Machine Learning-Driven Optimisation of Operational Spaces for Uncertainty Management in Process Industries ⋆

Sam Kay, Mengjia Zhu, Oliver Pennington, Dongda Zhang *

* *Department of Chemical Engineering, University of Manchester, Manchester, M13 9PL, UK (e-mail: dongda.zhang@manchester.ac.uk).*

Abstract: Process optimisation and quality control are crucial in process industries to minimise waste and enhance economics. However, common uncertainties ranging from feedstock variability to human error can cause significant deviations in product quality leading to batch discards. This study introduces a novel framework combining machine learning with optimisation strategies to identify optimal operational spaces under uncertainty. Using a process model, the framework screens a broad operational space, isolating promising sub-regions and control trajectories. Machine learning techniques are used to cluster these sub-regions by displayed control patterns, and a dynamic optimisation framework identifies the maximum operable design space, ensuring constraints are met under uncertainty. Two case studies, involving a fermentation process and a formulation manufacturing process, were conducted to demonstrate the high efficiency of the proposed framework and to showcase its strong potential for industrial applications.

Keywords: Artificial intelligence, Batch processing, Dynamic optimisation, Design space identification, Process control, Uncertainty estimation.

1. INTRODUCTION

Product quality control is essential in industries such as pharmaceuticals, specialty chemicals, and formulation processes, where one aims to strictly regulate process conditions for a myriad of reasons, ranging from ensuring the safety and efficacy of process operations, to process optimisation for improved economic performance and sustainability (Hicks et al., 2021). In recent years, there has been an increase in pressure for industries to attain tighter control over their product qualities; this is primarily driven by global and local governmental initiatives to reduce environmental emissions, alongside increasing operation costs. Due to this, it is highly desirable to eliminate any sources of wasted product which is often in considerable volumes, particularly in the pharmaceutical sectors (Amrih and Damayanti, 2022).

In order to address such problems and maintain an advantage in a competitive market, it is necessary that new formulations and greener processes are developed (Hill, 2007). Due to this, process optimisation and control becomes a key focal point, especially in addressing inefficiencies, reducing waste, and enhancing process flexibility to adapt to manufacturing demands and constraints. Many existing processes rely on traditional set-point control methods which, whilst effective for maintaining specific process conditions known to provide reliable operation, lack the adaptability needed to respond effectively to uncertainties and dynamic process requirements which may result in frequent batch rejections. Furthermore, the identified optimal set-point strategy is, by definition, rigid in its control and thus could require significant energy consumption.

Naturally, uncertainties pose a great challenge which remains to be addressed in the implementation of the current safety, economic and environmental directives, since the quality of process control solutions are heavily influenced by them. Uncertainties arise within standard processing for a multitude of reasons, and although possible to mitigate certain sources, it is not possible to eliminate all uncertainties so, it is good practice to identify and account for these when aiming to implement robust control solutions (Geletu et al., 2013). Such sources of uncertainty are inherent to operations, and can be exhibited through variation of one's ability to achieve a control action's set-point, changes in feed concentrations or compositions, variability in measurements, and even operator actions.

A promising alternative to set-point control is operational space control, which provides greater flexibility by defining an operational space rather being bound to a single set-point. This approach offers distinct advantages over set-point control, as it enables processes to adapt dynamically, within certain process constraints, to account for variability and uncertainties.

1.1 Optimisation Under Uncertainty

Robust Optimisation Robust optimisation tackles uncertainty in a deterministic manner, focusing on ensuring that constraints remain satisfied even in the worst-case scenarios corresponding to the most adverse operating conditions and uncertainties. This is achieved by incorporating uncertainty sets, which define the ranges of variability for both the objective function and constraints

⋆ This project is partially supported by the EPSRC Grant of *Artificial Intelligence Enabling Future Optimal Flexible Biogas Production for Net-Zero* (EP/Y005600/1)

(García and Peña, 2018). The goal is to develop an optimal control strategy that performs adequately under all possible conditions within these uncertainty sets, with a particular emphasis on the worst-case outcome. While effective in maintaining feasibility, this often results in overly conservative solutions, as the method prioritises reliability over performance (Bomze and Gabl, 2023). This is usually considered in the event that any probability of constraint violation is unacceptable (Gabrel et al., 2014).

