Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

Identification and Self-optimization of Robust Nominal Operating Ranges Using Proximal Policy Optimization

Ashish Yewale[a], and Brahim Benyahia[a]*

[a] Department of Chemical Engineering, Loughborough University, Leicestershire, LE11 3TU, UK
* Corresponding Author: b.benyahia@lboro.ac.uk

ABSTRACT

Reliable process operation under uncertainty remains a fundamental challenge in chemical and pharmaceutical manufacturing. Variability arising from feed fluctuations, material properties, external disturbances, and uncertain model parameters can significantly impact feasibility and performance when operating ranges are designed under nominal assumptions. Design Space Identification (DSI) addresses these limitations by defining a Probabilistic Design Space (PDS), within which process constraints are satisfied with a prescribed confidence level. However, identifying and extracting robust and practically usable nominal operating ranges (NORs) from complex, high-dimensional, and nonconvex PDSs remains computationally demanding. This work proposes a novel reinforcement learning (RL)–based framework for automated identification of the largest robust hyper-rectangular NOR fully contained within a given PDS. The design centering problem is formulated as a sequential decision-making task and solved using the Proximal Policy Optimization (PPO) algorithm. By interacting with a probabilistic feasibility environment, the RL agent adaptively adjusts the location and size of the NOR to maximize its area while enforcing probabilistic constraint satisfaction. Deep neural network policies enable efficient exploration of nonlinear feasibility boundaries without exhaustive sampling or gradient-based optimization. The approach is demonstrated on a tablet lubrication case study with parametric uncertainty. Results show rapid convergence to well-located, near-maximal NORs using substantially fewer probabilistic feasibility evaluations than conventional grid-based or Monte Carlo methods. The proposed PPO-based framework provides a scalable and data-efficient solution for robust operating range identification in uncertainty-aware process design.

Keywords: Probabilistic design space (PDS), Nominal Operating Range (NOR), Reinforcement Learning, Operational flexibility, Proximal Policy Optimization

1. INTRODUCTION

Chemical and pharmaceutical manufacturing is undergoing rapid transformation driven by the need for increased process efficiency, reduced development timelines, and improved product quality [1-2]. Within this evolving context, the Quality by Design (QbD) paradigm has emerged as a central framework for modern process development, emphasizing systematic process understanding and risk-based decision-making [3-4]. Central to QbD is the establishment of quantitative relationships among Critical Quality Attributes (CQAs), process parameters, and material attributes, enabling reliable and reproducible operation across a range of conditions [5-6].

Achieving such robustness is inherently challenging due to multiple sources of uncertainty, including raw material variability, feed disturbances, unmeasured process dynamics, and uncertain model parameters. To explicitly account for these uncertainties, Design Space Identification (DSI) methodologies have been developed to characterize the regions of feasible operation that satisfy all process and quality constraints. In recent years, DSI extends this concept by defining a Probabilistic Design Space (PDS), comprising operating conditions that satisfy all process and quality constraints with a prescribed confidence level α (e.g., ≥ 95 %) [7-10].

While probabilistic DSI provides a rigorous uncertainty-aware framework, the resulting PDSs are often irregular, nonconvex, and disconnected, particularly in nonlinear and high-dimensional systems. As a result, PDSs are difficult to use directly in industrial practice. Instead, operation typically relies on a Nominal Operating Range (NOR), commonly defined as a hyper-rectangular region fully contained within the PDS. NORs are interpretable, easily implementable, and well aligned with control strategies and regulatory documentation.

Identifying the largest feasible NOR within a PDS corresponds to the classical design centering problem. However, ensuring probabilistic feasibility throughout an entire operating region is not merely at isolated points, it renders this problem computationally prohibitive for conventional approaches such as exhaustive grid searches, Monte Carlo sampling, or heuristic expansion, especially as dimensionality increases. Moreover, point-based optimization strategies often yield operating conditions that appear optimal in simulation but lack robustness under real-world disturbances [11-12].

