Efficient Experimental Design with an AI-Driven Tool Based on Bayesian Optimization

M. Pournajar[1], L. Wölfel[1], R. Q. Albuquerque[2], V. Eberhardt[3], A Lehmann[3] and C. Kuenneth[1]*

[1] Department of Computational Materials Science (Materials Informatics), Faculty of Engineering Science, University of Bayreuth, Bayreuth, Germany,

[2] Department of Polymer Engineering, Faculty of Engineering Science, University of Bayreuth, Bayreuth, Germany,

[3] Fraunhofer-Institut für Angewandte Polymerforschung IAP, Potsdam, Germany

* *Corresponding author (Christopher.Kuenneth@uni-bayreuth.de)*

Keywords: *machine learning; Bayesian optimization; sustainability; mechanical properties*

1. Abstract

Optimizing fiber properties in wet-spinning is challenging because elongation at break depends on many strongly coupled process and formulation variables. In this work, we investigated a multidimensional dataset from cellulose fiber production that included key formulation and process variables (such as cellulose content, die geometry, temperatures, airflow, humidity, and take-up speed) as well as resulting fiber properties (including titer, tenacity, E-modulus, and elongation at break). To efficiently identify promising process settings, Bayesian optimization was applied to the existing experimental dataset as a data-driven strategy for navigating the complex nonlinear relationships between processing variables and elongation at break. Within this proof-of-concept study, Bayesian optimization was used on the existing experimental dataset to assess whether target elongation values could be reached with fewer trials than required in the historical experimental campaign. A Gaussian-process surrogate model and acquisition function were employed to iteratively select informative and promising experiments from the dataset. This approach is well suited for process optimization because it can reduce the number of experiments needed to identify favorable operating conditions in a high-dimensional and expensive experimental space. The results demonstrate that Bayesian optimization is an effective framework for accelerating the optimization of cellulose fiber spinning processes.

2. Introduction

Designing polymer-based fibers for specific applications requires multi-objective optimization of often competing properties within a narrow target window [1]. In high-performance cellulose and other artificial fibers, elongation at break is a key descriptor of toughness, damage tolerance, and textile processing behavior, and must be engineered together with other mechanical properties to meet application-specific requirements [2, 3]. Achieving such balance is particularly challenging in wet-spinning and related fiber-forming processes, where small changes in dope composition, spinneret geometry, coagulation and air-gap conditions, or take-up speed can produce large and highly nonlinear changes in fiber properties [4, 5]. Consequently, identifying optimal processing conditions within large design spaces becomes difficult, and intuition-driven trial-and-error approaches or conventional design-of-experiments methods become increasingly impractical as the number of relevant variables grows [6]. Over the past decade, machine-learning (ML) and Bayesian optimization (BO) have emerged as powerful tools for accelerating materials discovery and process optimization under data-sparse and experimentally expensive conditions [7]. In a typical BO framework, a probabilistic surrogate model is trained on experimental data and iteratively updated through active learning, while an acquisition function balances exploration of uncertain regions against exploitation of promising conditions. This paradigm has been successfully applied across diverse materials problems, including composition and microstructure optimization, mixed qualitative–quantitative design spaces, and autonomous or semi-autonomous laboratory workflows [8, 9]. In polymer science, BO-based active-learning strategies have been used to optimize polymerization protocols, design polymers with targeted rheological responses, and accelerate the discovery of polymer systems with enhanced thermal and mechanical performance [10, 11]. BO-driven active learning has also begun to impact fiber-forming processes and fiber property optimization. For example, rapid Bayesian optimization has been used to guide a multivariate fluidic platform toward processing conditions yielding short polymer fibers with targeted length, diameter, and morphology, achieving desired property windows in only a few tens of experiments [1]. Similarly, Bayesian optimization of dope preparation and spinning conditions has significantly improved the tensile strength, modulus, and toughness of wet-spun silk fibers using only a limited number of carefully selected experiments. These studies highlight the potential of BO-based active learning to efficiently navigate complex processing–structure–property relationships in fiber systems, particularly when quantitative and qualitative objectives must be balanced [12]. Machine learning–based optimization has been widely employed in polymer-related applications [13, 14, 15]. Albuquerque et al. [16] applied Bayesian optimization to maximize the glass-transition temperature (T_g) of an epoxy resin system containing one resin and seven amino acid curing agents at a unity stoichiometric ratio (R), achieving the maximum T_g after only nine additional experiments. More broadly, BO-based approaches have increasingly been applied to material formulations, primarily focusing on optimization of single target properties and attracting growing industrial interest [17]. In formulation design, Gaussian-process (GP) surrogate models combined with acquisition functions such as expected improvement (EI), probability of improvement (PI), and upper confidence bound (UCB) have been successfully used to balance exploration and exploitation. Albuquerque et al. [18] reported that EI provided the most effective trade-off in their formulation studies. Pruksawan et al. [15] demonstrated the use of active learning and Bayesian optimization for the optimization of epoxy-based adhesives using a small dataset with four variables. After 47 experiments, they identified an adhesive joint strength approximately 13% higher than the maximum value obtained in the preceding experiments. Kim et al. [19] employed Gaussian-process regression within an active-learning framework to discover polymers with high glass-transition temperatures, comparing exploitation, exploration, and balanced selection strategies against random sampling. They have reported a virtual experiment using different active learning (AL) strategies to screen 736 different polymers to find high-Tg ones. Active learning extends conventional ML by enabling the model to iteratively guide selection of the next experiment, thereby improving learning efficiency and reducing experimental burden [6][20]. In the present work, we extend an active-learning framework based on Bayesian optimization to maximize elongation at break in cellulose-based fibers. Starting from an experimentally generated dataset of 98 wet-spinning trials varying cellulose content, spinneret and air-gap geometry, thermal and flow conditions, and take-up speed, we train a Gaussian-process surrogate model to capture relationships between process parameters and mechanical performance, with particular emphasis on elongation at break. Within an iterative BO framework, acquisition functions propose new process conditions that balance exploration of poorly characterized regions with exploitation of settings predicted to yield high elongation. Our results demonstrate that this strategy efficiently identifies processing windows yielding substantially improved elongation relative to baseline conditions while requiring far fewer experiments. Although demonstrated here for cellulose-based fibers, the proposed framework is general and may be extended to other fiber-forming processes and multi-objective optimization problems in advanced polymeric materials.