Stochastic Optimisation An alternative method used for optimisation under uncertainty is stochastic optimisation. It is a generic branch of optimisation which refers to the use of random variables to represent uncertain parameters (Zheng et al., 2014). The probability distributions can be tailored to accurately account for different uncertainty types, and their expected characteristics (*e.g.*, Gaussian) such that real operations are reflected. This can be used to provide information about the expected performance of a solution during an optimisation procedure. In practice, it is intractable to optimise over the range of a probability distribution. Hence, it is common that Monte Carlo and scenario-based methods are implemented to reformulate the problem into a deterministic one (Zheng et al., 2014). Naturally, this usually revolves around unique sampling methodologies. When dealing with optimisation under uncertainty, stochastic optimisation is often chosen over robust optimisation because it offers more flexibility, better application to non-convex optimisation of complex real non-linear systems, and usually more practical solutions for industrial use.

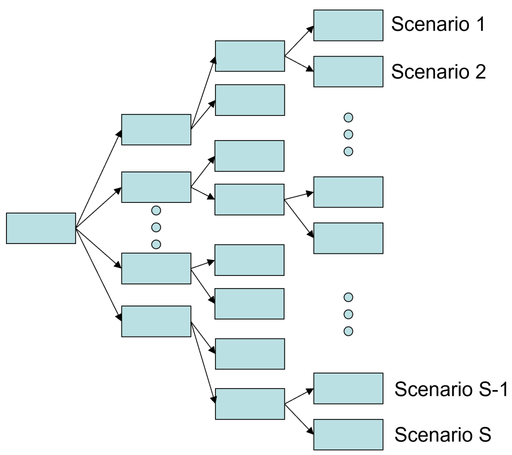

Fig. 1. Diagram of an example scenario tree.

Scenario tree optimisation, a form of stochastic optimisation, can be employed to transform complex probabilistic problems into more manageable deterministic ones, thus avoiding the computational challenges of directly solving probabilistic systems. In this approach, uncertainties are propagated across an entire process, with multiple different realisations of these uncertainties considered (Silvente et al., 2019; Kammammettu and Li, 2023). Each full trajectory of the process, defined by a unique combination of uncertainty realisations, is referred to as a scenario. One can represent all considered uncertain trajectories as a single system named a scenario tree where uncertainty unfolds stagewise, an example of this is shown in Fig. 1. This method provides a structured way to account for a diverse range of time-varying uncertainties, enabling better decision-making across a set of potential future outcomes.

In this paper, we propose a novel scenario tree optimisation framework that combines the concepts of flexible and optimal operational spaces to simultaneously enhance process adaptability, robustify the process to uncertainties and achieve operational optimality, therefore, representing a novel advancement in process quality control.

1.2 Process Description

We assume a mathematical model is available to describe the process dynamics:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(t, \boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{\theta}, \boldsymbol{\xi}), \quad (1)$$

where, $\boldsymbol{x}(t_j) \in \mathbb{R}^n$ is the vector of state variables, $t_j \in [0, t_f]$ is a time instance of the operational time horizon, and $\dot{\boldsymbol{x}}(t_j)$ is the corresponding time derivative. $\boldsymbol{u}(t_j) \in \mathbb{R}^m$ is the vector of control actions. $\boldsymbol{\theta} \in \mathbb{R}^p$ is the vector of fixed parameters, and $\boldsymbol{\xi} \in \mathbb{R}^r$ is the vector of uncertain parameters. In this work, each uncertainty sources, ξ_i, for $i = 1, \ldots, r$, is approximated as a random variable, and assigned a unique Gaussian distribution, *i.e.*, $\xi_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, with mean value, μ_i, equal to its nominal value, and standard deviation, σ_i, equal to the deviation one might expect during operations. Both, μ_i and σ_i, are assumed to be known. The goal is to identify a practical, optimal, and flexible operational space for $\boldsymbol{u}(t)$, ensuring that product quality is consistently maintained despite the uncertainties introduced by $\boldsymbol{\xi}$.

2. METHODOLOGY

The proposed methodology is described in Fig. 2. In general, the approach can be split into 4 steps, which are detailed in the following subsections.