These limitations motivate the development of methods that directly characterize feasible regions and efficiently navigate complex probabilistic design spaces. In this work, we propose a novel reinforcement learning (RL)–based framework for automated design centering within PDSs. Recent advances in RL for process systems engineering have demonstrated its potential for control, scheduling, and optimization under uncertainties [3, 13-15]; however, its application to probabilistic design space exploration and robust operating range identification remains largely unexplored. This work bridges that gap by introducing a scalable, data-efficient, and uncertainty-aware RL methodology for robust NOR identification, offering a practical pathway for translating probabilistic design concepts into actionable operating guidelines. By formulating NOR identification as a sequential decision-making problem, an RL agent iteratively refines the location and size of an operating region candidate while enforcing probabilistic feasibility constraints. Proximal Policy Optimization (PPO) is employed to enable stable and sample-efficient exploration without requiring explicit gradients, convexity assumptions, or exhaustive enumeration of feasible points.

2. METHOD

2.1 Problem Statement and Scope

This work addresses the design centering problem within a given PDS. The PDS is assumed to be available *a priori*, having been identified using established probabilistic design space methodologies, as reported in previous work [9]. The objective is to determine the NOR that is fully contained within the given PDS while satisfying a prescribed probabilistic confidence level.

The NOR is parameterized as a hyper-rectangular region in the space of operating variables and is defined by its center and half-widths in each dimension. The design centering task therefore consists of identifying these parameters such that the resulting NOR maximizes operational flexibility while remaining probabilistically feasible throughout its entire extent. The NOR is defined as

$$NOR = \{x \in \mathbb{R}^n | \mu - \Delta \leq x \leq \mu + \Delta\} \quad (1)$$

where μ denotes the center of the operating region and Δ represents the vector of half-widths.

2.2 Design Centering Optimization Problem

The design centering objective is to identify the NOR parameters (μ, Δ) that maximize the area (or hypervolume) of the hyper-rectangular operating region while ensuring probabilistic feasibility throughout the entire NOR. This objective can be formulated as

$$\max_{\mu,\Delta} V(\Delta) = \prod_{i=1}^{n} 2\Delta_i \quad (2)$$

Subject to

$$\mathbb{P}_\theta(C(x,\theta) \leq 0) \geq \alpha, \forall x \in NOR \quad (3)$$

Unlike pointwise feasibility checks, constraint (3) requires that all operating conditions within the NOR satisfy the probabilistic constraints, rendering the problem highly challenging. The feasibility boundary is implicit, nonconvex, and typically requires repeated uncertainty propagation. As a result, conventional optimization, sampling-based, or gradient-based approaches become computationally prohibitive, particularly in high-dimensional settings. These challenges motivate the use of adaptive, data-driven approaches that can directly characterize and exploit the feasible region, rather than relying on dense sampling of the full design space.

3. METHODOLOGY: REINFORCEMENT LEARNING FRAMEWORK FOR DESIGN CENTERING

To address the design centering problem defined above, the NOR identification is formulated as a sequential decision-making problem and solved using RL. This formulation enables efficient exploration of complex probabilistic design spaces without relying on convexity assumptions or explicit gradient evaluations.

3.1 Markov Decision Process Formulation

The design centering task is reformulated as a continuous-state, continuous-action Markov Decision Process (MDP), in which an agent iteratively modifies the parameters of a candidate NOR and receives feedback based on probabilistic feasibility and operating range size.

State Space. At iteration t, the state is defined as,

$$s_t = [\mu_t, \Delta_t],$$

where μ_t denotes the current NOR center at iteration t and Δ_t represents the corresponding half-widths along each dimension.

Action Space. The action space corresponds to incremental updates of the NOR parameters,

$$a_t = [\Delta\mu_t, \Delta\Delta_t],$$

allowing the agent to shift the NOR center and expand or contract its dimensions.

State Transition. Given an action, the environment updates the NOR parameters and evaluates probabilistic feasibility using constraint (3). The updated parameters constitute the next state s_{t+1}. Infeasible transitions result in penalization and episode termination.

This formulation allows the agent to iteratively refine the location and size of the NOR while interacting directly with the probabilistic feasibility landscape.