3. Materials and Methods

3.1. Lyocell Technology

Lyocell is an established process for producing regenerated cellulosic fibres, with high-purity wood pulp as the usual feedstock and alternative or recycled cellulose sources expected to become more relevant. At roughly 0.4 million tonnes per year it is still a small share of the some 8.4 million tonnes of regenerated cellulosic fibres produced in 2024, but it is a growing one, used in apparel, home textiles, nonwovens and selected technical applications.

What sets lyocell apart from viscose is that the cellulose is dissolved directly, without any chemical derivatisation, in N-methylmorpholine N-oxide (NMMO). The resulting viscous dope is extruded through a spinneret, and the filaments are drawn across a short air gap before entering an aqueous coagulation bath, a configuration known as dry-jet wet spinning. The draw in this air gap orients the cellulose chains and sets the crystallinity of the fibre, and it is largely this orientation that governs the mechanical properties, including the elongation at break we optimise here [2]. In the bath the cellulose precipitates as cellulose II; the filaments are then washed, and the NMMO is recovered and recycled at rates above 99%. Finishing, drying and, where required, crimping and cutting complete the process.

For optimisation the key point is that no single step determines the final fibre quality. Pulp, dope preparation, spinneret geometry, air-gap draw, coagulation, washing and drying all interact, which is precisely the kind of coupled, high-dimensional problem where data-driven methods can help.

3.2. *Method*

The dataset was generated by the IAP institute using the above-mentioned Lyocell technology. We begin by selecting three experiments from a dataset of 98 wet-spinning trials, excluding samples that meet the target elongation threshold, and assume that elongation is initially known only for these selected samples. These data form the initial training set used to fit a Gaussian process regression (GPR) model, where the processing parameters serve as inputs and elongation at break is the target response. The remaining 95 experiments are not provided to the model at this stage. Using the BO workflow illustrated in Figure 1., we iteratively identify the most informative next experiment from the pool of unobserved samples to efficiently discover process conditions yielding high-elongation fibers. In this proof-of-concept study, no new fibers are physically produced; instead, the experimentally measured elongation values are sequentially revealed from the existing dataset. After each selection, the corresponding data point is appended to the training set, the GPR model is retrained, and the remaining candidates are re-evaluated. This iterative procedure continues until a sample meeting the predefined elongation threshold is identified. The BO process follows a closed-loop cycle in which an initial subset of experiments is selected and used to fit the GPR surrogate, after which the model predicts both the mean elongation and its uncertainty for all unobserved candidates. An acquisition function is then computed, and the next experiment is selected by maximizing this criterion. The corresponding elongation is obtained, the dataset is updated, and the surrogate model is refit before repeating the cycle until the target elongation threshold is reached. For experiment selection, we adopt a balanced exploration–exploitation strategy based on the expected improvement (EI) acquisition function. This approach prioritizes candidates that offer either high predicted elongation or high uncertainty, thereby enabling efficient navigation of the design space. As a result, regions that are poorly characterized are explored while simultaneously exploiting areas already predicted to yield favorable outcomes. Within this GPR-based framework, the EI criterion adaptively guides the search toward high-performing samples. To assess the statistical robustness of the proposed active-learning strategy, the entire procedure is repeated 10 times with different randomly selected initial samples. For each run, we record the number of iterations required to reach the elongation threshold, enabling a quantitative evaluation of the efficiency and variability of the approach.