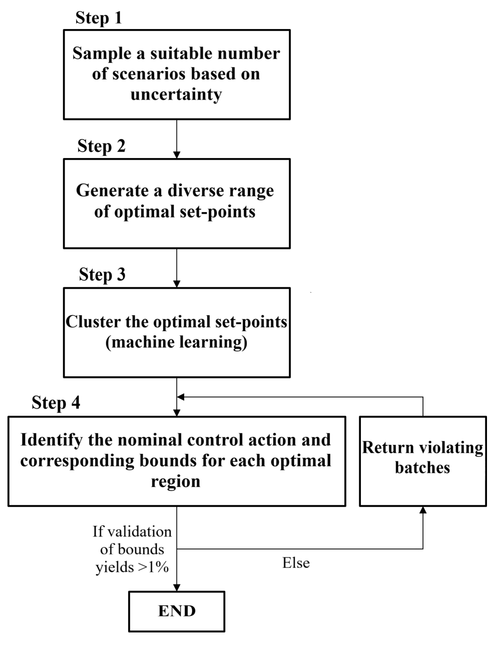

Fig. 2. Proposed methodology for optimal space identification.

2.1 Step 1: Uncertainty Sampling

We adopt the scenario tree analysis, as described in Section 1.1, to represent the possible range of uncertainties one may experience in processing. The entire scenario tree can be realised by sampling each uncertainty source $\boldsymbol{\xi}_i$ from their corresponding distributions.

It is important to note that a good representation of uncertainty is necessary to achieve a robust control action. Hence, it is recommended that a large enough number of uncertainty samples is taken from each source of uncertainty. However, this must be considered in conjunction with the increased computational cost one faces when increasing the size of the scenario tree. In practice, it is common to find a trade-off between the two. Once the scenario tree is realised, all optimisation strategies will be optimised over the entire scenario tree such that any defined constraints are satisfied for any of the considered uncertainty realisations.

2.2 Step 2: Filling out the Optimal Region

Since different pathways may exist to achieve control goals, in Step 2, we focus on narrowing the broad search space to identify specific operating spaces that are likely to contain the optimal control actions. These selected spaces will be further reduced to identify the operational regions in Step 4. The optimal operational space is characterised by control trajectories that meet a predefined standard of process optimality while also adhering to process constraints under any considered uncertainties realised in the scenario tree. Specifically, we want to minimise some process cost, C, while maintaining tight control over end product quality, y_{t_f}, close to the desired set point, y_{SP}. In addition, we aim to locate different operational spaces, if they exist, to characterise the process.

We discretise the operation time horizon into N intervals, with the control actions, $\boldsymbol{u}(t)$, assigned as piecewise constants within each time interval, *i.e.*, $\boldsymbol{u}(t) = \boldsymbol{u}_k$, for $t \in [t_k, t_{k+1})$, $\forall k \in [0, N-1]$, where $t_k = k\Delta t$ and $\Delta t = \frac{t_f}{N}$. We assume that the system states, $\boldsymbol{x}_k = \boldsymbol{x}(t_k)$, are fully observable, and the initial states, $\boldsymbol{x}_0$, are known. Often, the process is subject to constraints, which are denoted as $\boldsymbol{g}(\boldsymbol{x}_k)$. Additionally, the control actions are bounded by physical limitations, denoted as $\underline{\boldsymbol{u}}_k$ and $\overline{\boldsymbol{u}}_k$ for $k = 0, \ldots, N-1$. The problem formulation is summarised as follows:

$$\begin{aligned} \min_{\substack{\boldsymbol{u}_k \\ k=0,\ldots,N-1}} \quad & \lambda_1 \cdot \mathrm{obj}_1 + \lambda_2 \cdot C \\ \text{s.t.} \quad & \mathrm{obj}_1 = \sum_{s=1}^{S} (y_{t_f,s} - y_{\mathrm{SP}})^2 \\ & \boldsymbol{x}_{k+1} = \boldsymbol{f}(\Delta t, \boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}, \boldsymbol{\xi}) \\ & \boldsymbol{x}_{\mathbf{0}} = \boldsymbol{x}(0) \\ & \boldsymbol{g}(\boldsymbol{x}_k) \leq 0 \\ & \underline{\boldsymbol{u}}_k \leq \boldsymbol{u}_k \leq \overline{\boldsymbol{u}}_k, \end{aligned} \tag{2}$$

where λ_1 and λ_2 represent the weighting parameters for each objective; S is the number of uncertain scenarios considered in the scenario tree; and $\boldsymbol{f}(\cdot)$ is the selected numerical integrator (*e.g.*, implicit/explicit Euler, Runge-Kutta, *etc.*).