3.2 Reward Function Design

The reward function is designed to promote identification of large, well-located NORs while strictly enforcing probabilistic feasibility. At each iteration, the RL agent receives a scalar reward defined as

$$R_t = w_1 V(\Delta_t) - w_2 \mathbb{I}(\text{violation}) - w_3 d_{\text{boundary}}, \tag{4}$$

where $V(\Delta_t) = \prod_{i=1}^{n} 2\Delta_{t,i}$ represents the hypervolume (area in this 2D problem) of the nominal operating region, $\mathbb{I}(\text{violation})$ is an indicator function equal to one if probabilistic feasibility is violated, and d_{boundary} is a boundary-based shaping term that penalizes proximity to infeasible regions.

The first term directly encodes the design centering objective, while the penalty terms ensure strict constraint satisfaction and promote robustness margins. The weighting coefficients w_1, w_2, and w_3 balance the competing objectives of maximizing operating range size and maintaining robust feasibility. Specifically, w_1 scales the hypervolume objective to encourage expansion of the NOR, w_2 imposes a strong penalty on probabilistic feasibility violations to discourage infeasible actions, and w_3 penalizes proximity to the PDS boundary to maintain a robustness margin. The weights are selected empirically to ensure that feasibility violations are strongly discouraged while still allowing the agent to expand the operating region within the feasible domain. Infeasible actions terminate the episode, ensuring that learned policies prioritize feasibility over aggressive expansion.

In addition to reward-based penalization, hard termination criteria are enforced to ensure strict feasibility. Episodes are terminated if (i) the NOR center lies outside the probabilistic design space at initialization, (ii) an action results in the NOR center escaping the PDS, or (iii) the updated hyper-rectangular region intersects or extends beyond the PDS boundary. These termination rules prevent infeasible policy updates and focus learning on robust and interpretable operating regions.

The resulting learning problem involves continuous state and action spaces, nonconvex feasibility boundaries, and discontinuities due to constraint violations. To address these challenges, PPO is employed. PPO provides stable and sample-efficient learning through a clipped surrogate objective that limits overly aggressive policy updates, making it well suited for exploration in complex PDSs. Without specialized constraint-handling mechanisms, the PPO agent adaptively refines the NOR parameters and converges toward a large, well-located operating region within the given PDS.

Case study: Tablet Lubrication process

Lubrication plays a critical role in tablet manufacturing, particularly in direct compression process, as it facilitates tablet ejection during compression. However, excessive lubrication can adversely affect tablet mechanical properties, most notably tensile strength. The interaction between lubrication extent and compaction pressure therefore represents a key trade-off in tablet manufacturing, where increased lubrication may reduce tensile strength, a loss that can be partially compensated by higher compaction pressures.

In this work, a tablet lubrication and compaction model previously reported in the literature is adopted [16-17]. Tablet tensile strength is modeled as:

$$\sigma = \sigma_{sf=0.85,0}\big((1-\beta) + \beta exp(-\gamma k)\big) \tag{5}$$

$$\sigma_{sf=0.85,0} = a_1 exp\big(b_1(1-sf)\big) \tag{6}$$

$$\beta = a_2(1-sf) + b_2 \tag{7}$$

where $\sigma_{sf=0.85,0}$ denotes the tensile strength at a solid fraction of 0.85 in the absence of lubrication, β represents the fraction of tensile strength loss attributable to lubrication, γ is the lubrication rate constant, k denotes the lubrication extent, and sf denotes the solid fraction. The model involves five uncertain parameters, $\theta = [a_1(MPa);\ b_1(-);\ a_2(-);\ b_2(-);\ \gamma(dm^{-1})]$ and captures the effects of solid fraction and lubrication extent on tablet tensile strength.

In the present study, solid fraction and lubrication extent are treated as CPPs, while tablet tensile strength is considered as a CQA. The PDS for this process, accounting for parameter uncertainty and probabilistic constraint satisfaction, has been previously identified and reported in Yewale et al. (2025)[9]. Here, this probabilistic design space is taken as a fixed feasibility domain, and the proposed reinforcement learning framework is applied to automatically identify a robust and well-located nominal operating range.

