UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E ELETRÔNICA
CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Gustavo Schewinski

Machine Learning Surrogate-Based Lookahead Bayesian Optimization for Non-Isothermal Glass Molding

(The complete thesis, written in English, is included in the attachments)
(O trabalho completo, escrito em inglês, está incluído nos anexos)

Aachen, Alemanha
2024

Gustavo Schewinski

Machine Learning Surrogate-Based Lookahead Bayesian Optimization for Non-Isothermal Glass Molding

Trabalho de Conclusão de Curso do Curso de Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica e Eletrônica da Universidade Federal de Santa Catarina para a obtenção do título de Bacharel em Engenharia Elétrica.
Orientador Prof. Danilo Silva, Dr.
Coorientador Hendrik Mende, M.Sc.

Aachen, Alemanha
2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Schewinski, Gustavo
Machine Learning Surrogate-Based Lookahead Bayesian Optimization for Non-Isothermal Glass Molding / Gustavo Schewinski ; orientador, Danilo Silva, coorientador, Hendrik Mende, 2024.
92 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Engenharia Elétrica, Florianópolis, 2024.

Inclui referências.

1. Engenharia Elétrica. 2. Otimização Bayesiana Antecipada. 3. Otimização de processos. 4. Aprendizado de máquina. I. Silva, Danilo. II. Mende, Hendrik. III. Universidade Federal de Santa Catarina. Graduação em Engenharia Elétrica. IV. Título.

Gustavo Schewinski

Machine Learning Surrogate-Based Lookahead Bayesian Optimization for Non-Isothermal Glass Molding

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de Bacharel em Engenharia Elétricae aceito, em sua forma final, pelo Curso de Graduação em Curso de Graduação em Engenharia Elétrica.

Aachen, Alemanha, 28 de fevereiro de 2024.

Documento assinado digitalmente
Miguel Moreto
Data: 22/03/2024 08:56:42-0300
CPF: ***.850.100-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Miguel Moreto, Dr.
Coordenador do Curso

Banca Examinadora:

Documento assinado digitalmente
Danilo Silva
Data: 22/03/2024 07:25:29-0300
CPF: ***.557.924-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Danilo Silva, Dr.
Orientador

Documento assinado digitalmente
Richard Demo Souza
Data: 22/03/2024 07:27:00-0300
CPF: ***.267.379-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Richard Demo Souza, Dr.
Avaliador
Universidade Federal de Santa Catarina

Lars Leyendecker

Digitally signed by Lars Leyendecker
Date: 2024.03.20 11:18:31 +01'00'

Lars Leyendecker, M.Sc..
Avaliador
Fraunhofer-Institut für Produktionstechnologie IPT

RESUMO

Na indústria de manufatura, o desenvolvimento de novos produtos frequentemente requer a criação de processos complexos e caros. Esta tese visa criar um *framework* de Otimização Bayesiana Antecipada Multiobjetivo *(multiobjective Lookeahead Bayesian Optimization)* para otimizar o processo de moldagem de vidro não isotérmico (NGM) no Fraunhofer IPT. O método NGM é utilizado para a fabricação de lentes a partir de material de vidro maciço ou fino, moldando-o sob calor e força externa. A tecnologia é promissora por sua capacidade de formar formatos complexos de maneira econômica e eficiente em termos de energia. No entanto, devido à complexidade do processo de fabricação e aos inúmeros fatores que podem impactá-lo, identificar os parâmetros ideais para alcançar precisão exata no formato das lentes é desafiador. Métodos tradicionais para identificar os parâmetros ideais são caros e dependem fortemente de experimentação extensiva, consumo de material, possuem um extenso espaço de parâmetros e dependem do elemento de acaso. A Otimização Bayesiana visa abordar essa complexidade conduzindo experimentos sequenciais para encontrar os parâmetros ideais, reduzindo assim o número de experimentos necessários através de um raciocínio preditivo sobre os possíveis resultados de cada experimento e mantendo um equilíbrio entre *exploration* e *exploitation*. O estudo escolheu um Processo gaussiano para a Otimização Bayesiana como modelo substituto, e *Expected Hypervolume Improvement (qEHVI)* juntamente com *Decaying Prior-Weighted qEHVI* foram selecionados como funções de aquisição após um benchmark. Um modelo de aprendizado de máquina foi desenvolvido para avaliar o *framework* seguindo a metodologia CRISP-DM, atuando como um substituto para o processo real. Este modelo empregou um regressor Lasso e foi baseado em um conjunto de dados existente com muitos experimentos. O desenvolvimento do modelo de aprendizado de máquina envolveu etapas de pré-processamento de dados, extração, engenharia, limpeza, seleção de características e treinamento do modelo. Finalmente, o *framework* e o modelo substituto foram integrados para realizar a otimização. A otimização levou a uma melhoria significativa nos valores de pico a vale (PV), os quais indicam a precisão do formato, com um ligeiro aumento no desvio do formato (RMS). Esses resultados foram alcançados com uma redução de mais de 50% no número de experimentos em comparação com o conjunto de dados existente, e identificou com sucesso os parâmetros ideais do processo.

Palavras-chave: Otimização Bayesiana Antecipada. Otimização de processos. Aprendizado de máquina.

RESUMO

In the manufacturing industry, the development of new products often requires the creation of complex and expensive processes. This thesis aims to create a multiobjective Lookahead Bayesian Optimization (BO) framework to optimize the non-isothermal glass molding (NGM) process at Fraunhofer IPT. NGM is utilized for manufacturing glass optics from bulk or thin glass material by shaping the glass under heat and external force. The technology is promising for its ability to form complex shapes in a cost- and energy-efficient manner. However, due to the complexity of the manufacturing process and the numerous factors that can impact it, identifying the optimal parameters for achieving precise lens shape accuracy is challenging. Traditional methods to identify the optimum parameters are costly and rely heavily on extensive experimentation, material consumption, large parameter spaces, and an element of chance. Bayesian Optimization aims to address this complexity by conducting sequential experiments to find the optimal parameters, thereby reducing the number of required experiments through better predictive reasoning about future outcomes and maintaining a balance between exploration and exploitation. The study chose a Gaussian process for the BO framework as the surrogate model, and parallel Expected Hypervolume Improvement (qEHVI) along with Decaying Prior-Weighted qEHVI were selected as acquisition functions after a benchmark. A Machine Learning model was developed to evaluate the framework following the CRISP-DM methodology, serving as a surrogate for the real-world process. This model employed a Lasso regressor and was based on an existing dataset with a lot of experiments. The Machine Learning model's development involved stages of data preprocessing, feature extraction, engineering, cleaning, selection, and model training. Finally, the framework and the surrogate model were integrated to perform the optimization. The optimization led to a significant improvement in peak-to-valley (PV) values, indicating form accuracy, with a slight increase in shape deviation (RMS). These outcomes were achieved with more than a 50% reduction in the number of experimental tests compared to the existing dataset, and successfully identified the optimal process parameters.

Keywords: Lookahead Bayesian Optimization. Process Optimization. Machine Learning.

SUMÁRIO

1 INTRODUÇÃO

A tese desenvolve um *framework* de Otimização Bayesiana Antecipada (*Lookahead Bayesian Optimization*) para otimizar o processo de moldagem de vidro não isotérmico no Instituto Fraunhofer IPT. Ela aborda os desafios do processo de produção, visando diminuir o tempo e recursos gastos para a identificação dos parâmetros ideais para a produção das lentes. Além disso, um modelo de aprendizado de máquina foi utilizado para simular o processo real (como se fosse um modelo substitutivo) e, posteriormente, para a validação do *framework*, visto que o processo de avaliação e produção das lentes é demorado e depende da disponibilidade.

O trabalho foi realizado no Instituto Fraunhofer, Alemanha, dentro do departamento de produção inteligente. Ele abrange um estudo teórico de conceitos fundamentais para a otimização bayesiana, aprendizado de máquina, desenvolvimento do *framework* com funções de aquisição customizadas e modelos substitutos, modelo de aprendizado de máquina para validação, e realização da otimização utilizando tanto o *framework* quanto o modelo.

É importante ressaltar que a tese completa foi desenvolvida na língua inglesa e está disponível, na íntegra, no ANNEX A – ENGLISH THESIS. A presente versão em português é somente um resumo, ressaltando os principais pontos do trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo aborda o embasamento teórico essencial para a moldagem de vidro, amostragem, aprendizado de máquina e Otimização Bayesiana. Essas informações servem como uma base para entender os conceitos e metodologias empregados na tese.

2.1 MOLDAGEM DE VIDRO

2.1.1 Visão geral da produção de ópticas complexas

O vidro é cada vez mais utilizado em relação aos polímeros devido às suas propriedades superiores, como dureza e resistência a danos térmicos e por arranhões, melhorando a qualidade óptica e durabilidade (VU; GRUNWALD; BERGS, 2020; VU; KREILKAMP; DAMBON *et al.*, 2016). A fabricação tradicional de ópticas é ineficiente devido ao desperdício de material e alto consumo de energia. A Moldagem de Vidro Não Isotérmica (NGM) e a Moldagem de Vidro de Precisão (PGM) são dois métodos que abordam esses desafios, com a NGM oferecendo vantagens em tempo de processamento e viabilidade econômica para a produção de ópticas complexas de vidro (BLIEDTNER; GRÄFE, 2010; JIANG, C. *et al.*, 2022; VU; HELMIG *et al.*, 2020).

2.1.2 Moldagem de vidro não isotérmica

A NGM emprega uma técnica de formação a quente onde as temperaturas do vidro e do molde diferem, possibilitando tempos de processamento mais curtos e mantendo alta precisão (VU; KREILKAMP; KRISHNAMOORTHI *et al.*, 2016; NONISOTHERMAL..., 2016). Este método é distinguido da PGM pela separação das longas etapas de aquecimento e resfriamento do ciclo de moldagem, tornando-o uma opção econômica para a produção de componentes ópticos complexos (VU; GRUNWALD; BERGS, 2020).

2.2 AMOSTRAGEM

2.2.1 Latin Hypercube Sampling

O método *Latin Hypercube Sampling* (LHS) de amostragem é projetado para superar a amostragem aleatória simples, garantindo uma cobertura completa do intervalo das variáveis de entrada (MCKAY; BECKMAN; CONOVER, 1979; DEUTSCH; DEUTSCH, 2012). O LHS é popular por sua capacidade de cobrir efetivamente espaços de design pequenos e grandes e é adequado para otimizar designs para modelos sofisticados dentro de um número fixo de execuções (LOH, 1996; HELTON; DAVIS, 2003; VIANA, 2016).

2.3 APRENDIZADO DE MÁQUINA

2.3.1 Fundamentos do Aprendizado de Máquina

Aprendizado de máquina, um subconjunto da Inteligência Artificial (IA), é "a ciência (e arte) de programar computadores para que eles possam aprender a partir de dados"(GERON, 2019). Abrange vários domínios e é principalmente categorizado em aprendizado supervisionado, não supervisionado e por reforço (GERON, 2019). Suas aplicações abrangem vários domínios, incluindo robótica, jogos de computador, previsão de tráfego, diagnóstico médico e recomendação de produtos (RAY, 2019).

2.3.2 Framework CRISP-DM

O *framework* CRISP-DM define seis fases para projetos de mineração de dados, desde entender os objetivos de negócios até a implantação do modelo para benefícios comerciais (CHAPMAN, 2000; MARTÍNEZ-PLUMED *et al.*, 2021). Ele enfatiza o entendimento dos aspectos de negócios e dados, preparação dos dados, modelagem, avaliação e implantação, fornecendo uma abordagem estruturada para desenvolver soluções baseadas em dados.