Since we want to identify a diverse set of plausible control trajectories to find different operational spaces (if multiple exist), once an initial optimisation of Problem (2) is complete, we can identify the next control trajectory by re-optimising Problem (2), with a penalty function, p, added to the original objective function (*i.e.* $\lambda_1 \cdot \mathrm{obj}_1 + \lambda_2 \cdot C + p$), to maximise the differences between the existing control trajectories and the ones to be identified. The form of such a penalty could be

$$p = \lambda_3 \cdot \sum_{d=1}^{D_{\mathrm{curr}}} \sum_{j=1}^{m} \sum_{k=0}^{N-1} \frac{(u_{j,k} - u^*_{j,k,d})}{\Delta \overline{u}_{j,k}}, \tag{3}$$

where, λ_3 is the weighting parameter used to dictate the strength of the penalty function, for which larger values aim to identify more varied optimums and so, a larger optimal region. D_{curr} is the number of identified control trajectories thus far. $u^*_{j,k,d}$ represents the discretised control actions from an identified trajectory, and $u_{j,k}$ is the corresponding control action within the current optimisation. $\Delta \overline{u}_{j,k}$ is the maximum range of control action $u_{j,k}$ among identified ones, which is used to normalise the values for each control variable which may otherwise be of different orders of magnitude.

Using this methodology, with a large enough number of optimal solutions, $D_{\max}$, it is possible to approximate regions of optimal solutions (control strategies) which satisfy all the feasibility constraints under uncertainty.

2.3 Step 3: Clustering Optimal Regions

Following Step 2, different operational spaces may be identified. Clustering algorithms are applied to identify and characterise these spaces. As clustering is an unsupervised learning approach, validating the results across different clustering algorithms and analysing cluster characteristics is essential to assess distinctions in control strategy behaviours.

In this work, multiple clustering algorithms: k-means (Na et al., 2010), DBSCAN (Khan et al., 2014), and spectral clustering (Jia et al., 2014), were employed and compared to ensure consistency in the recommended number of clusters.

2.4 Step 4: Nominal Control and Bound Estimation Strategy

In Steps 2 and 3, we have approximations for a number of distinct optimal spaces (clusters). We can now determine the optimal control action for each cluster, as well as their corresponding operational regions. Specifically, the operational region of an optimal control trajectory is defined by upper and lower bounds for each control action, such that, if we operate within, the process is feasible under considered uncertainties. The optimal set-points are established using the dynamic optimisation formulation described in (2), where the search space is now restricted to that of the chosen optimal region to ensure that the solution does indeed fall within the correct cluster, *i.e.*, $\underline{\boldsymbol{u}}_k$ and $\overline{\boldsymbol{u}}_k$ in (2) are updated to reflect the cluster characteristics.

Once the nominal (optimal) set-points have been found, we propose a two step algorithm to identify the upper and lower bounds independently by maximising the distance between the two whilst satisfying process constraints. Eqn. (4) shows the optimisation strategy to find the lower bound:

$$\begin{aligned}
\max_{\substack{\boldsymbol{u}_k^{\text{lb}}\\ k=0,\dots,N-1}} \quad & \min_{j=1}^{m} \min_{k=0}^{N-1} \ \underline{w}_{j,k} \cdot (u_{j,k}^{\text{nominal}} - u_{j,k}^{\text{lb}}) \\
\text{s.t.} \quad & \underline{w}_{j,k} = \frac{1}{u_{j,k}^{\text{nominal}} - \underline{u}_{j,k}} \\
& \boldsymbol{x}_{k+1} = \boldsymbol{f}(\Delta t, \boldsymbol{x}_k, \boldsymbol{u}_k^{\text{lb}}, \boldsymbol{\theta}, \boldsymbol{\xi}) \\
& \boldsymbol{x_0} = \boldsymbol{x}(0) \\
& \boldsymbol{g}(\boldsymbol{x}_k) \leq 0 \\
& \underline{\boldsymbol{u}}_k \leq \boldsymbol{u}_k^{\text{lb}} \leq \boldsymbol{u}_k^{\text{nominal}},
\end{aligned} \tag{4}$$

where, (4) aims to maximise the distance between the control actions at the lower bound, $u_{j,k}^{\text{lb}}$, and the nominal control actions, $u_{j,k}^{\text{nominal}}$, across each control action, j, for each time-step, k. $\underline{w}_{j,k}$ is used to normalise the magnitude of the objective contribution for different control actions. In order to ensure that the area of the bounds is spread as evenly as possible across the entire control trajectory for each variable, the objective function is designed such that the minimum distance between $\boldsymbol{u}^{\text{nominal}}$ and $\boldsymbol{u}^{\text{lb}}$ is maximised. This way, it is more likely that the resultant bounds will be useful in operation since there will be no individual time-step for which set point control is effectively required.