4. RESULTS

4.1. Probabilistic Design Space Characterization

The probabilistic design space (PDS) for the tablet lubrication case study, previously identified using nested sampling [9], is shown in Fig. 1 in the space of porosity and lubrication extent. Each sampled operating condition is classified according to the probability of satisfying all quality constraints, highlighting regions of high confidence (α ≥ 0.95), intermediate feasibility, and infeasible operation. The resulting PDS exhibits a distinctly nonconvex and asymmetric geometry, with feasibility rapidly degrading near the upper porosity boundary and at high lubrication extents. This irregular structure highlights the limitations of conventional pointwise feasibility analysis and local expansion strategies, which are ill-suited to capture the global structure of such complex probabilistic design spaces.

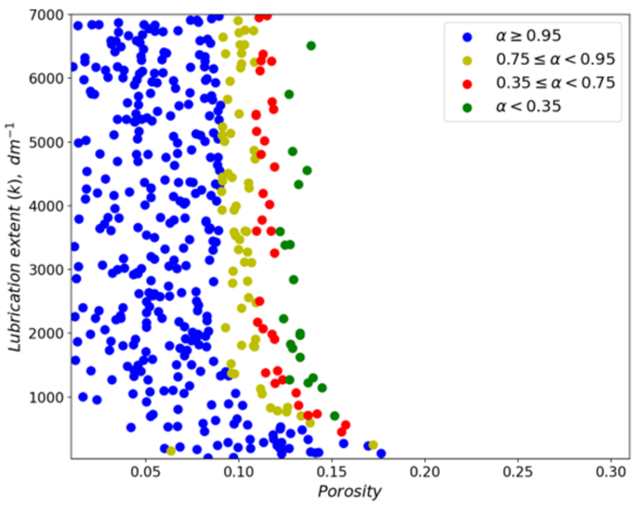

Figure 1. PDS identified using Nested sampling (Yewale et al 2025 [9])

4.2. Learning Dynamics and Convergence Behavior

The training and evaluation performance of the PPO agent are summarized in Fig. 2(a) and Fig. 2(b), respectively. During training, the mean episodic reward increases rapidly within the initial phase, indicating effective early exploration and rapid identification of feasible regions within the PDS. This phase is characterized by relatively high reward variance, reflecting exploratory behavior and frequent interactions with feasibility boundaries.

As training progresses, reward trajectories gradually stabilize, reflecting a transition from exploration toward exploitation and policy refinement. Occasional reward fluctuations persist due to stochastic exploration, but the overall trend reflects consistent improvement in the agent's ability to identify large, feasible operating regions.

The results shown in Fig. 2(b) exhibit significantly reduced variance and sustained high reward levels, confirming that the learned policy generalizes well beyond individual training episodes. This behavior confirms that the learned policy generalizes well beyond individual training episodes and reliably identifies large, feasible operating regions. Importantly, convergence is achieved using substantially fewer probabilistic feasibility evaluations than would be required by exhaustive grid-based or Monte Carlo design centering approaches, underscoring the computational efficiency of the proposed RL framework.

