for measured and unmeasured uncertain parameters. Comput. Chem. Eng. 135, http://dx.doi.org/10.1016/j.compchemeng.2020.106727.

Ostrovsky, G.M., Datskov, I.V., Achenie, L.E., Volin, Y.M., 2003. Process uncertainty: Case of insufficient process data at the operation stage. AIChE J. 49 (5), 1216–1232. http://dx.doi.org/10.1002/aic.690490513.

Persson, J., Berntsson, T., 2009. Influence of seasonal variations on energy-saving opportunities in a pulp mill. Energy 34 (10), 1705–1714. http://dx.doi.org/10.1016/J.ENERGY.2009.07.023.

Pistikopoulos, E.N., Mazzuchi, T.A., 1990. A novel flexibility analysis approach for processes with stochastic parameters. Comput. Chem. Eng. 14 (9), 991–1000. http://dx.doi.org/10.1016/0098-1354(90)87055-T.

Pulsipher, J.L., Zavala, V.M., 2018. A mixed-integer conic programming formulation for computing the flexibility index under multivariate Gaussian uncertainty. Comput. Chem. Eng. 119, 302–308. http://dx.doi.org/10.1016/j.compchemeng.2018.09.005.

Raspanti, C., Bandoni, J., Biegler, L., 2000. New strategies for flexibility analysis and design under uncertainty. Comput. Chem. Eng. 24 (9–10), 2193–2209. http://dx.doi.org/10.1016/S0098-1354(00)00591-3.

Rooney, W.C., Biegler, L.T., 1999. Incorporating joint confidence regions into design under uncertainty. Comput. Chem. Eng. 23 (10), 1563–1575. http://dx.doi.org/10.1016/S0098-1354(99)00311-7.

Rooney, W.C., Biegler, L.T., 2003. Optimal process design with model parameter uncertainty and process variability. AIChE J. 49 (2), 438–449. http://dx.doi.org/10.1002/aic.690490214.

Saboo, A.K., Morari, M., Woodcock, D.C., 1985. Design of resilient processing plants-VIII. A resilience index for heat exchanger networks. Chem. Eng. Sci. 40 (8), 1553–1565. http://dx.doi.org/10.1016/0009-2509(85)80097-X.

Sahinidis, N.V., 2017. BARON 21.1.13: Global optimization of mixed-integer nonlinear programs, User's Manual.

Straub, D.A., Grossmann, I.E., 1990. Integrated stochastic metric of flexibility for systems with discrete state and continuous parameter uncertainties. Comput. Chem. Eng. 14 (9), 967–985. http://dx.doi.org/10.1016/0098-1354(90)87053-R.

Straub, D.A., Grossmann, I.E., 1993. Design optimization of stochastic flexibility. Comput. Chem. Eng. 17 (4), 339–354. http://dx.doi.org/10.1016/0098-1354(93)80025-I.

Swaney, R.E., Grossmann, I.E., 1985. An index for operational flexibility in chemical process design. Part I: Formulation and theory. AIChE J. 31 (4), 621–630. http://dx.doi.org/10.1002/aic.690310412.

Zhang, Q., Grossmann, I.E., Lima, R.M., 2016. On the relation between flexibility analysis and robust optimization for linear systems. AIChE J. 62 (9), 3109–3123. http://dx.doi.org/10.1002/aic.15221.

Zheng, C., Zhao, F., Zhu, L., Chen, X., 2021. Analytical solution of volumetric flexibility through symbolic computation. Chem. Eng. Sci. 239, 116643. http://dx.doi.org/10.1016/J.CES.2021.116643.