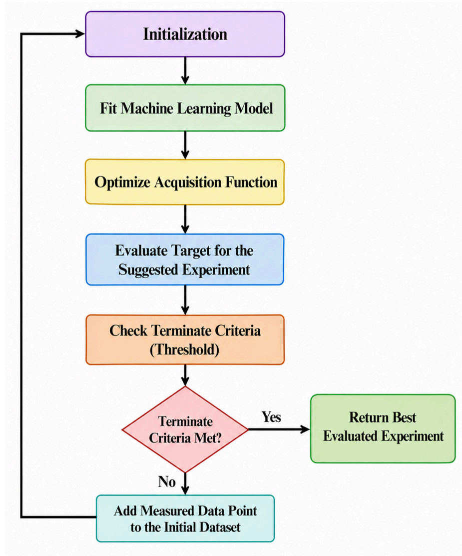

Figure 1. Workflow of Bayesian optimization

4. Results and discussion

Figure 2 summarizes the distribution of elongation at break for the full set of 98 cellulose-based fiber samples. The box plot shows a relatively broad spread of elongation values, with most samples clustered around intermediate elongations and only a few reaching the upper end of the acces- sible range. The corresponding histogram indicates that high-elongation fibers close to the dataset maximum (13.40%) are relatively rare. The three samples used to initialize the active-learning procedure are highlighted in Figure 2. as Initial training data.

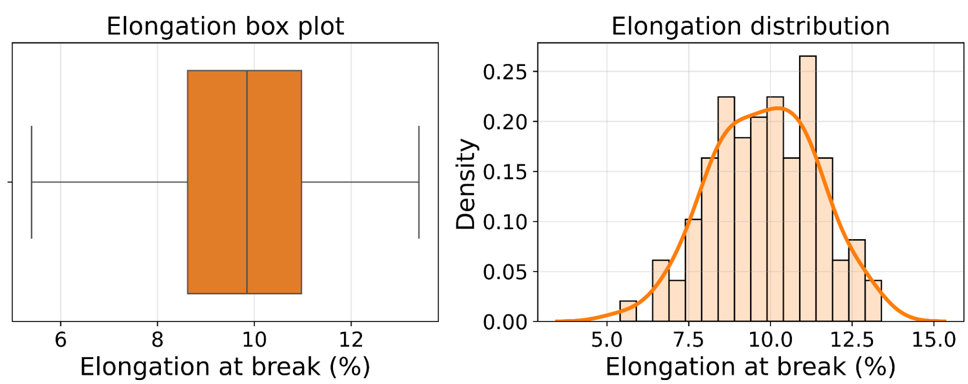

Figure 2. Distribution of elongation at break (%) represented by a box plot and histogram.

To evaluate the efficiency and robustness of the Bayesian-optimization strategy, we repeated the closed-loop workflow 10 times, each time starting from a different initial set of three samples drawn from the available experiments. In every run, a Gaussian-process surrogate model was trained on the currently available data, an expected-improvement acquisition function was optimized to select the next experiment, and the true elongation of the suggested fiber was revealed from the dataset and added to the training set. Figure 3. shows the evolution of the average actual elongation at break as a function of experiment number across these 10 BO runs. The blue curve corresponds to the mean elongation of the fibers selected by BO at each iteration, and the shaded region denotes one standard deviation across the repeated runs. The horizontal dashed line marks the maximum elongation in the dataset (13.40%), which is taken as the optimization target, while the orange curve indicates the elongation values of the three initial training samples. Several features are apparent. First, after initialization with three moderate-elongation samples, the BO policy quickly proposes fibers with elongation values that are competitive with, or higher than, the initial points. Second, the spread between different runs is relatively large in the early iterations, reflecting the combination of high model uncertainty and the different initial training sets. As more measurements are incorporated, the Gaussian-process model is updated and the predictive uncertainty decreases, leading to a progressive tightening of the shaded region and a steady increase in the mean elongation selected by the algorithm. On average, the BO workflow requires about 20 experiments to identify the sample with maximum elongation in the available dataset. This is substantially smaller than the total number of candidates (98), indicating a significant reduction in experimental effort relative to trial and error. The overall increase of the blue curve toward the dashed threshold line, together with the shrinking uncertainty band, suggests that the acquisition function effectively balances exploration of poorly characterized regions and exploitation of promising conditions throughout the optimization process.