2.4 OTIMIZAÇÃO BAYESIANA

A Otimização Bayesiana (BO) tem suas raízes em métodos estatísticos do início do século XX, evoluindo para uma ferramenta poderosa para otimizar funções desconhecidas ou *"black box"*. Ganhou impulso com a adoção de Processos Gaussianos (GPs) como modelos substitutos, introduzidos por Kushner (1962), e expandiu-se ainda mais em aplicações através do trabalho de Jones, Schonlau e Welch (1998). Apesar de sua eficácia em otimização multidimensional e com restrições, a BO enfrenta limitações em espaços de alta dimensão e sensibilidade a *outliers*.

Na Otimização Bayesiana, ocorre a melhoria iterativa do modelo substituto, principalmente usando GPs, e a seleção estratégica de pontos de observação por meio de funções de aquisição. Esse processo identifica eficientemente soluções ótimas, equilibrando a exploração de novas áreas e a explotação de regiões promissoras conhecidas, também conhecido como o equilíbrio entre *exploration* e *exploitation*. Funções de aquisição avançadas, como a *Expected Hypervolume Improvement (EHVI)*, são fundamentais para realçar a capacidade da BO, especialmente para tarefas de otimização multiobjetivo. EHVI foca em melhorar a frente de Pareto (em inglês, *Pareto front*), uma coleção de soluções ótimas em cenários multiobjetivo, maximizando o indicador de hipervolume que mede a qualidade dessas soluções.

Um avanço significativo na BO é a incorporação de estratégias de antecipação, que consideram os benefícios de longo prazo de observações potenciais em vez de tomar decisões

com base em ganhos imediatos. Essa abordagem, exemplificada por "Efficient Nonmyopic Bayesian Optimization via One-Shot Multi-Step Trees"por Shali Jiang *et al.* (2020), usa uma função de aquisição não míope que planeja vários passos à frente, otimizando a seleção de pontos de observação futuros. Esse método, que pode prever a utilidade de amostras futuras com mais precisão, representa um avanço significativo em tornar a BO mais eficaz para problemas de otimização complexos, onde o custo e o tempo dos experimentos são considerações críticas.

2.5 FUNÇÕES DE BENCHMARK

Funções de benchmark, como a família DTLZ e o problema de Segurança Veicular, são utilizadas para avaliar métodos de otimização.

3 METODOLOGIA

Este capítulo apresenta a metodologia para desenvolver um modelo de Otimização Bayesiana Antecipatória e Aprendizado de Máquina para otimizar o processo de fabricação de lentes. A abordagem começa com uma configuração experimental envolvendo amostragem LHS para projetar experimentos e coletar dados sobre a qualidade das lentes e os parâmetros do processo. Um modelo de regressão de aprendizado de máquina simula o processo de fabricação, avaliando a eficácia da Otimização Bayesiana Antecipatória para a otimização dos parâmetros do processo. A metodologia abrange a preparação de dados, processamento, extração e seleção de características para garantir a prontidão dos dados para o treinamento do modelo.

3.1 CONFIGURAÇÃO EXPERIMENTAL

Para a coleta e preparação de dados, os experimentos são projetados usando amostragem LHS, focando na coleta de dados de qualidade das lentes e dados dos sensores do processo. Os experimentos visam cobrir o espaço de parâmetros efetivamente, garantindo uma cobertura representativa para o modelo de aprendizado de máquina.

3.2 PIPELINE DE APRENDIZADO DE MÁQUINA

A pipeline de aprendizado de máquina envolve o pré-processamento dos dados de entrada e objetivo, incluindo extração, engenharia e limpeza de características. Logo após, inclui a seleção de características significativas, treinamento do modelo com regressão Lasso e ajuste de hiperparâmetros para minimizar erros de predição, garantindo treinamento eficaz do modelo.

3.3 FRAMEWORK DE OTIMIZAÇÃO BAYESIANA ANTECIPATÓRIA

O *framework* de Otimização Bayesiana Antecipatória emprega um regressor de Processo Gaussiano e várias funções de aquisição, incluindo funções personalizadas desenvolvidas para incorporar conhecimento especializado no processo de otimização. Testes de *benchmark* avaliam diferentes funções de aquisição e passos antecipatórios, determinando estratégias eficazes para a otimização do processo real. O *framework* visa selecionar experimentos altamente informativos para o processo de fabricação de lentes durante a otimização, maximizando os *insights* sobre a função desconhecida representada pelo processo real de fabricação.

4 RESULTADOS E DISCUSSÃO

O capítulo apresenta os resultados da integração da Otimização Bayesiana Antecipatória com um modelo de Aprendizado de Máquina de Regressão Lasso para otimizar o processo de fabricação de lentes.

4.1 MODELO DE APRENDIZADO DE MÁQUINA

O modelo de aprendizado de máquina alcançou Erros Absolutos Médios (MAEs) para valores de pico a vale (PV) e desvio no formato (RMS) significativamente inferiores ao objetivo de 80 micrômetros, indicando um excelente desempenho. A boa generalização do modelo para dados não vistos foi destacada pela semelhança das métricas de erro entre os conjuntos de validação e teste.

4.2 INTEGRAÇÃO DA OTIMIZAÇÃO BAYESIANA ANTECIPATÓRIA E MODELO DE APRENDIZADO DE MÁQUINA

A Otimização Bayesiana Antecipatória, integrada ao modelo substituto de aprendizado de máquina, foi executada em 20 iterações, cada uma avaliando 10 candidatos. Essa configuração, totalizando 220 experimentos, teve como objetivo minimizar tanto os valores de PV quanto de RMS. A otimização alcançou com sucesso uma frente de Pareto abrangente, indicando um conjunto diversificado de soluções ótimas com valores mínimos de PV e RMS de menos de 32,5 micrômetros e 4,5 micrômetros, respectivamente.

Uma comparação com os melhores resultados de 500 experimentos usando amostragem LHS mostrou que a Otimização Bayesiana melhorou significativamente os valores de PV, embora os valores de RMS tenham aumentado ligeiramente. Notavelmente, essas melhorias foram alcançadas com menos experimentos do que a LHS.

5 CONCLUSÃO

Este capítulo demonstra que o framework de Otimização Bayesiana Antecipatória, proposto nesta tese, melhora a fabricação de lentes de vidro no Fraunhofer IPT, requerendo menos experimentos. Um modelo de regressão Lasso, treinado com dados da amostragem LHS, simula o processo com alta precisão, atendendo aos padrões de desempenho. A eficácia é confirmada por testes de benchmark e pela simulação do processo real usando Processo Gaussiano e funções de aquisição específicas, como o qDecayWeightedEHVI. Isso resultou em melhorias significativas nos valores de pico a vale (PV), com um pequeno aumento nos valores RMS, em comparação com dados prévios. A otimização efetiva com menos experimentos evidencia o sucesso da tese.

REFERÊNCIAS

BLIEDTNER, Jens; GRÄFE, Günter. **Optiktechnologie: Grundlagen – Verfahren – Anwendungen – Beispiele**. 2., aktualisierte Auflage. [*S.l.*]: Carl Hanser Verlag GmbH & Co. KG, 2010. P. 419. ISBN 978-3-446-42215-5. DOI: 10.3139/9783446424661.

CHAPMAN, Peter. CRISP-DM 1.0: Step-by-step data mining guide. *In:* Available from: https://api.semanticscholar.org/CorpusID:59777418.

DEUTSCH, Jared L.; DEUTSCH, Clayton V. Latin hypercube sampling with multidimensional uniformity. **Journal of Statistical Planning and Inference**, v. 142, n. 3, p. 763–772, 2012. ISSN 0378-3758. DOI: https://doi.org/10.1016/j.jspi.2011.09.016. Available from: https://www.sciencedirect.com/science/article/pii/S0378375811003776.

GERON, Aurelien. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 2nd. [*S.l.*]: O'Reilly Media, Inc., 2019. ISBN 1492032646.

HELTON, J.C.; DAVIS, F.J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. **Reliability Engineering System Safety**, v. 81, n. 1, p. 23–69, 2003. ISSN 0951-8320. DOI: https://doi.org/10.1016/S0951-8320(03)00058-9. Available from: https://www.sciencedirect.com/science/article/pii/S0951832003000589.

JIANG, Cheng *et al.* Simulation of the Refractive Index Variation and Validation of the Form Deviation in Precisely Molded Chalcogenide Glass Lenses (IRG 26) Considering the Stress and Structure Relaxation. **Materials**, v. 15, n. 19, 2022. ISSN 1996-1944. DOI: 10.3390/ma15196756. Available from: https://www.mdpi.com/1996-1944/15/19/6756.

JIANG, Shali *et al.* **Efficient Nonmyopic Bayesian Optimization via One-Shot Multi-Step Trees**. [*S.l.*: *s.n.*], 2020. arXiv: 2006.15779 [cs.LG].

JONES, Donald R.; SCHONLAU, Matthias; WELCH, William J. Efficient Global Optimization of Expensive Black-Box Functions. **Journal of Global Optimization**, v. 13, n. 4, p. 455–492, dez. 1998. ISSN 1573-2916. DOI: 10.1023/A:1008306431147. Available from: https://doi.org/10.1023/A:1008306431147.

KUSHNER, Harold J. A versatile stochastic model of a function of unknown and time varying form. **Journal of Mathematical Analysis and Applications**, v. 5, n. 1, p. 150–167, 1962. ISSN 0022-247X. DOI: https://doi.org/10.1016/0022-247X(62)90011-2. Available from: https://www.sciencedirect.com/science/article/pii/0022247X62900112.

LOH, Wei-Liem. On Latin hypercube sampling. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 24, n. 5, p. 2058–2080, 1996. DOI: 10.1214/aos/1069362310. Available from: https://doi.org/10.1214/aos/1069362310.

MARTÍNEZ-PLUMED, Fernando *et al.* CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. **IEEE Transactions on Knowledge and Data Engineering**, v. 33, n. 8, p. 3048–3061, 2021. DOI: 10.1109/TKDE.2019.2962680.

MCKAY, M. D.; BECKMAN, R. J.; CONOVER, W. J. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. **Technometrics**, [Taylor Francis, Ltd., American Statistical Association, American Society for Quality], v. 21, n. 2, p. 239–245, 1979. ISSN 00401706. Available from: http://www.jstor.org/stable/1268522. Acesso em: 19 dez. 2023.

NONISOTHERMAL glass molding for the cost-efficient production of precision freeform optics. [*S.l.*: *s.n.*], 2016. Accessed on 2024-01-28. DOI: 10.1117/1.OE.55.7.071207. Available from: https://publica.fraunhofer.de/handle/publica/246935.

RAY, Susmita. A Quick Review of Machine Learning Algorithms. *In:* 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). [*S.l.*: *s.n.*], 2019. P. 35–39. DOI: 10.1109/COMITCon.2019.8862451.

VIANA, Felipe A. C. A Tutorial on Latin Hypercube Design of Experiments. **Quality and Reliability Engineering International**, v. 32, n. 5, p. 1975–1985, 2016. DOI: https://doi.org/10.1002/qre.1924. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qre.1924. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.1924.

VU, Anh Tuan; GRUNWALD, Tim; BERGS, Thomas. Thermo-viscoelastic Modeling of Nonequilibrium Material Behavior of Glass in Nonisothermal Glass Molding. **Procedia Manufacturing**, v. 47, p. 1561–1568, 2020. 23rd International Conference on Material Forming. ISSN 2351-9789. DOI: https://doi.org/10.1016/j.promfg.2020.04.350. Available from: https://www.sciencedirect.com/science/article/pii/S2351978920314177.