The upper bounds can be identified with similar strategies as follows:

$$\begin{aligned}
\max_{\substack{\boldsymbol{u}_k^{\text{ub}}\\ k=0,\dots,N-1}} \quad & \min_{j=1}^{m} \min_{k=0}^{N-1} \ \overline{w}_{j,k} \cdot (u_{j,k}^{\text{ub}} - u_{j,k}^{\text{nominal}}) \\
\text{s.t.} \quad & \overline{w}_{j,k} = \frac{1}{\overline{u}_{j,k} - u_{j,k}^{\text{nominal}}} \\
& \boldsymbol{x}_{k+1} = \boldsymbol{f}(\Delta t, \boldsymbol{x}_k, \boldsymbol{u}_k^{\text{ub}}, \boldsymbol{\theta}, \boldsymbol{\xi}) \\
& \boldsymbol{x_0} = \boldsymbol{x}(0) \\
& \boldsymbol{g_{x_k}} \leq 0 \\
& \boldsymbol{u}_k^{\text{nominal}} \leq \boldsymbol{u}_k^{\text{ub}} \leq \overline{\boldsymbol{u}}_k.
\end{aligned} \tag{5}$$

Using this two-step algorithm, one can obtain an upper and lower bound that each accommodate uncertainties whilst respecting process constraints which, provides an optimal operational region in which control of the design variables can be loosened, and still achieve good process performance and reach desired product quality requirements.

Because the two-step algorithm addresses the upper and lower bounds independently, each bound is guaranteed to meet all constraints individually. However, this does not ensure that any control trajectories sampled between these bounds will also comply. Therefore, it is essential to validate these bounds by randomly sampling control actions within them and recording any instances that violate the constraints. These recorded samples can then be returned to the framework using a well designed penalty, which can be added to the objective function in (4) and (5), to reduce the bounds such that these violating control trajectories cannot be sampled. Specifically, the penalty function, q, is defined as follows:

$$q = \sum_{j=1}^{m} \sum_{k=0}^{N-1} \sum_{z=1}^{c_v} Q_{j,k,z}, \tag{6}$$

where, c_v is the total number of violating control trajectories recorded in the validation step, and $Q_{j,k,z}$ is the penalisation parameter, which equals 0, if $u_{j,k}^{\text{ub}} \leq u_{j,k,z}$, and equals 1, otherwise, $\forall j, k, z$. As noted in Fig. 2, we terminate the process if the violation of the validation set is less than or equal to 1%.

3. CASE STUDIES

3.1 Fed-batch Bioprocess Optimisation

The proposed method is first examined with a case study of astaxanthin production in a fed-batch bioprocess under parameter uncertainty (Vega-Ramon et al., 2021), *i.e.*, the source of uncertainty arises from the model parameters.

The kinetic model of the process is taken from Vega-Ramon et al. (2021), which is converted to a fed-batch model by including the feed flow rate $F_{\text{in}}(t)$ as the control input:

$$\begin{aligned}
\frac{dV(t)}{dt} &= F_{\text{in}}(t) \\
\frac{dX(t)}{dt} &= \mu_m(t)X(t) - \mu_d X(t) - \frac{X(t)}{V(t)} F_{\text{in}}(t) \\
\frac{dS(t)}{dt} &= \frac{F_{\text{in}}(t)}{V(t)} S_{\text{in}} - Y_S \mu_m(t) X(t) - \frac{S(t)}{V(t)} F_{\text{in}}(t) \\
\frac{dP(t)}{dt} &= \alpha \mu_m(t) X(t) + \beta X(t) - k_d X^2(t) \left(\frac{P(t)}{P(t) + 0.01} \right) - \frac{P(t)}{V(t)} F_{\text{in}}(t),
\end{aligned}$$