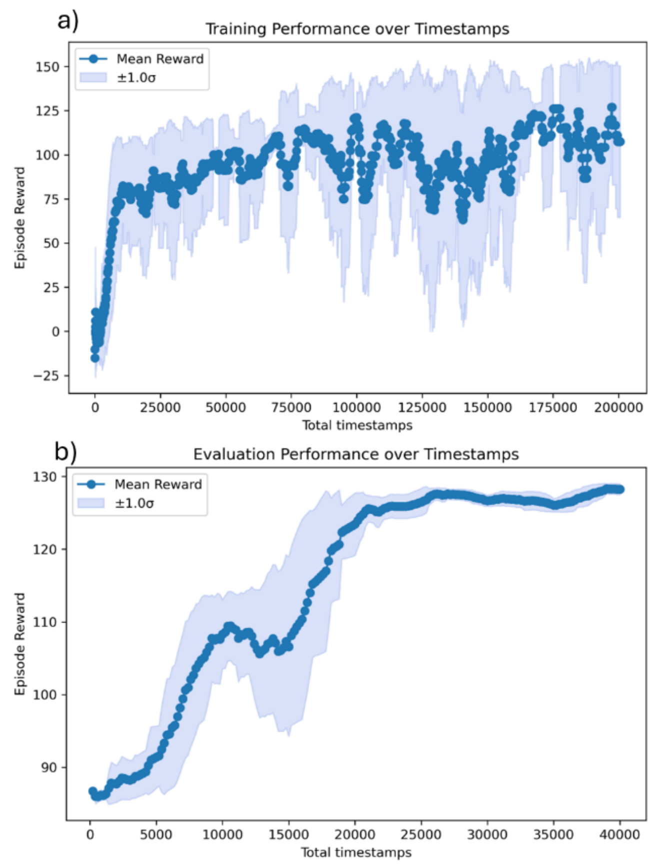

Figure 2. Training and evaluation performance of the PPO agent

4.3. Policy Behavior and Trajectory Analysis

A representative sample of the PPO trajectories across multiple episodes is shown in Fig. 3, illustrating the episodic refinement of the NOR within the PDS during learning. Each episode begins from a randomly initialized NOR center, indicated by a blue circular marker, located within the probabilistic design space, resulting in diverse starting points across the feasible region. The red cross denotes the terminal NOR center at the end of the episode, corresponding to the final operating region returned by the policy after the sequence of updates.

During early learning episodes, the agent executes a sequence of incremental updates to the NOR center and dimensions, guided by the reward structure and feasibility feedback. The resulting trajectories exhibit exploratory behavior, including corrective movements away from nearby infeasible boundaries and directional adjustments toward regions of higher feasibility. These exploratory steps reflect the agent's interaction with the underlying PDS geometry, as it learns to balance NOR expansion with constraint satisfaction. Importantly, the trajectories remain confined to feasible regions, indicating that boundary violations are effectively penalized during learning.

As learning progresses, the trajectories become shorter, smoother, and more directed, converging toward a common region near the interior of the PDS. This transition reflects policy stabilization, where the agent has acquired an internal representation of the feasible region and no longer requires extensive exploration. The resulting trajectories consistently terminate at similar NOR centers, demonstrating convergence toward a well-located and robust operating region.

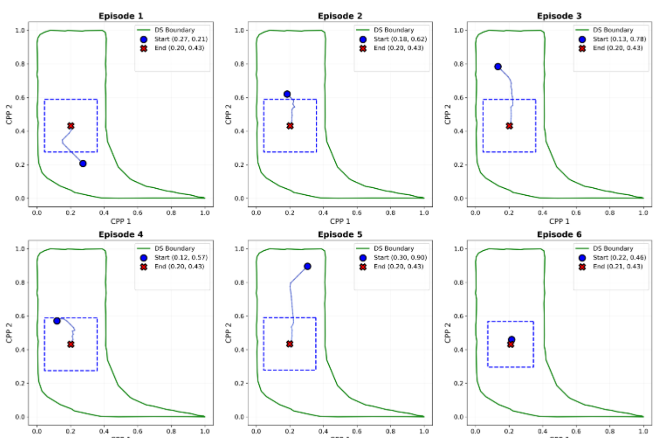

Figure 3. Representative PPO trajectories showing episodic refinement of the NOR within the PDS

The final and best-performing NOR identified by the trained PPO agent is shown on Fig. 4. The resulting hyper-rectangular operating region is fully contained within the PDS and is well-located relative to the feasibility boundaries. The identified NOR achieves a near-maximal feasible area while maintaining a clear margin from infeasible regions, highlighting the effectiveness of the boundary-aware reward design. Notably, repeated training runs converge to nearly identical NOR centers and dimensions, providing strong evidence that the PPO-based framework identifies a globally robust solution rather than a local or initialization-dependent optimum. This repeatability is particularly important for industrial deployment, where reproducibility and interpretability of operating ranges are critical.