VU, Anh Tuan; HELMIG, Thorsten *et al.* Numerical and experimental determinations of contact heat transfer coefficients in nonisothermal glass molding. **Journal of the American Ceramic Society**, v. 103, n. 2, p. 1258–1269, 2020. DOI: https://doi.org/10.1111/jace.16756. eprint: https://ceramics.onlinelibrary.wiley.com/doi/pdf/10.1111/jace.16756. Available from: https://ceramics.onlinelibrary.wiley.com/doi/abs/10.1111/jace.16756.

VU, Anh-Tuan; KREILKAMP, Holger; DAMBON, Olaf *et al.* Nonisothermal glass molding for the cost-efficient production of precision freeform optics. **Optical**

Engineering, SPIE, v. 55, n. 7, p. 071207, 2016. DOI: 10.1117/1.OE.55.7.071207. Available from: https://doi.org/10.1117/1.OE.55.7.071207.

VU, Anh-Tuan; KREILKAMP, Holger; KRISHNAMOORTHI, Bharathwaj Janaki *et al.* A hybrid optimization approach in non-isothermal glass molding. **AIP Conference Proceedings**, v. 1769, n. 1, p. 040006, out. 2016. ISSN 0094-243X. DOI: 10.1063/1.4963428. eprint: https://pubs.aip.org/aip/acp/article-pdf/doi/10.1063/1.4963428/13726112/040006_1_online.pdf. Available from: https://doi.org/10.1063/1.4963428.

APÊNDICE A – VERSÃO EM INGLÊS

FEDERAL UNIVERSITY OF SANTA CATARINA
ELECTRICAL AND ELETRONIC ENGINEER DEPARTMENT
UNDERGRADUATE COURSE IN ELECTRICAL ENGINEERING

Gustavo Schewinski

Machine Learning Surrogate-Based Lookahead Bayesian Optimization for Non-Isothermal Glass Molding

Aachen, Germany
2024

Gustavo Schewinski

Machine Learning Surrogate-Based Lookahead Bayesian Optimization for Non-Isothermal Glass Molding

Undergraduate Course Final Project in ELECTRICAL ENGINEERING from the ELECTRICAL AND ELECTRONIC ENGINEERING DEPARTMENT of the Federal University of Santa Catarina for obtaining the degree of bachelor in Electrical Engineering.
Supervisor: Prof. Danilo Silva, Dr.
Co-supervisor: Hendrik Mendes, M.Sc.

Aachen, Germany
2024

ACKNOWLEDGEMENTS

I would like to express my gratitude to my parents, Alceu and Mônia, and my sister, Selena, for all their support and encouragement through my life and academic journey. Your presence was fundamental, and without you, none of this would have been meaningful.

I am also grateful for the opportunity I had during my internship at Fraunhofer Institute for Production Technology IPT. A special thanks to my internship supervisor, Maximilian Brochhaus, for his mentorship and support throughout the year.

To Hendrik Mende, my thesis advisor at IPT, thank you for your invaluable support and guidance during the development of this project; it has been an enriching experience.

To Danilo Silva, my academic advisor from UFSC, thank you for your contribution and valuable feedback on this work.

I am also grateful to the friends I made during my exchange year in Germany. Meeting each of you was a fantastic experience that I will always remember. Additionally, I would like to thank all my friends from the Hiwi room on the 4th floor for making my days funnier and more enjoyable, including Davi, Fabian, Fernando, Laura, Luiza, Pedro, Rafael, Sharaang, and Victor.

ABSTRACT

In the manufacturing industry, the development of new products often requires the creation of complex and expensive processes. This thesis aims to create a multiobjective Lookahead Bayesian Optimization (BO) framework to optimize the non-isothermal glass molding (NGM) process at Fraunhofer IPT. NGM is utilized for manufacturing glass optics from bulk or thin glass material by shaping the glass under heat and external force. The technology is promising for its ability to form complex shapes in a cost- and energy-efficient manner. However, due to the complexity of the manufacturing process and the numerous factors that can impact it, identifying the optimal parameters for achieving precise lens shape accuracy is challenging. Traditional methods to identify the optimum parameters are costly and rely heavily on extensive experimentation, material consumption, large parameter spaces, and an element of chance. Bayesian Optimization aims to address this complexity by conducting sequential experiments to find the optimal parameters, thereby reducing the number of required experiments through better predictive reasoning about future outcomes and maintaining a balance between exploration and exploitation. The study chose a Gaussian process for the BO framework as the surrogate model, and parallel Expected Hypervolume Improvement (qEHVI) along with Decaying Prior-Weighted qEHVI were selected as acquisition functions after a benchmark. A Machine Learning model was developed to evaluate the framework following the CRISP-DM methodology, serving as a surrogate for the real-world process. This model employed a Lasso regressor and was based on an existing dataset with a lot of experiments. The Machine Learning model's development involved stages of data preprocessing, feature extraction, engineering, cleaning, selection, and model training. Finally, the framework and the surrogate model were integrated to perform the optimization. The optimization led to a significant improvement in peak-to-valley (PV) values, indicating form accuracy, with a slight increase in shape deviation (RMS). These outcomes were achieved with more than a 50% reduction in the number of experimental tests compared to the existing dataset, and successfully identified the optimal process parameters.

Keywords: Lookahead Bayesian Optimization. Process Optimization. Machine Learning.

RESUMO

Na indústria de manufatura, o desenvolvimento de novos produtos frequentemente requer a criação de processos complexos e caros. Esta tese visa criar um *framework* de Otimização Bayesiana Antecipada *(Lookeahead Bayesian Optimization)* para otimizar o processo de moldagem de vidro não isotérmico (NGM) no Fraunhofer IPT. O método NGM é utilizado para a fabricação de lentes a partir de material de vidro maciço ou fino, moldando-o sob calor e força externa. A tecnologia é promissora por sua capacidade de formar formatos complexos de maneira econômica e eficiente em termos de energia. No entanto, devido à complexidade do processo de fabricação e aos inúmeros fatores que podem impactá-lo, identificar os parâmetros ideais para alcançar precisão exata no formato das lentes é desafiador. Métodos tradicionais para identificar os parâmetros ideais são caros e dependem fortemente de experimentação extensiva, consumo de material, possuem um extenso espaço de parâmetros e dependem do elemento de acaso. A Otimização Bayesiana visa abordar essa complexidade conduzindo experimentos sequenciais para encontrar os parâmetros ideais, reduzindo assim o número de experimentos necessários através de um raciocínio preditivo sobre os possíveis resultados de cada experimento e mantendo um equilíbrio entre *exploration* e *exploitation*. O estudo escolheu um Processo gaussiano para a Otimização Bayesiana como modelo substituto, e *Expected Hypervolume Improvement (qEHVI)* juntamente com *Decaying Prior-Weighted qEHVI* foram selecionados como funções de aquisição após um benchmark. Um modelo de aprendizado de máquina foi desenvolvido para avaliar o *framework* seguindo a metodologia CRISP-DM, atuando como um substituto para o processo real. Este modelo empregou um regressor Lasso e foi baseado em um conjunto de dados existente com muitos experimentos. O desenvolvimento do modelo de aprendizado de máquina envolveu etapas de pré-processamento de dados, extração, engenharia, limpeza, seleção de características e treinamento do modelo. Finalmente, o *framework* e o modelo substituto foram integrados para realizar a otimização. A otimização levou a uma melhoria significativa nos valores de pico a vale (PV), os quais indicam a precisão do formato, com um ligeiro aumento no desvio do formato (RMS). Esses resultados foram alcançados com uma redução de mais de 50% no número de experimentos em comparação com o conjunto de dados existente, e identificou com sucesso os parâmetros ideais do processo.

Palavras-chave: Otimização Bayesiana Antecipada. Otimização de processos. Aprendizado de máquina.

LIST OF FIGURES

LISTINGS

LIST OF TABLES

CONTENTS

1 INTRODUCTION

In the field of engineering, the process of developing a material or product is often time-consuming, traditionally relying on a trial-and-error approach. This method can significantly prolong the time required to identify optimal parameters. (LIU, 2017)

Bayesian Optimization has become popular in engineering due to its potential to accelerate the process by intelligently selecting the right experiments sequentially. This approach aims to optimize an otherwise complex and poorly understood function, often referred to as a "black box" function (GREENHILL et al., 2020). The objective of this thesis is to optimize a real production process of optical lenses using a Looakhead Bayesian Optimization approach and assess its performance with a Machine Learning model, serving as a surrogate for the actual experiment.

This initial chapter presents the motivation and context of the thesis, followed by an introduction to the project's background. It then outlines the objectives of the bachelor thesis, including a description of the work's scope, and concludes with an overview of the document's structure.

1.1 MOTIVATION AND CONTEXT

The demand for glass lenses has been steadily increasing, driven by their extensive application across various industries and in numerous modern technologies. Glass lenses are integral components in diverse optical systems, such as camera optics, LED lighting, sensor devices, and interior components in automobiles (ZHU et al., 2012; VU; HELMIG, et al., 2020).

Despite the widespread use, the production process remains costly and involves extended experimentation cycles, This is primarily due to the limited understanding of how different steps in the process influence model glass to shrink, as well as the behavior of glass in terms of heat, flow, and physical changes (VU; KREILKAMP; KRISHNAMOORTHI, et al., 2016). This complexity poses significant obstacles in comprehending how these parameters affect the various phenomena involved in glass molding.

Traditionally, non-isothermal glass molding (NGM) technology has been used to manufacture simple products without high requirements, such as drinking cups, ashtrays, and salad bowls. However, at IPT, there is an ambition to push the NGM beyond its conventional applications, aiming to achieve high precision and energy-efficient manufacturing with small shape deviation, potentially revolutionizing the industry with a cost-effective and highly accurate production method.

These challenges can be effectively addressed through the application of Bayesian Optimization. This method progressively refines the production process and significantly reduces the number of necessary experiments, demonstrating a sample-efficient approach. It is utilized in various "black box" functions encountered in areas such as AI, engineering,

and material science, thus becoming an increasingly attractive methodology for accelerated materials research and optimizing material properties (SOLOMOU et al., 2018).

1.2 PROJECT CONTEXT

The thesis project was conducted at Fraunhofer IPT during a one-year internship program, *BRAACHEN*. The necessary materials, machinery, and data were provided by the institute.

1.2.1 Institute

1.2.1.1 Fraunhofer-Gesellschaft

The *Fraunhofer-Gesellschaft* is the world's leading applied research organization. It is based in Germany and is deeply committed to the development of new technologies through interdisciplinary research collaborations with partners and companies from across the globe. Nowadays it operates with more than 80 research facilities, including 76 institutes, and more than 30.000 employees. (FRAUNHOFER, 2023c)

The name of the *Gesellschaft* was named after Joseph Von Fraunhofer, a successful researcher, inventor, and entrepreneur, who was therefore chosen as the role model and namesake of the organization. He is considered one of the founders of modern optics due to his contributions to the development of new glass production and processing methods, as well as manufacturing the best telescopes of his time. He was recognized as the first German exponent of precision optics, and for his contribution to the research on the spectral composition of light. (FRAUNHOFER, 2023a)

Established in Munich on March 26^{th}, 1949, the organization has a rich history of innovation, with several notable achievements. One of the most renowned Fraunhofer developments is the MP3 audio data compression process, which has garnered widespread recognition over the years (FRAUNHOFER, 2023b).