where $\mu_m(t) = \mu_m \frac{S(t)}{S(t)+K_c X(t)}$. Here, $X(t)$ is the biomass concentration (g/L), $S(t)$ is the substrate concentration (g/L), S_{in} is the substrate concentration in the feed flow, $P(t)$ is the product concentration (mg/L), t is the time (h), and $V(t)$ is the volume of the reactor. Model parameters $\mu_m, K_c, \mu_d, Y_S, \alpha, \beta$, and k_d are uncertain, whose nominal values are shown in Table 1. The standard deviations of the parameters are taken as 2% variations around their nominal values.

The process begins with an initial volume, $V(0)$, of 0.4 L, an initial biomass concentration, $X(0)$, of 0.1 g/L, an initial substrate concentration, $S(0)$, of 6 g/L, and an initial product concentration, $P(0)$, of 0 mg/L. The substrate concentration in the feed, S_{in}, is set to 12 g/L, and the reactor's volume is 2 L (V_{reactor}). The process is run for 168 hours (T), The feed flow rate, $F_{\text{in}}(t)$, is adjusted every 12 hours, leading to 14 discretised time segments, (N). The control goal in this case is to ensure that the mass production of astaxanthin at the end of the time horizon is higher than 58 mg, *i.e.*, $y_{\text{SP}} = 58$ in Eqn. (2). Several path constraints are involved: (i) all the state variables are nonnegative, *i.e.*, $\boldsymbol{x}(t) \geq 0$, for $t \in [0, T]$, where $\boldsymbol{x}(t) = [V(t), X(t), S(t), P(t)]^\top$; ($ii$) maximum working volume should not exceed 95% of the reactor's volume, *i.e.*, $V(t) \leq 0.95 V_{\text{reactor}}$; ($iii$) no feed in the first and last days (24 hours), *i.e.*, $u_k = 0$, for $k \in \{0, 1, N-3, N-2\}$; (iv) the substrate concentration should not change dramatically, *i.e.*, $\frac{dS(t)}{dt} \leq 3$ g/L; and (v) the mass of the product astaxanthin should increase throughout the process, *i.e.*, $\frac{dP(t)}{dt} \geq 0$.

Table 1. Model parameters for astaxanthin production (Vega-Ramon et al., 2021).

Parameter	Nominal Value
μ_m (h^{-1})	0.43
K_c	63.7
μ_d (h^{-1})	0.0021
Y_S (g/g)	2.58
α (mg/g)	7.88
β (mg/g h^{-1})	0.236
k_d (mg L g^{-2} h^{-1})	0.0648

In this case, 1,000 samples are used to construct the scenario tree. One cluster is identified, whose identified operational region is shown in Fig. 3. The identified operational region is validated with 10,000 samples, where no violation was observed. The proposed method effectively identifies the operational space for the feed flow rate, $F_{\text{in}}(t)$, under parameter uncertainties, optimising both the lower and upper control bounds. This approach ensures that the process constraints are satisfied across all sampled uncertainties while maximising the optimal and flexible operational space.

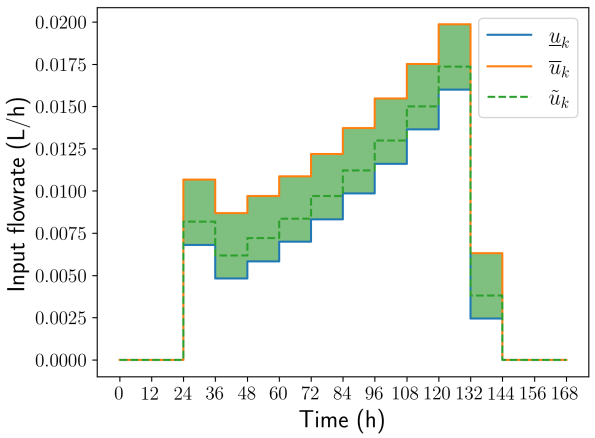

Fig. 3. Designed operational space (shaded) using the proposed method.

3.2 Formulation Product Quality Control

In order to examine the applicability of the framework to more complex, non-linear systems, we make use of a model proposed in (Rogers et al., 2024) which describes the material transformation between liquid products in a batch dynamic mixing process. The formulated product is constructed over time through a series of ingredient additions made to the process. The product composition is described through a series of 3 equations (material transformations) as functions of 3 operating conditions (control variables), and 11 inherent process states. A full description of the model is found in (Rogers et al., 2024).