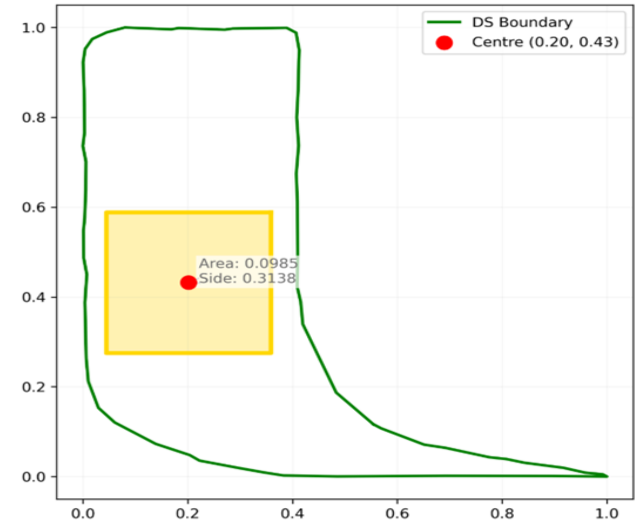

Figure 4. Best-performing nominal operating range identified by the PPO agent, fully contained within the probabilistic design space

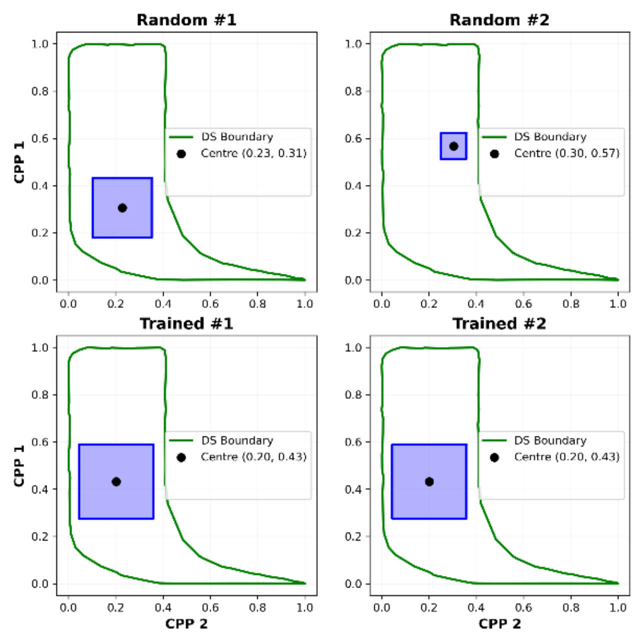

Figure 5. Comparison between randomly initialized operating regions and NORs identified by the trained PPO policy

To further assess robustness, Fig. 5 compares NORs obtained from randomly initialized centers with those identified by the trained PPO policy. Random initialization leads to highly variable operating regions, many of which either intersect the PDS boundary or yield overly conservative ranges that fail to exploit available operational flexibility. In contrast, the PPO-based approach consistently produces similar NOR centers and dimensions across independent runs, demonstrating robustness with respect to initialization and stochasticity.

5. CONCLUSION

This work introduced a novel RL-based framework for automated identification of robust nominal operating ranges within the probabilistic design spaces. By formulating the design centering problem as a sequential decision-making task and solving it using PPO, the proposed approach enables efficient exploration of complex, non-convex feasibility regions under uncertainty.

Application to a tablet lubrication case study demonstrated rapid convergence to well-located, near-maximal NORs using substantially fewer probabilistic feasibility evaluations than conventional methods. The learned policy exhibited strong robustness, repeatability, and insensitivity to initialization, all of which are critical for industrial deployment.

The proposed framework provides a scalable and data-efficient alternative to traditional design centering techniques and offers a practical pathway for translating probabilistic design spaces into actionable and robust operating ranges. Future work will extend the approach to higher-dimensional systems and investigate integration with digital twins and adaptive process design workflows.

ACKNOWLEDGEMENTS

This project has received funding under the call HORIZON-HLTH-2021-IND-07 (SusPharma, grant agreement no. 101057430) and under the UKRI Horizon Europe Guarantee scheme (SusPharma, project reference no. 10038378).