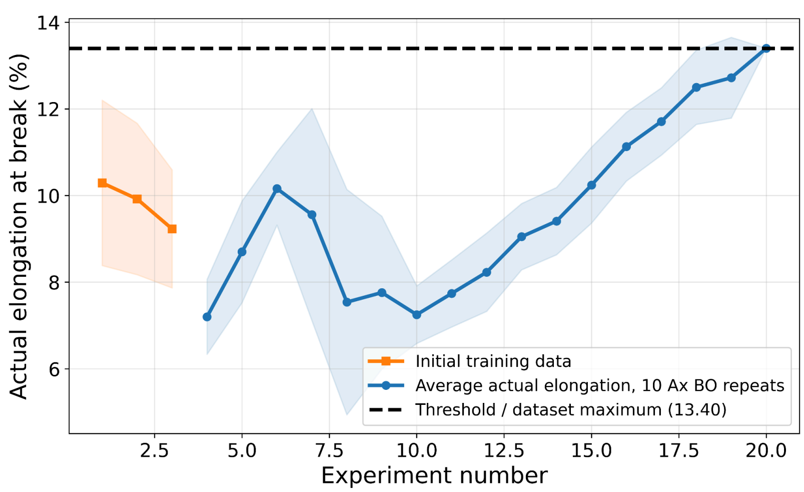

Figure 3. Evolution of actual elongation at break (%) during Bayesian optimization experiments.

From a process-design perspective, these results demonstrate that the proposed active-learning framework can locate high-elongation wet-spinning conditions using only a small fraction of the available experiments. The fact that the workflow performs reliably even when initialized with only three non-optimal fibers suggests that the method is well suited for realistic small-data scenarios, where only limited mechanical characterization is available. In the wet-spinning dataset, the maximum elongation-at-break value (13.40%) was first achieved only after 83 individual experiments. In contrast, the BO workflow identified a fiber at the same threshold after 20 iterations on average across ten independent runs (including the three initial training points). This corresponds to a reduction from 83 to 20 experiments, i.e. a factor of 83/20 ≈ 4.2 fewer experiments, or about a 76% reduction in experimental effort. In practical terms, if one wet-spinning and characterization cycle is taken as the unit of experimental time, BO saves roughly three out of every four experiments that would otherwise be required to reach the same elongation level. This gain arises because the Gaussian-process surrogate and the expected-improvement acquisition function continuously steer new experiments toward regions in parameter space that are predicted to be both promising and informative, instead of sampling conditions uniformly or at random.

5. Conclusion and outlook

We have demonstrated that Bayesian optimization with a Gaussian-process surrogate and an expected-improvement acquisition function can efficiently identify high-elongation wet-spinning conditions for cellulose-based fibers using only a small fraction of the available experiments. In a dataset of 98 cellulose fiber samples, high-elongation fibers close to the maximum (13.40%) are rare, and the three initial training samples used to start the active-learning procedure lie in the mid-range of the elongation distribution, far below the dataset maximum. Despite this realistic small-data, non-optimal starting scenario, the BO workflow identified a fiber at or above the target elongation threshold after an average of 20 iterations across ten independent runs, compared to 83 experiments required in the historical wet-spinning experiments. This corresponds to a reduction by a factor of approximately 4.2, or about a 76% reduction in experimental effort. The results show that the BO policy quickly proposes fibers with elongation values competitive with or higher than the initial points, and that the mean elongation generally increases as the Gaussian-process model is updated and predictive uncertainty decreases. The shrinking uncertainty band across iterations confirms that the acquisition function effectively balances exploration of poorly characterized regions with exploitation of promising conditions.

The present work is a proof-of-concept study in which the BO workflow was evaluated using virtual updates on an existing dataset. In future work, the same framework can be coupled directly to laboratory wet-spinning experiments, where the elongation of each recommended candidate would be measured experimentally and fed back into the optimization loop. This would enable true closed-loop active learning for fiber process development. Further extensions could include multi-objective Bayesian optimization to simultaneously optimize elongation together with other critical fiber properties such as tenacity, E-modulus, and titer, rather than focusing on elongation alone.