The main challenge in this case study was the difficulties in reaching the desired end product quality subject to a myriad of process uncertainties that one faces in standard operation. The main sources of uncertainty were characterised into three classes; the first being variation in the input feed composition, which would be reflected in the estimated parameters of the model proposed by (Rogers et al., 2024). The second source of uncertainty is identified as human error, this is introduced to the systems primarily through difficulties in exactly following the times at which ingredient additions should be made to the process as indicated by the plant process flow diagram. The final source of uncertainty is system control error, that is deviations in the value of control variables from the set-points at which they are required to reach. The uncertainty levels assumed for these variables are 10%, 20% and 5% per standard deviation respectively. It should be stated that 100 uncertain scenarios were taken to assemble the scenario tree.

Specifically, one has access to dynamic control over 3 control variables, each of which are allowed to change 5 times, in order to maintain consistent product qualities.

The methodology displayed in Section 2 was applied in full, with 40 optimal set-points being established in step 2. In step 3, it was found that 2 clusters, containing distinct control pathways, were derived. The resulting control variable 1's nominal control actions and corresponding bounds for each cluster are shown in Fig. 4, alongside their reduced bounds (right) after refining them as described in step 4. Similarly, Fig. 5 shows the results for the control of control variable 2.

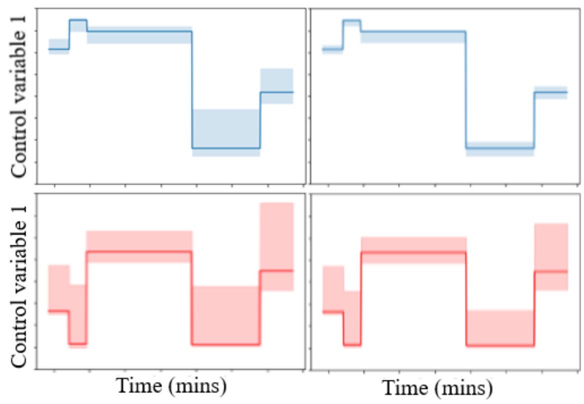

Fig. 4. Control variable 1 nominal set-point and bounds for cluster 1 (blue) and cluster 2 (red).

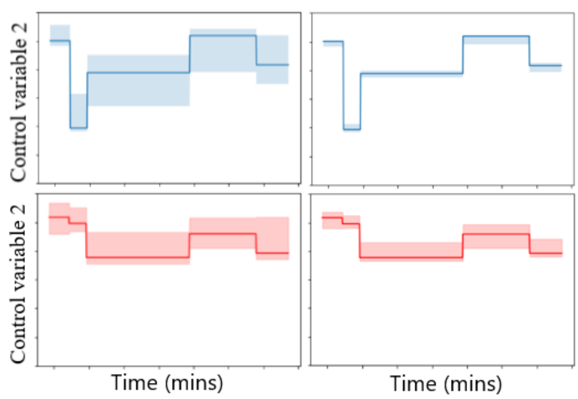

Fig. 5. Control variable 2 set-point and bounds for cluster 1 (blue) and cluster 2 (red).

The bounds (before refinement) were validated for both clusters by randomly sampling 1000 control trajectories from between the control variables' bounds for 100 unseen uncertain scenarios (i.e., 100,000 simulations). When these sampled control trajectories were simulated through the process model, it was found that cluster 1 had a product quality violation rate of 22% whilst cluster 2 possessed one of 9.5%. However, it should be noted that for each the maximum and mean number of scenarios violating for a given control action (only considering those that did violate) was 3 and 1.1 (out of 100) respectively. This is supportive that the initial bounds estimated are indeed robust to uncertainties. Once the bounds were refined, these violation rates dropped to 0% and 0.9% respectively.

An important note to make is that the first cluster, when refined, loses much of its operability and reaches somewhat of a set-point control status. In contrast, the second cluster retains it's operability in large, only reducing its control area by 24% and 39% for control variables 1 and 2, respectively, in comparison to the decrease in area of 51% and 79% for the first cluster. Therefore, there is an argument to be made for allowing a small violation rate, as is 0.9% for cluster 2, in order to improve the operability of the process.