1.2.1.2 Fraunhofer IPT

The Fraunhofer Institute for Production Technology IPT is based on the premises of the RWTH Aachen campus in Aachen, focusing on the development of systems solutions for the sustainable, resilient, and digitalized production of resource-saving products. (IPT, 2024)

1.2.1.3 The Production Quality Department

The internship was realized within the Production Quality department, which places a strong emphasis on consistent digitalization and cross-linking of production data, as well as maximizing resource efficiency. During the internship, one of the focuses was on

projects in Machine Learning and Data Science, with an emphasis on applications relevant to the manufacturing industry.

1.3 OBJECTIVE OF THE THESIS

1.3.1 General objective

The bachelor thesis aims to develop a Lookahead Bayesian Optimization framework to optimize the production process of glass lenses. This method has distinct advantages over traditional BO approaches. One of the key benefits of Lookahead BO is its ability to anticipate the probable outcomes of future tests. By its anticipation, the framework can selectively focus on those that are most likely to yield significant improvement, particularly in scenarios where the number of experiments that can be conducted is limited. Another benefit is the evaluation of multiple candidates in each iteration, allowing for the planning of an entire set of experiments in a single optimization step. This is particularly valuable as it saves time in the measurement and analysis phases required before proceeding to the next set of candidates.

This framework involves integrating and selecting a Gaussian processes and an Acquisition Functions to streamline the optimization process. As well as building a Machine Learning model to simulate real experiments and assess the performance of this optimization technique. The ultimate goal is to improve shape deviation beyond what was achieved in prior experiments.

The decision to build a Machine Learning model for simulating the actual process is driven by practical considerations, such as providing flexibility for future experiments and projects that extend beyond the scope of this thesis. Furthermore, the limited timeframe for project development and the availability of machinery were key influencing factors. Using this model provides a viable way to perform a comprehensive evaluation of the optimization technique.

One of the major challenges of this work is ensuring that each component of the framework functions efficiently and cohesively. This requires a lot of extensive research, exploration, and development. Additionally, the final framework should be user-friendly, facilitating its adoption in future experiments at Fraunhofer IPT.

1.3.2 Specific objectives

The specific objectives were outlined based on the main ones, and aim to define an optimization that satisfies all requirements:

- Study the literature about Lookahead Bayesian Optimization and Machine Learning and how they are applied.

- Develop an efficient Lookahead Bayesian Optimization pipeline, validated through benchmark testing, which includes a Surrogate Model and an Acquisition Function.
- Implementation of the Lookahead Bayesian Optimization framework into code.
- Develop a Machine Learning model to serve as a surrogate for validation, accurately mimicking the production of real lenses.
- Realization of the optimization process and validation of the results.

It's important to note that this thesis employs two distinct surrogate models. The first is integrated into the Lookahead Bayesian Optimization framework, aiding in the optimization process. The second surrogate developed independently, validates the framework by mimicking real-world lens production.

1.4 DOCUMENT OUTLINE

This document is organized into five chapters, detailing the development of a Lookahead Bayesian Optimization framework and its assessment through a Machine Learning model.

Chapter 1 provides important information about the project, including its motivation and context. It also outlines the objectives and scope of the student's work.

Chapter 2 highlights the theoretical background necessary to understand the key concepts utilized in this thesis. These concepts include the fundamentals of glass molding, sampling techniques, Bayesian optimization, and Machine Learning.

Chapter 3 focuses on the methodology. It describes how the experimental data was generated for the Machine Learning model, the steps involved in developing the model, and the creation of the Lookahead Bayesian Optimization framework.

Chapter 4 is dedicated to presenting the results of the Machine Learning model and evaluating the performance of the Lookahead Bayesian Optimization.

Finally, **Chapter 5** concludes the thesis by discussing the results and assessing the initial objectives of the work. This chapter also suggests future research direction in modeling the lens manufacturing process and advancing Lookahead Bayesian Optimization in production settings.

2 THEORETICAL BACKGROUND

This chapter outlines the essential theoretical background of glass molding, sampling, machine learning, and Bayesian Optimization. This information serves as a foundation for understanding the concepts and methodologies employed in the thesis work.

2.1 GLASS MOLDING

2.1.1 Overview of the production of complex optics

Nowadays, glass is increasingly replacing polymers in many applications. This is predominantly due to its superior hardness, enhanced thermal and scratch resistance which together increase optical quality under any condition and extend durability (VU; GRUNWALD; BERGS, 2020; VU; KREILKAMP; DAMBON, et al., 2016).

The market of LED-based products, encompassing automotive lighting, street illumination, and sensor components, is projected to grow at a Compound Annual Growth Rate (CAGR) of 11% from 2023 to 2030, as reported by Grand View Research (2023). This growth represents a rising demand for these products in the coming years, being important to be cost-efficient.

The conventional optics manufacturing is very old and remained the same for years. These lenses pass through a process of grinding and polishing, leading to significant material waste and high energy consumption, rendering them inefficient for large-scale production. (BLIEDTNER; GRÄFE, 2010)

Furthermore, the industry's demand for complex shapes necessitates a costly and challenging manufacturing process with high requirements for form accuracy and surface roughness. To address these demands, two primary methods are utilized: Precision Glass Molding (PGM) and the focus of this thesis, Nonisothermal Glass Molding (NGM).

PGM is a specialized isothermal technique wherein the glass temperature is maintained uniformly during the pressing process (JIANG, C. et al., 2022). This method is utilized to produce highly accurate optical components, such as lenses for digital cameras and smartphones, benefiting from improved form and surface precision. However, PGM requires advanced system technology, expensive high-precision equipment, and lengthy processing time (VU; HELMIG, et al., 2020).

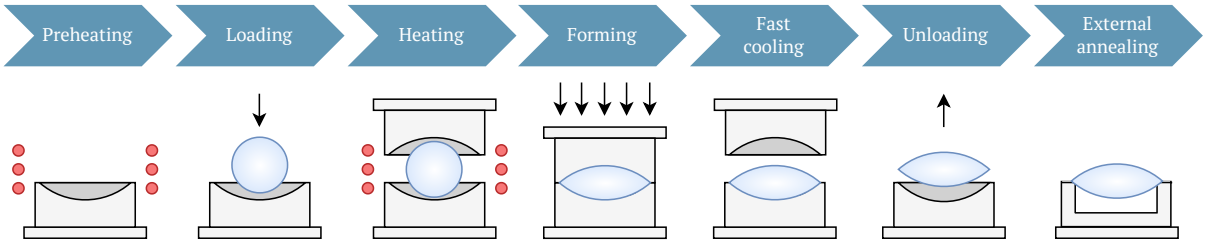
2.1.2 Non-isothermal glass molding

Non-isothermal glass molding, the method utilized in this thesis, employs a replication-based hot forming technique for manufacturing. This method allows the temperatures of the glass and the molding tools to differ during the process. This technique allows the process to reduce the processing time per unit molded part, while still maintaining high precision and good surface quality of the molded optics as reported

in Vu, Kreilkamp, Krishnamoorthi, et al. (2016). This short processing time is possible because it differs from the PGM method in the separation of the long heating and cooling stages from the molding cycle (NONISOTHERMAL..., 2016). Additionally, it is also an economically advantageous method for producing complex glass optics (VU; GRUNWALD; BERGS, 2020), as it doesn't need grinding and polishing.

The NGM process can be outlined in 10 steps, as depicted in Figure 1 and described in Mende et al. (2023). In summary, the glass lens production process begins with heating the furnace. The lens is then placed into the mold tool, and both are positioned inside the furnace. During the homogenizing phase inside the furnace, the temperature within the mold and the glass is evenly distributed. After this phase, the mold and glass are removed from the furnace and moved to the forming stage. Here, the machine applies a pressing force to bring the two halves of the mold together. Finally, the lens undergoes rapid cooling, is unloaded from the mold, and then subjected to external annealing.

Figure 1 – Non-isothermal glass forming process.



Source: Author, based on Fiorini (2023).

2.2 SAMPLING

In this work, the primary objective of designing samples is to carefully select a limited number of experiments for execution in the actual manufacturing process. The goal is to ensure that these experiments capture all essential information and accurately represent the system. This sampling will form the basis for creating the machine learning model, and ensure good generalization capabilities. The chosen method must minimize the inclusion of non-informative simulation points, thereby enhancing the model's representatives of the real-world process (LEVY; ADAMS; STEINBERG, 2010).

2.2.1 Latin Hypercube Sampling

The Latin Hypercube Sampling (LHS) was created with the objective of outperforming other methods such as random sampling (SRS) by ensuring full coverage of the range of the input variables (MCKAY; BECKMAN; CONOVER, 1979). It is a technique that focuses on uniform sampling of the univariate distributions by stratifying the cumulative

distribution function and randomly sampling within the strata (DEUTSCH; DEUTSCH, 2012).

This method is one of the most popular in research (LOH, 1996), and it is both theoretically and experimentally proven that LHS is better than random sampling (HELTON; DAVIS, 2003). Additionally, as mentioned in (VIANA, 2016), LHS offers several benefits that make it a preferred choice in the design of experiments:

- LHS is capable of covering both small and large design spaces effectively.
- It is suitable for optimizing designs for sophisticated models within a fixed number of runs.
- The method is flexible because the dimensions of the design space are independent.

It works by dividing the subspace into equally probable intervals, wherein some points are chosen to have good coverage. It creates the samples by randomly combining the chosen points without replacement, which won't be chosen again in the next iteration. Repeatedly doing this iteration, LHS would produce a set of samples that are representative of the real variability (SHIELDS; ZHANG, 2016).

2.3 MACHINE LEARNING

2.3.1 Fundamentals of Machine Learning

Machine learning, a subset of Artificial Intelligence (AI), is "the science (and art) of programming computers so they can learn from data" (GERON, 2019). Nowadays, its applications span various domains, including robotics, computer games, traffic prediction, medical diagnosis, and product recommendation (RAY, 2019).

It is primarily categorized into three types: supervised, unsupervised, and reinforcement learning, though they can be combined when needed.

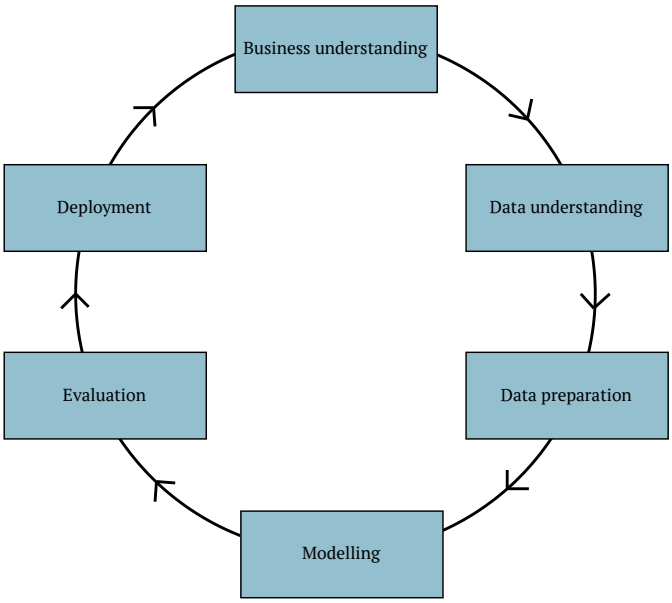
- Supervised Learning: This occurs when the training data includes the target solutions or labels (GERON, 2019). It encompasses two tasks: classification for categorical target data and regression for numerical target data.
- Unsupervised Learning: This takes place when the training data don't have the target solutions or labels.
- Reinforcement Learning: Here, a model (or agent) learns over time to develop an optimal strategy (policy) for maximizing rewards (GERON, 2019). The learning process involves receiving rewards or penalties based on its actions.