REFERENCES

1. Gao S, Benyahia B. Robust techno-economic analysis, life cycle assessment, and quality and sustainability by digital design of three alternative continuous pharmaceutical tablet manufacturing processes. Systems and Control Transactions 4:2568-2573 (2025). https://doi.org/10.69997/sct.104102
2. Campbell, T. J. S.; Rielly, C. D.; Benyahia, B. Digital Design and Optimization of an Integrated Reaction-Extraction-Crystallization-Filtration Continuous Pharmaceutical Process. Comput. -Aided Chem. Eng. 2022, 51, 775–780. https://doi.org/10.1016/B978-0-323-95879-0.50130-2
3. Peterson JJ, Yahyah M. A bayesian design space approach to robustness and system suitability for pharmaceutical assays and other processes. Statistics in Biopharmaceutical Research 1:441-449 (2009). https://doi.org/10.1198/sbr.2009.0037
4. Peterson JJ. A bayesian approach to the ICH Q8 definition of design space. Journal of Biopharmaceutical Statistics 18:959-975 (2008). https://doi.org/10.1080/10543400802278197
5. von Stosch M, Schenkendorf R, Geldhof G, Varsakelis C, Mariti M, Dessoy S, Vandercammen A, Pysik A, Sanders M. Working within the design space: do our static process characterization methods suffice?. Pharmaceutics 12:562 (2020). https://doi.org/10.3390/pharmaceutics12060562
6. Duarte JG, Duarte MG, Piedade AP, Mascarenhas-Melo F. Rethinking pharmaceutical industry with quality by design: application in research, development, manufacturing, and quality assurance. AAPS J 27: (2025). https://doi.org/10.1208/s12248-025-01079-w
7. Laky D, Xu S, Rodriguez JS, Vaidyaraman S, García Muñoz S, Laird C. An optimization-based framework to define the probabilistic design space of pharmaceutical processes with model uncertainty. Processes 7:96 (2019). https://doi.org/10.3390/pr7020096
8. Meng Q, Bogle D, Charitopoulos VM. Probabilistic design space exploration and optimization via bayesian approach for a fluid bed drying process. European Journal of Pharmaceutical Sciences 210:107116 (2025). https://doi.org/10.1016/j.ejps.2025.107116
9. Yewale A, Yuan X, Benyahia B. Robust pharmaceutical tableting process through combined probabilistic design space and flexibility analysis. Systems and Control Transactions 4:637-643 (2025). https://doi.org/10.69997/sct.128065
10. Chiplunkar R, Pauzi SM, Sachio S, Papathanasiou MM, Kontoravdi C. Probabilistic design space identification for upstream bioprocesses under limited data availability?. Systems and Control Transactions 4:2561-2567 (2025). https://doi.org/10.69997/sct.166359
11. Kusumo KP, Morrissey J, Mitchell H, Shah N, Chachuat B. A design centering methodology for probabilistic design space. IFAC-PapersOnLine 54:79-84 (2021). https://doi.org/10.1016/j.ifacol.2021.08.222
12. Moshiritabrizi I, McMullen JP, Wyvratt BM, McAuley KB. A comparative study of strategies for incorporating uncertainty in design space determination for pharmaceutical manufacturing. Ind. Eng. Chem. Res. 64:21658-21668 (2025). https://doi.org/10.1021/acs.iecr.5c03037
13. Yewale A, Yang Y, Nazemifard N, Papageorgiou CD, Rielly CD, Benyahia B. Deep reinforcement learning-based self-optimization of flow chemistry. ACS Eng. Au 5:247-266 (2025). https://doi.org/10.1021/acsengineeringau.5c00004
14. Anandan, P.D.; Rielly, C. D.; Benyahia, B.; Optimal

Control Policies of a Crystallization Process Using Inverse Reinforcement Learning, Computer Aided Chemical Engineering, 2021 51, 1093-1098 . https://doi.org/10.1016/B978-0-323-95879-0.50183-1
15. Anandan, P.D.; Rielly, C. D.; Benyahia, B.; Control of Batch and Continuous Crystallization Processes using Reinforcement Learning. Computer Aided Chemical Engineering, 2021, 50, 1371-1376. https://doi.org/10.1016/B978-0-323-88506-5.50211-4
16. Kushner J IV, Moore F. Scale-up model describing the impact of lubrication on tablet tensile strength. International Journal of Pharmaceutics 399:19-30 (2010). https://doi.org/10.1016/j.ijpharm.2010.07.033
17. Nassar J, Williams B, Davies C, Lief K, Elkes R. Lubrication empirical model to predict tensile strength of directly compressed powder blends. International Journal of Pharmaceutics 592:119980 (2021). https://doi.org/10.1016/j.ijpharm.2020.119980