4. CONCLUSION

In conclusion, operational space design offers a unique solution to accommodate uncertainties into the development of robust controls when one wishes to reliably meet key process constraints. The combination of flexibility and optimality into the development of operational spaces allows for both reliable and high-performance control strategies. The increase in operability was shown to be a key advantage over classic set-point control, where one is restricted to the operation of a process within tight guidelines. Furthermore, the systematic approach of identifying different operating regions within the entire optimal space provides a distinct advantage over traditional set-point control, in which one may overlook the existence of such regions. This way, the methodology can be also used for knowledge gain about how a process may be operated; which can provide benefits for process operators to identify regions in which the process may be more stable and easier to control within.

The proposed framework's effectiveness was demonstrated using case studies on dynamic batch processes, including a fed-batch fermentation process and the production of a consumer goods formulation product. Results indicate that our approach effectively manages uncertainties, offering economically advantageous process designs and operational strategies while upholding critical process constraints on top of improving process operability. Overall, this work demonstrates the novel combination of optimal and flexible design space identification and dynamic optimisation under uncertainty. Moreover, the potential of the methodology for situations in which consistent violations of process constraints exist due to unmanageable uncertainties, or in cases where it is difficult to achieve consistent product quality has been demonstrated.

REFERENCES

Amrih, P. and Damayanti, R.W. (2022). Pharma 4.0 quality management challenge: A literature review. In *Proceedings of 12th Annual International Conference on Industrial Engineering and Operations Management (IEOM) Conference.*

Bomze, I.M. and Gabl, M. (2023). Optimization under uncertainty and risk: Quadratic and copositive approaches. *European Journal of Operational Research*, 310(2), 449–476.

Gabrel, V., Murat, C., and Thiele, A. (2014). Recent advances in robust optimization: An overview. *European journal of operational research*, 235(3), 471–483.

García, J. and Peña, A. (2018). Robust optimization: concepts and applications. *Nature-inspired methods for stochastic, robust and dynamic optimization*, 7.

Geletu, A., Klöppel, M., Zhang, H., and Li, P. (2013). Advances and applications of chance-constrained approaches to systems optimisation under uncertainty. *International Journal of Systems Science*, 44(7), 1209–1232.

Hicks, A., Johnston, M., Mowbray, M., Barton, M., Lane, A., Mendoza, C., Martin, P., and Zhang, D. (2021). A two-step multivariate statistical learning approach for batch process soft sensing. *Digital Chemical Engineering*, 1, 100003.

Hill, K. (2007). Industrial development and application of biobased oleochemicals. *Pure and Applied Chemistry*, 79(11), 1999–2011.

Jia, H., Ding, S., Xu, X., and Nie, R. (2014). The latest research progress on spectral clustering. *Neural Computing and Applications*, 24, 1477–1486.

Kammammettu, S. and Li, Z. (2023). Scenario reduction and scenario tree generation for stochastic programming using sinkhorn distance. *Computers & Chemical Engineering*, 170, 108122.

Khan, K., Rehman, S.U., Aziz, K., Fong, S., and Sarasvady, S. (2014). Dbscan: Past, present and future. In *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*, 232–238. IEEE.

Na, S., Xumin, L., and Yong, G. (2010). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In *2010 Third International Symposium on intelligent information technology and security informatics*, 63–67. Ieee.

Rogers, A.W., Lane, A., Mendoza, C., Watson, S., Kowalski, A., Martin, P., and Zhang, D. (2024). Integrating knowledge-guided symbolic regression and model-based design of experiments to accelerate process flow diagram development. *IFAC-PapersOnLine*, 58(14), 127–132.

Silvente, J., Papageorgiou, L.G., and Dua, V. (2019). Scenario tree reduction for optimisation under uncertainty using sensitivity analysis. *Computers & Chemical Engineering*, 125, 449–459.

Vega-Ramon, F., Zhu, X., Savage, T.R., Petsagkourakis, P., Jing, K., and Zhang, D. (2021). Kinetic and hybrid modeling for yeast astaxanthin production under uncertainty. *Biotechnology and Bioengineering*, 118(12), 4854–4866.

Zheng, Q.P., Wang, J., and Liu, A.L. (2014). Stochastic optimization for unit commitment—a review. *IEEE Transactions on Power Systems*, 30(4), 1913–1924.