6. References

[1] C. Li, D. Rubín de Celis Leal, S. Rana, S. Gupta, A. Sutti, S. Greenhill, T. Slezak, M. Height, and S. Venkatesh. Rapid bayesian optimisation for synthesis of short polymer fiber materials. 8Scientific Reports, 7(1):5683, 2017.

[2] K. Moriam, D. Sawada, K. Nieminen, Y. Ma, M. Rissanen, N. Nygren, C. Guizani, M. Hum- mel, and H. Sixta. Spinneret geometry modulates the mechanical properties of man-made cellulose fibers. Cellulose, 28(17):11165–11181, 2021.

[3] M. Ypma, J. B. Westerink, H. Maatman, H. Boerstoel, and J. Veurink. Cellulose fibres and filaments having a high elongation at break, 2000.

[4] Y. Jin and P. V. Kumar. Bayesian optimisation for efficient material discovery: a mini review. Nanoscale, 15(15):10975–10984, 2023.

[5] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, and T. Lookman. Accelerated search for materials with targeted properties by adaptive design. Nature Communications, 7(1):11241, 2016.

[6] Turab Lookman, Prasanna V. Balachandran, Dezhen Xue, and Ruihao Yuan. Active learn- ing in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Computational Materials, 5:21, 2019.

[7] D. Rubín de Celis Leal, D. Nguyen, P. Vellanki, C. Li, S. Rana, N. Thompson, S. Gupta, K. Pringle, S. Subianto, S. Venkatesh, T. Slezak, M. Height, and A. Sutti. Efficient bayesian function optimization of evolving material manufacturing processes. ACS Omega, 4(24):20571–20578, 2019.

[8] K. Wang and A. W. Dowling. Bayesian optimization for chemical products and functional materials. Current Opinion in Chemical Engineering, 36:100728, 2022.

[9] H. Zhang, W. W. Chen, A. Iyer, D. W. Apley, and W. Chen. Uncertainty-aware mixed-variable machine learning for materials design. Scientific Reports, 12(1):19760, 2022.

[10] R. Zhang, J. Xu, H. Zhang, G. Xu, and T. Luo. Active learning-guided exploration of ther- mally conductive polymers under strain. Digital Discovery, 4(3):812–823, 2025.

[11] S. Jiang and M. A. Webb. Generative active learning across polymer architectures and solvo-phobicities for targeted rheological behavior. npj Computational Materials, 12(1):28, 2026.

[12] Y. Yao, B. J. Allardyce, R. Rajkhowa, D. Hegh, et al. Improving the tensile properties of wet spun silk fibers using rapid bayesian algorithm. ACS Biomaterials Science & Engineering, 6(5):3197–3207, 2020.

[13] J. R. Deneault, J. Chang, J. Myung, D. Hooper, A. Armstrong, M. Pitt, and B. Maruyama. Toward autonomous additive manufacturing: Bayesian optimization on a 3d printer. MRS Bulletin, 46:566–575, 2021.

[14] K. Park, Y. Kim, M. Kim, C. Song, J. Park, and S. Ryu. Designing staggered platelet com- posite structure with gaussian process regression based bayesian optimization. Composites Science and Technology, 220:109254, 2022.

[15] S. Pruksawan, G. Lambard, S. Samitsu, K. Sodeyama, and M. Naito. Prediction and opti- mization of epoxy adhesive strength from a small dataset through active learning. Science and Technology of Advanced Materials, 20:1010–1021, 2019.

[16] R. Q. Albuquerque, F. Rothenhäusler, and H. Ruckdäschel. Designing formulations of bio- based, multicomponent epoxy resin systems via machine learning. MRS Bulletin, 49:59–70, 2024.

[17] S. G. Baird, J. R. Hall, and T. D. Sparks. Compactness matters: Improving bayesian opti- mization efficiency of materials formulations through invariant search spaces. Computational Materials Science, 224:112134, 2023.

[18] Rodrigo Q. Albuquerque, Florian Rothenhäusler, Philipp Gröbel, and Holger Ruckdäschel. Multi-objective optimization of sustainable epoxy resin systems through bayesian optimiza- tion and machine learning. ACS Applied Engineering Materials, 1(12):3298–3308, 2023.

[19] Chiho Kim, Anand Chandrasekaran, Anurag Jha, and Rampi Ramprasad. Active-learning and materials design: the example of high glass transition temperature polymers. MRS Com- munications, 9(3):860–866, 2019.

[20] Rama Vasudevan, Ghanshyam Pilania, and Prasanna V. Balachandran. Machine learning for materials design and discovery. Journal of Applied Physics, 129(7):070401, 2021.