2.3.2 CRISP-DM Framework

The CRISP-DM (Cross Industry Standard Process for Data Mining) framework, introduced by Chapman (2000), guides and documents all the key steps in data mining projects (MARTÍNEZ-PLUMED et al., 2021). In recent years, it has gained popularity in data science for deriving insights from data. It is commonly employed in AI and machine learning projects, where the goal is to develop high-quality, data-driven solutions.

The framework has six phases, outlined below based on Chapman (2000), and illustrated in Figure 2.

Figure 2 – CRISP-DM steps.



Source: Author, based on Chapman (2000).

1. Business understanding: This initial step focuses on understanding the project's objectives and requirements from a business perspective.

2. Data understanding: It involves data collection, exploratory analysis, identifying quality issues, and having first insights.

3. Data preparation: This stage finalizes the dataset, including data cleaning and transformations.

4. Modeling: It entails creating and refining the model to achieve the desired performance.

5. Evaluation: The model undergoes a business evaluation to ensure it meets the specified requirements.

6. Deployment: The final phase, where the model is utilized for business benefits, either as a report or a software component.

Regarding the CRISP-DM framework, two steps are particularly relevant to this thesis and will have their theoretical background outlined: modeling, and evaluation.

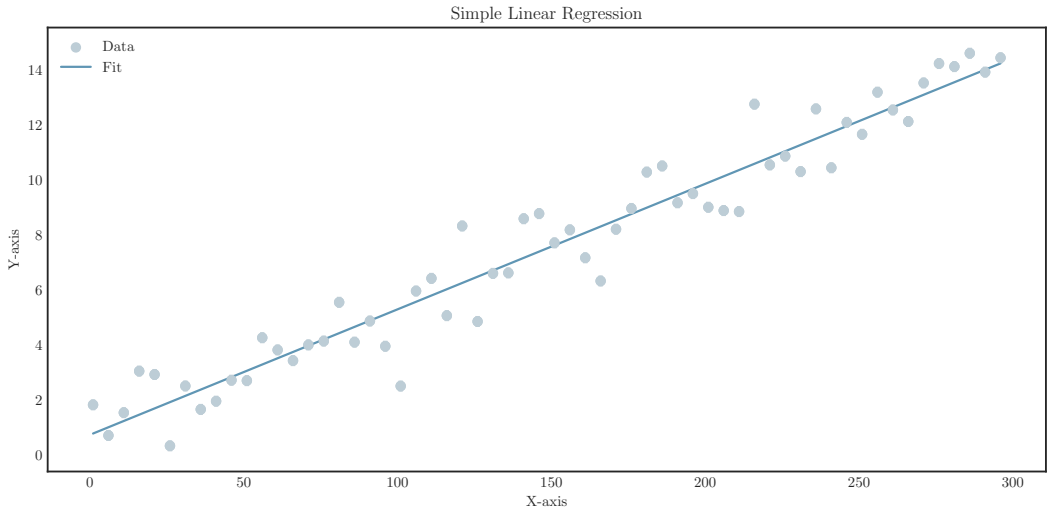
2.3.2.1 Modelling: linear regression

Linear regression is one of the simplest ways to determine the relationship between two continuous variables. It is widely used for predictive analysis methods when the outcomes of prediction can be quantified and modeled in connection to several independent variables (LIM, 2019). There are two types: simple and multiple regression.

Simple regression, depicted in Figure 3, analyses the relationship between a dependable y variable and an independent variable x (HOPE, 2020). The relationship is expressed in Equation 1, where C is a constant number or "intercept", β the slope or "coefficient", x the predictor, and ϵ represents the noise.

$$y = C + \beta x + \epsilon \tag{1}$$

Figure 3 – Simple Linear Regression.



Source: Author.

Multiple regression extends to multiple predictors (HOPE, 2020), as shown in Equation 2. In the Equation, β values are regression coefficients, representing the relationship between input features x_i and the output result y (LIM, 2019).

$$y = C + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + \varepsilon \tag{2}$$

2.3.2.1.1 Linear Regression with Lasso regularization

In linear regression, regularization aims to reduce overfitting by constraining the model's weights. Overfitting occurs when a model performs well on the training data but fails to generalize effectively. (GERON, 2019)

Lasso, or "Least absolute shrinkage and selection operator", adds a regularization term based on the $l1$ norm of the weights to its cost function. This approach reduces the weights of the less important features, effectively selecting a simpler model that may prevent overfitting. (GERON, 2019)

2.3.2.2 Evaluation: performance metrics

Assessing regression model performance relies on various metrics, including the Mean Squared Error, Root Mean Squared Error, and Mean Absolute Error.

2.3.2.2.1 Mean Squared Error

Mean squared error (MSE) measures the average squared difference between actual and predicted values, as shown in Equation 3. The squaring of the differences makes this metric sensitive to larger errors. (NASER; ALAVI, 2023)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

2.3.2.2.2 Root Mean Squared Error (RMSE)

The root mean squared error (RMSE) metric is the square root of MSE, as depicted in Equation 4. This metric has the same unit as the target values, making it more understandable. (NASER; ALAVI, 2023)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{4}$$

2.3.2.2.3 Mean Absolute Error

Mean absolute error (MAE) measures the average absolute difference between actual and predicted values, as shown in Equation 5. (NASER; ALAVI, 2023)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{5}$$

2.3.3 Challenges of Machine Learning

Machine Learning model development faces several key challenges, as outlined in (GERON, 2019):

1. Insufficient quantity of data: Quality data is crucial for most models. Research by Microsoft Banko and Brill (2001) found that various models can perform similarly with enough data even in complex tasks like natural language disambiguation.

2. Non-Representative Training Data: The training dataset must include all necessary cases for proper generalization. Effective sampling is important to guarantee great coverage.

3. Poor-Quality data: Data cleaning, including removing outliers and noise, is essential for quality training.

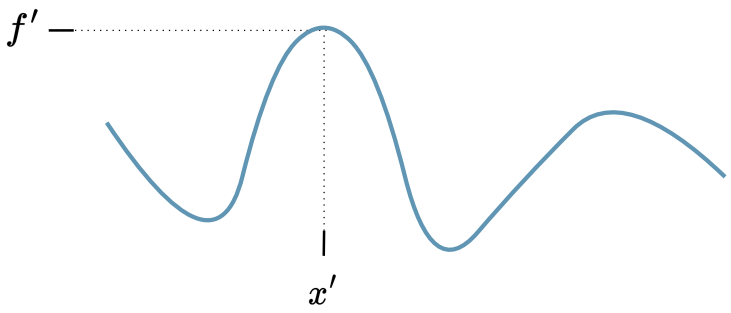
4. Irrelevant features: Feature engineering is critical for selecting, combining, or creating features that a model can learn from. This step involves feature selection and feature extraction.

2.4 OPTIMIZATION

According to Greenhill et al. (2020), the goal of every optimization process is to search in the domain one point x' which has the globally maximal value of f'. This concept is mathematically represented in Equation 6 and visually depicted in Figure 4.

$$f(x') = \max_{x \in X} f(x) = f(x') \tag{6}$$

Figure 4 – Optimization of the function f.



Source: Author.

Optimization problems are typically categorized as either single-objective or multi-objective, depending on whether they involve optimizing a single function or multiple functions, respectively. Additionally, these problems vary in dimensionality based on the number of input variables involved (SHAN; WANG, 2010).

2.5 BAYESIAN OPTIMIZATION

Bayesian Optimization (BO) methods have significantly evolved and are fundamentally based on the principle that each experiment should be purposefully designed. This principle traces back to the experimental design for polynomial regression in 1918, as introduced by Smith (1918). These foundational ideas in statistical approaches and experimental design paved the way for the advanced strategies utilized today.

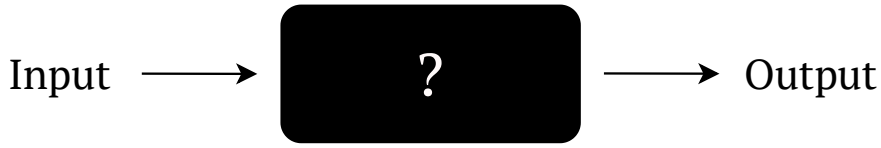
The method resurfaced with a publication by Kushner (1962), who introduced the Gaussian process model for the objective function and proposed the first optimization policies based on Bayesian Decision Theory. At the beginning of the 2000s, the method regained attention as evidenced in Jones, Schonlau, and Welch (1998), expanding its application in computer experiments and global optimization.

Nowadays, BO is a useful tool for finding global maxima of unknown functions, commonly referred to as "black box functions", where the underlying function is unknown, as illustrated in 5. This method is powerful when dealing with multidimensional problems,

multiple objectives, and constraints. Additionally, it is especially effective for expensive black-box systems due to its sample-efficient optimization algorithm (GREENHILL et al., 2020).

BO has its strengths, but it also comes with certain drawbacks and limitations. One of the primary constraints of BO is that it is most suited to continuous problems with fewer than 20 dimensions (FRAZIER, 2018). This limitation arises because BO does not scale well to high-dimensional problems or situations where a large number of samples is available (ERIKSSON et al., 2020). Another significant challenge with BO is its sensitivity to outliers. Since BO relies on probabilistic inference to make decisions, the presence of outliers can bias these inferences (MARTINEZ-CANTIN; TEE; MCCOURT, 2017).

Figure 5 – Black box function: a function where only the input and output values are known, but not its expression.



Source: Author.

BO has been an important tool for decades, being considered one of the most important statistical ideas of the past 50 years (GELMAN; VEHTARI, 2021). As stated by Shahriari et al. (2016), Bayesian Optimization has working use cases in all parts of engineering, such as A/B testing, Robotics, Reinforcement Learning, Natural Language Processing, Automatic Machine Learning, and Hyperparameter Tuning.

Generally, the iterative procedure follows what is shown in the Algorithm 1, based on Garnett (2023). With an initial dataset $\mathcal{D}$, in each iteration, the optimization policy selects the point for the next observation, which is then observed and appended to the existing dataset. This iteration continues until a termination condition is reached, such as achieving the objective or completing a set number of iterations.

Algorithm 1 Adaptive Sampling Algorithm

Require: initial dataset $\mathcal{D}$ ▷ can be empty
1: **repeat**
2: $x \leftarrow \textsc{Policy}(\mathcal{D})$ ▷ select the next observation location
3: $y \leftarrow \textsc{Observe}(x)$ ▷ observe at the chosen location
4: $\mathcal{D} \leftarrow \mathcal{D} \cup \{(x, y)\}$ ▷ update dataset
5: **until** termination condition reached
6: **return** $\mathcal{D}$

Specifically, the BO process involves several essential iterative steps, which are outlined below. The flow of these steps is visualized in Figure 6. The process fundamentally

relies on two main components: the Gaussian Processes and the Acquisition Function.

1. **Step 1**

   Initially, with a dataset $\mathcal{D}$ from the black-box function, the initial prior distribution is stored in a *objective function prior*, representing the probabilistic prior belief about the objective function.

   This step involves creating a Gaussian Process as the surrogate model to capture beliefs about the behavior of the unknown function, and an observation model to describe the data generation mechanism, which consists of noise and errors (SHAHRIARI et al., 2016).

2. **Step 2**

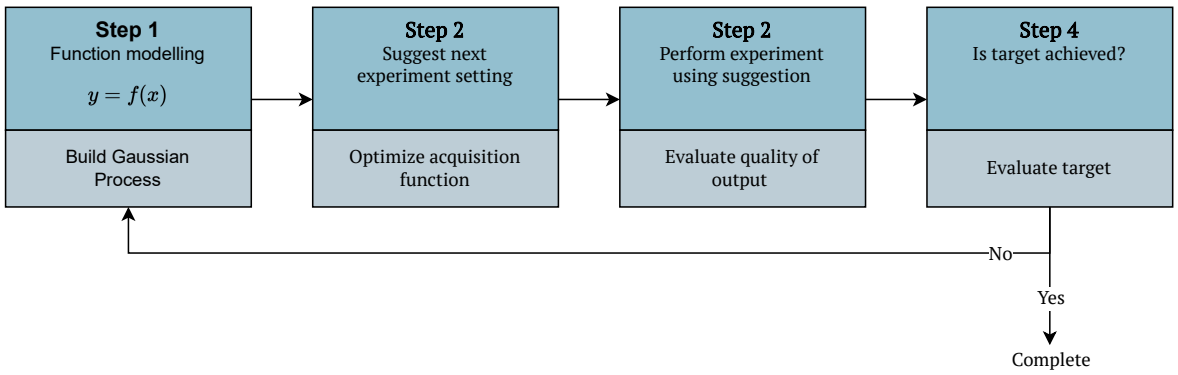
   The Bayesian optimization policies aim to select the next observation location, defined indirectly by optimizing an acquisition function that identifies potential observation locations, as stated in Garnett (2023).

   The acquisition function is optimized, identifying the most promising points for testing in the next experiment (GREENHILL et al., 2020).

3. **Step 3**

   Finally, the selected testing points are experimentally evaluated on the real process.

Figure 6 – Iterative steps of Bayesian Optimization.



Source: Author, based on Greenhill et al. (2020).

2.5.1 Surrogate Model

The surrogate model is employed in Bayesian Optimization to provide a stochastic approximation of the expensive objective function. Gaussian processes (GP) are often

chosen for this role due to their precise quantification of uncertainty (GRUVER et al., 2021), and it has been chosen for this purpose of the thesis.

However, GPs sometimes struggle with complex spaces and have certain limitations (EPPS et al., 2020). To address these challenges, alternative approaches are employed. One such alternative is the use of Random Forest, an ensemble of Decision Trees, which offers the advantage of being more computationally efficient (HUTTER; HOOS; LEYTON-BROWN, 2011). Another effective approach, especially for complex datasets, is the utilization of neural networks as the surrogate model. Employing an ensemble of neural networks can provide statistical properties similar to GPs, a method also known as Bayesian neural network, as described in (LIM et al., 2021).

2.5.1.1 Gaussian Process

Gaussian process regression is a probabilistic supervised machine learning technique used in both regression and classification tasks. As stated by Garnett (2023), it extends the familiar multivariate normal distribution to model functions on infinite domains. It assumes that the output y of a function f at input x can be described as $y = f(x) + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$ represents normally distributed noise.

It is characterized by a mean function $m(x)$, and a covariance function $k(x, x')$, which encode the prior knowledge about the function's expected value and variability, respectively (SCHULZ; SPEEKENBRINK; KRAUSE, 2018). This model allows for predictions that incorporate prior knowledge through kernels and provides uncertainty measures over those predictions (RASMUSSEN; WILLIAMS, 2005).

Specifically, a Gaussian process is defined as shown in Equation 7.

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')). \tag{7}$$

The covariance function, also called the kernel, represents the relationship between function values at different points, as can be seen in Equation 7.

$$k(x, x') = E\left[(f(x) - m(x))(f(x') - m(x'))\right]. \tag{8}$$

Many popular kernels exist in the literature, including radial basis, Matérn kernel, and rational quadratic. For the specific use case of glass lens manufacturing, the Matérn kernel is of particular interest and will be explained in detail. The Matérn kernel is a widely-used class of kernels often favored in BO due to its ability to specify the smoothness of the GP by controlling the differentiability of its sample paths, as stated in Borovitskiy et al. (2023).

The Matérn kernel is mathematically defined by Equation 9, where x and x' represent arbitrary input points.

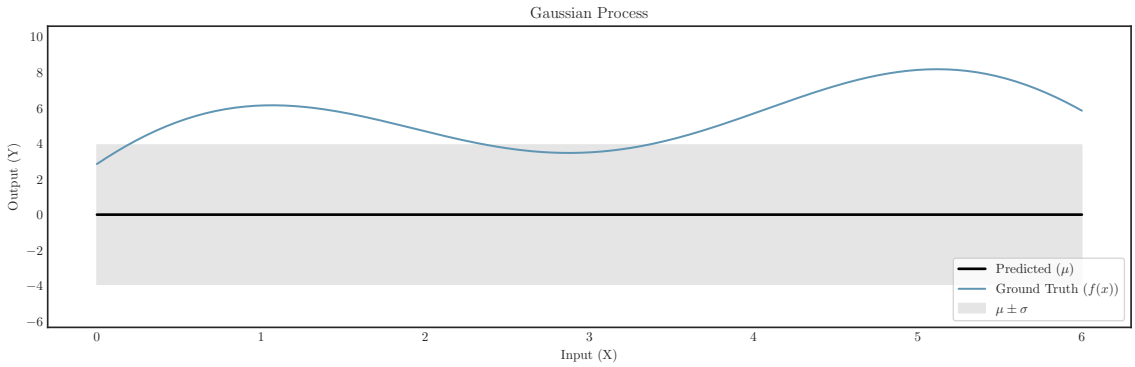
$$K(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|x - x'|}{\omega} \right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}|x - x'|}{\omega} \right), \tag{9}$$

This function has important hyperparameters that can be optimized, which are outlined below:

- ν: smoothness parameter, which governs the smoothness of the $\mathcal{GP}y$ (SANTNER; WILLIAMS; NOTZ, 2010).
- ω: scale parameter, which determines the spread of the correlation (RASMUSSEN; WILLIAMS, 2005).

In practice, the Gaussian process can be illustrated as seen in Figure 7, which illustrates a prior without any beliefs, usually starting with mean zero. It is expressed as a Gaussian Distribution with mean $m(x)$, and a covariance $k(x, x')$, shown in the graph $\mu \pm \sigma$. At this stage, the sequential optimization begins, where each new observation incrementally refines the model.

Figure 7 – Inital Gaussian Process Prior.



Source: (SCHEWINSKI, 2024).

After observing some data, the GP updates its predictions and uncertainty estimates. The optimization happens iteratively, selecting the most promising point for observations based on the acquisition function, which will be explained in the following section. This process converges the model towards the true objective, as each iteration refines the mean and reduces uncertainty, illustrated in Figure 8.

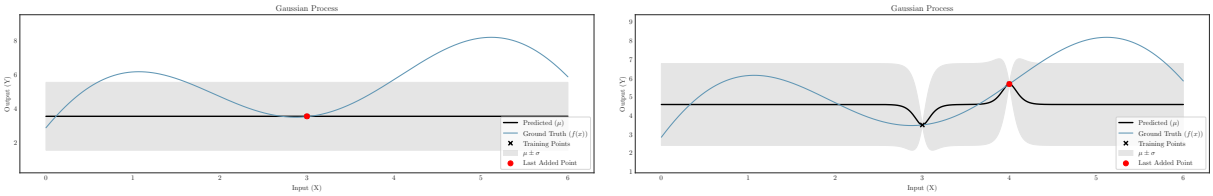
Finally, Figure 9 illustrates the Gaussian Process after multiple iterations. It reached a point of convergence, demonstrating its effectiveness in learning unknown functions.

2.5.2 Acquisition Function

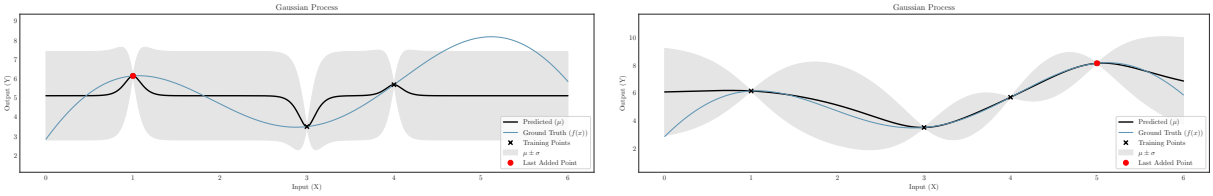
The acquisition functions aim to provide a score to each potential observation location (GARNETT, 2023) to help the optimization process, which can also be called an optimization policy. They are invaluable because of being inexpensive to evaluate compared to minimizing directly the objective function. As stated on Wilson, Hutter, and

Figure 8 – Gaussian Process after Observing Data.

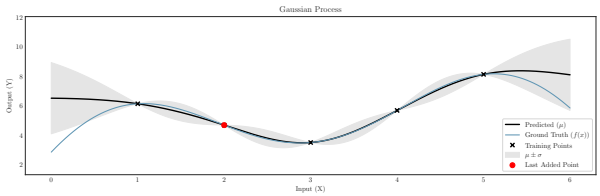
(a) Iteration 1 (b) Iteration 2



(c) Iteration 3 (d) Iteration 4

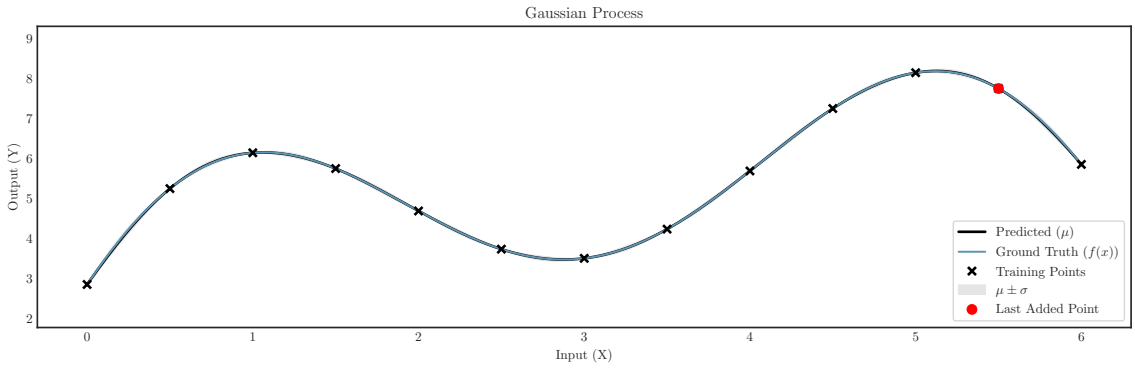


(e) Iteration 5



Source: (SCHEWINSKI, 2024).

Figure 9 – Refined Gaussian Process after Multiple Iterations.



Source: (SCHEWINSKI, 2024).

Deisenroth (2018), this function amounts to integrals defined in terms of a belief p over the unknown outcomes $y = \{y_1, \ldots, y_q\}$ revealed when evaluating a black box function f at corresponding input locations $X = \{x_1, \ldots, x_q\}$.

There are many acquisition functions with different goals and methods for selecting the next sample. However, to achieve this with as few function evaluations as possible, a balance between exploration and exploitation is required (GREENHILL et al., 2020).

Exploitation aims to sample where the objective function is expected to be high, while exploration intends to sample where there is uncertainty about the objective function, as mentioned in Garnett (2023).

Most acquisition functions are myopic policies, decided based on a one-step lookahead utility function (YUE; KONTAR, 2022), ignoring the long term impact of each selected sample. In contrast, in the emerging scenario of Reinforcement Learning, non-myopic methods have risen, creating multi-step lookahead algorithms that can perceive the informative gains obtainable by looking ahead into future steps (FIORE; MAININI, 2022). Detailed information about multi-step lookahead in Bayesian Optimization will be thoroughly examined in Section 2.5.2.3.

In this thesis, the focus is on multiobjective and nonmyopic acquisition functions, particularly due to their relevance in optimizing the real production of lenses at IPT. This emphasis is driven by the fact that the lens production process is costly and time-consuming, necessitating a limited number of precisely selected experiments. The choice of these specific types of acquisition functions is aimed at maximizing the efficiency and effectiveness of each experimental iteration.

For example, consider the visualization in Figure 10, which has one objective function and the acquisition function. The top graph represents the Gaussian Process of the function $sin(x)$. The bottom plot illustrates the acquisition function, showcasing a peak, highlighted by an arrow, indicating the most promising new point to sample next. This point represents the trade-off between exploration and exploitation, where the next evaluation would yield the most valuable information.

2.5.2.1 Hypervolume-Based Acquisition

The Hypervolume indicator is a metric used to evaluate the quality of solutions in multi-objective optimization, specifically known as the Pareto front.

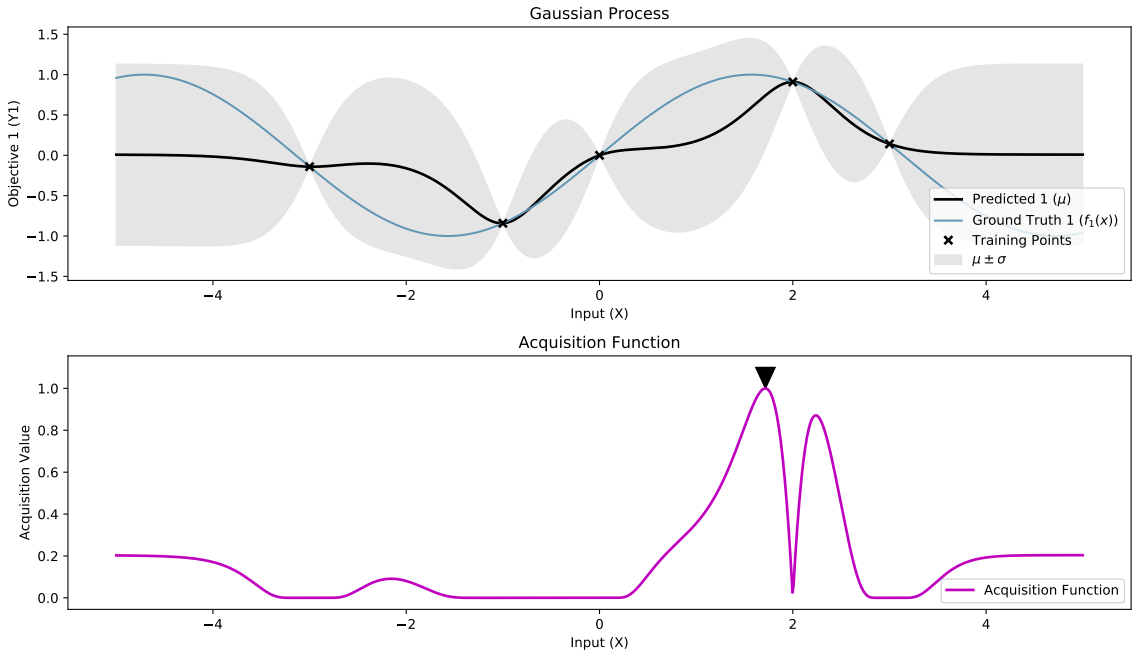
The Pareto front is a collection of "best" solutions in a multi-dimensional space, formally expressed as $P = \{y^{(1)}, \ldots, y^{(n)}\} \subset R^d$. Each point on this front represents a solution that is not outperformed by any other in all objectives, known as non-dominated solutions. (YANG et al., 2019)

In this context, a reference point r is used, typically representing a worst-case scenario in the objective space. The Dominated Space refers to all points inferior to the Pareto front solutions, effectively the space between each $y^{(i)}$ and r.

The Hypervolume Indicator is defined as the volume of the space that is not dominated by the Pareto front and bounded below by the reference point. It is formally represented in the Equation 10. This metric helps in the optimization process by quantifying how much "better" the Pareto front is compared to the worst-case scenario with a measurable quality of solution.

For example, consider the case in Figure 11 with the Pareto front as

Figure 10 – Gaussian Processes and Acquisition Function.



Source: (SCHEWINSKI, 2024).

$P = \{y^{(1)}, y^{(2)}, y^{(3)}\}$ and a reference point r. The Hypervolume would then be the region in light blue that lies above the reference point and below the Pareto front, highlighting the extent of optimal solutions available.

$$HV(P) = \lambda_d(\bigcup_{y \in P} [r, y]) \tag{10}$$

2.5.2.2 Expected Hypervolume Improvement

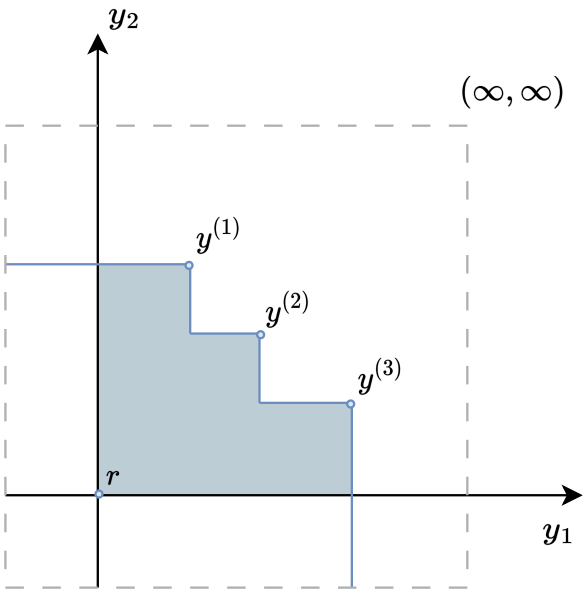
The Expected Hypervolume Improvement (EHVI) is an acquisition function based on hypervolume. It is designed to measure how much a new solution, y, can contribute to the existing Pareto front, P. (YANG et al., 2019)

Initially, the Hypervolume Improvement (HVI) of a new solution y to the Pareto Front P is calculated as the difference in hypervolume when y is included in P. It is mathematically expressed in Equation 11.

The HVI process can be seen in Figure 12 where the current Pareto front is denoted as $P = \{y^{(1)}, y^{(2)}, y^{(3)}\}$, and r the reference point. When a new point $y^{(+)}$ is added, the Hypervolume improvement $HVI(y^{(+)}, P, r)$ is the area of the dark blue polygon.

$$HVI(y, p) = HV(P \cup y) - HV(P) \tag{11}$$

Figure 11 – Hypervolume.



Source: Author.

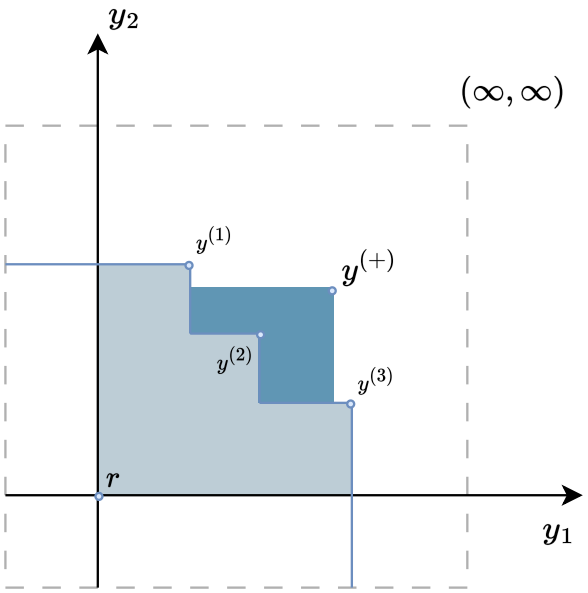
EHVI extends this concept by considering the predictive distribution (μ, σ) of potential new solutions given by a surrogate model, as expressed in Equation 12. Here, $\xi_{\sigma,\mu}(y)$ represents the probability density function of the predictive models's output. By integrating over all possible solutions, EHVI calculates the expected increase in hypervolume given the current knowledge of the surrogate model. Thus, the optimization selects regions expected to yield significant improvements to the Pareto front.

$$EHVI(\mu, \sigma, P, r) := \int_{R^d} HVI(y, P, r) \cdot \xi_{\sigma,\mu}(y)\, dy \tag{12}$$

Figure 13 illustrates EHVI, where the probability density function $\xi_{\sigma,\mu}(y)$ of the bivariate Gaussian distribution is shown as a 3D plot, indicating where new potential solution might exist, as well as the current Pareto front $P = \{y^{(1)}, y^{(2)}, y^{(3)}\}$, and the reference point r. EHVI samples points from the probability distribution, for example, sample $y^{(+)}$, and calculates the area of improvement indicated in the dark gray area. This process is crucial as it samples points that are both probable and valuable in improving the set of optimal solutions.

In one practice example, as shown in Figure 14, after three observations from a multi-objective problem, two posterior beliefs are leveraged to calculate the EHVI, quantifying the potential benefit of new observations to the Pareto front. The peak of the EHVI curve indicates the most promising location for the next observation.

Figure 12 – Hypervolume.



Source: Author.

2.5.2.3 Lookahead Bayesian Optimization

Lookahead Bayesian Optimization is a non-myopic auxiliary acquisition function that is used in conjunction with another. In recent years, non-myopic acquisition functions have been studied, and for this present thesis, the focus is on the "Efficient Nonmyopic Bayesian Optimization via One-Shot Multi-Step Trees" method presented by Shali Jiang et al. (2020), which is implemented by BoTorch (BALANDAT et al., 2020).

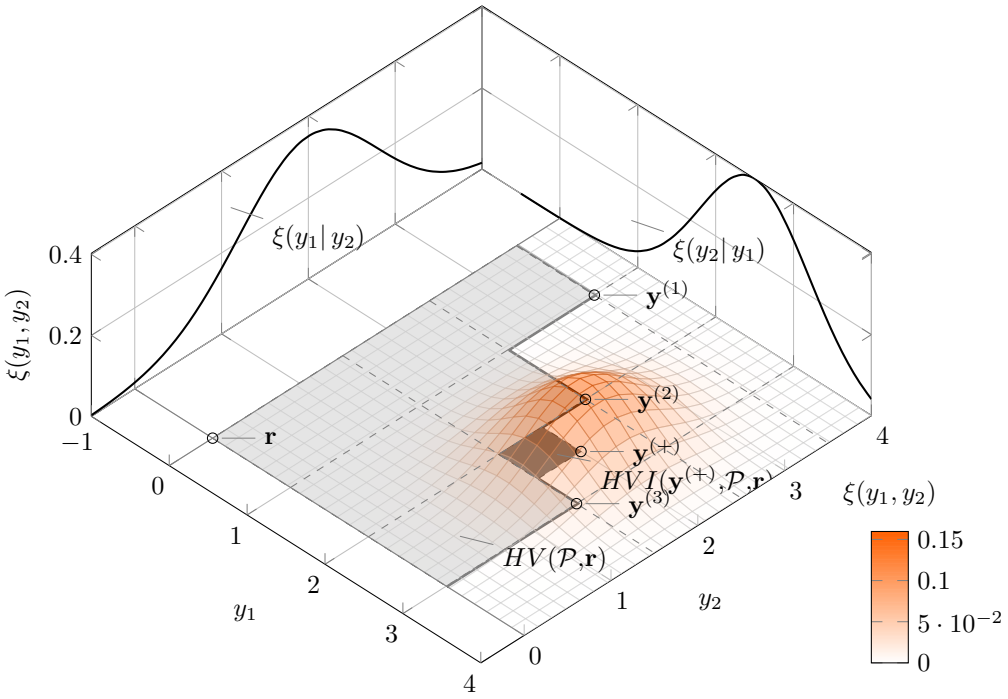
This method considers the expected improvement over a sequence of future samples. As Shali Jiang et al. (2020) states, it begins with an initial set of observations D_0 and a probabilistic surrogate model $p(Y|X, D_0)$. At each step k, the set of observations D_k is updated to maximize the utility function $u(D_k) = \max_{(x,y)\in D_k} y$, indicative of the highest peak of the function. This sequential process is guided by a policy, such as EHVI, determining the next location to query based on the current data.

Using Bellman recursion, it's possible to decompose the one-step marginal value, which determines the expected benefit of querying, into a multi-step framework for $t = 2, 3, ..., k$ (Equation 13). Consequently, the k-step lookahead acquisition function, $v_k(x|D)$, guides the next query point $\arg\max_x v_k(x|D)$.

$$v_1(x|D) = E_y \left[u(D \cup \{(x, y)\}) - u(D) \mid x, D \right]. \tag{13}$$

$$v_t(x|D) = v_1(x|D) + E_y \left[\max_{x'} v_{t-1}(x'|D \cup \{(x, y)\}) \right]. \tag{14}$$

Figure 13 – Hypervolume.



Source: Yang et al. (2019).

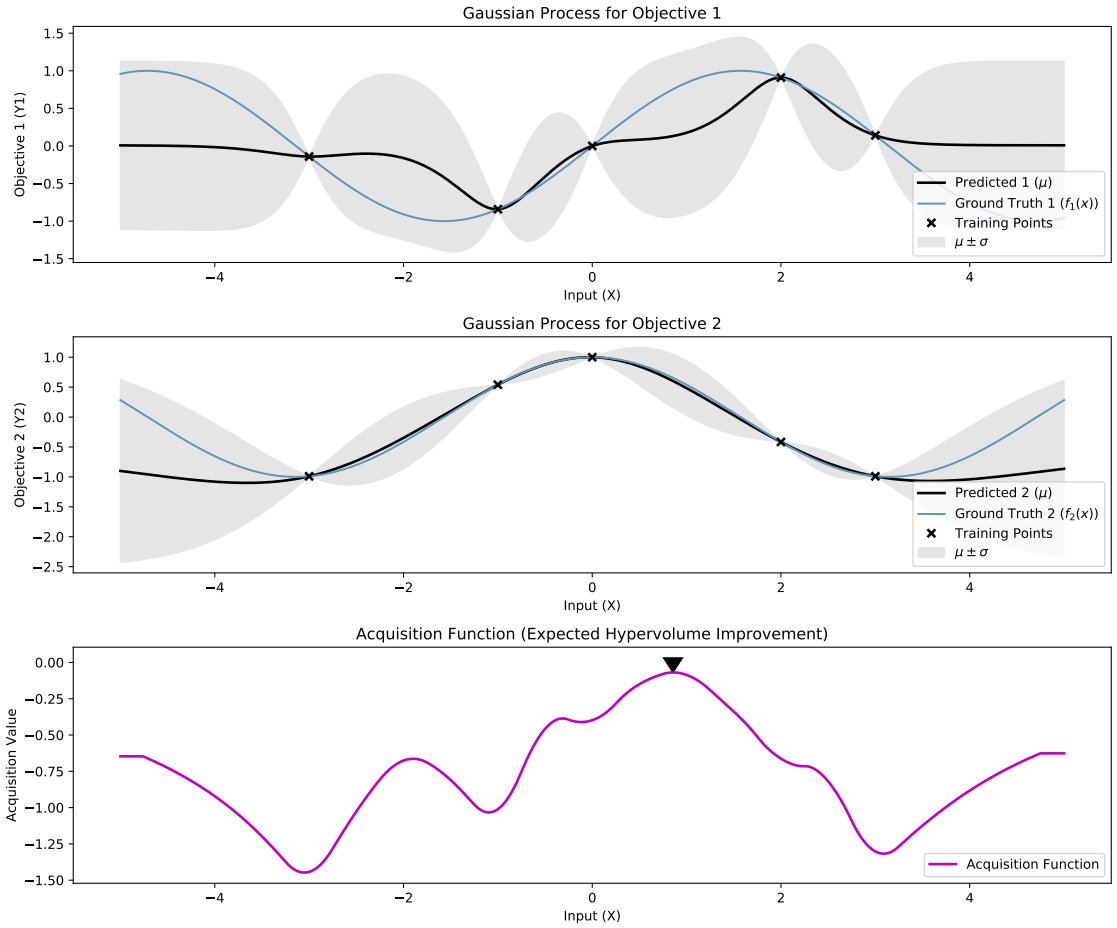
The library and paper introduce a one-shot multi-step approach to solving the k-step problem, $v_k(x|D)$, by solving a joint optimization problem, instead of other traditional ones that require nested optimization problems. This is very beneficial due to its reduced complexity and computational costs. (JIANG, S. et al., 2020)

2.6 BENCHMARK FUNCTIONS

Benchmark functions will be used to evaluate and select the most suitable acquisition function and its parameters to be used in the real-world lens optimization, as further discussed in Section 3.3.3.

This process will involve comparing the performance of all acquisition functions across various synthetic and non-synthetic functions to identify the optimal configuration. Subsequently, this optimal configuration will be applied to real-world lens optimization. Adopting this strategy is crucial, as it is impractical to test every possible configuration in real-world scenarios. These benchmark functions include the multi-objective DTLZ family and the Vehicle Safety function.

Figure 14 – Hypervolume.



Source: (SCHEWINSKI, 2024).

2.6.1 DTLZ

The family of DTLZ test functions, developed by Deb et al. (2002), are designed to be simple to construct, scalable, and capable of handling an arbitrary number of objectives. Another advantage of DTLZ functions is their well-defined solutions with simple shapes, such as spheres, curves, or simplices. A total of six test problems will be used, each testing a specific aspect of optimization capability. Each function presents its own challenges for optimization.

These test functions are constructed based on two main approaches: a test problem generator, and a constraint surface approach. The test problem generator allows for the creation of generic multiobjective test functions, where users can choose a h function related to the objective function, as shown in Equation 15. Additionally, a g function is chosen to construct the overall objective search space, as illustrated in Equation 16. (DEB et al., 2002)

$$
\begin{aligned}
&\text{Minimize} \quad f_1(\mathbf{x}_1), \\
&\vdots \\
&\text{Minimize} \quad f_{M-1}(\mathbf{x}_{M-1}), \\
&\text{Minimize} \quad f_M(\mathbf{x}) = g(\mathbf{x}_M)h(f_1, \ldots, f_{M-1}, g), \\
&\text{subject to} \quad \mathbf{x}_i \in R^{|\mathbf{x}_i|}, \quad \text{for } i = 1, 2, \ldots, M.
\end{aligned} \tag{15}
$$

$$
g_j(f_1, f_2, \ldots, f_M) \geq 0, \quad \text{for } j = 1, 2, \ldots, J. \tag{16}
$$

DTLZ1 through 3 are characterized by numerous local Pareto-optimal fronts, testing the algorithm's convergence ability (ZHANG et al., 2009). DTLZ4 analyses how the optimization handles a non-uniform distribution of points on the Pareto front (ZHANG et al., 2009). DTLZ5 challenges the optimization algorithm's ability to converge to a degenerated curve (LIU, 2017), while DTLZ7 complicates convergence by creating a surface composed of a straight line and a hyperplane.

2.6.2 Vehicle Safety

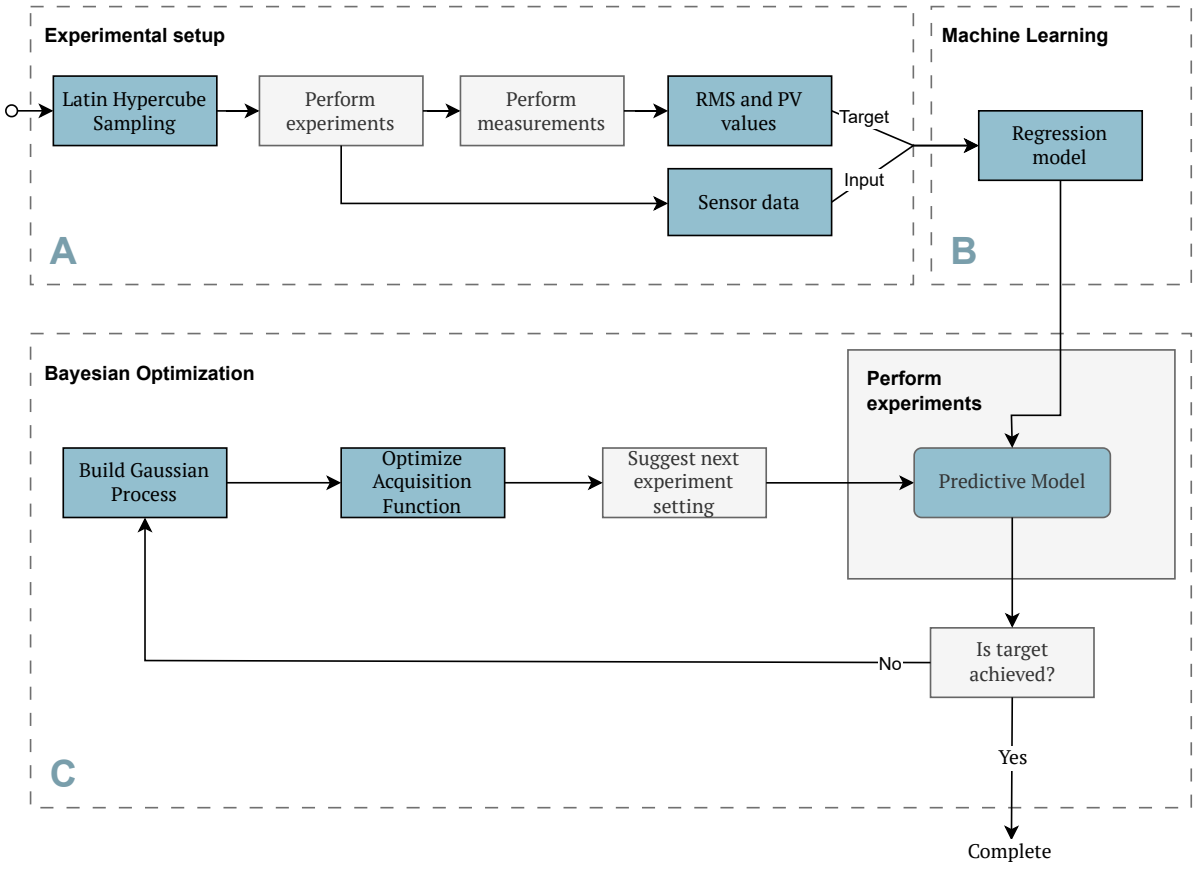
Vehicle safety represents a real-world multiobjective optimization problem. Benchmarks with real problems are also important, as synthetic test problems may include unrealistic properties, leading to overestimation or underestimation, as noted in Tanabe and Ishibuchi (2020). The vehicle safety optimization problem, proposed by Liao et al. (2008), models a full-width crash test and offset-frontal crash test. The model is purely mathematical and represents a real crash scenario. It involves a 5-dimensional parameter space with three objectives.

3 METHODOLOGY

In this chapter, the methodology and steps taken to develop a comprehensive Lookahead Bayesian Optimization and Machine Learning model for the surrogate for the validation will be detailed.

From a broader perspective, as depicted in Figure 15, the process begins with the experimental setup (step A in the Figure). This involves designing and conducting a set of experiments on the lens manufacturing machine, followed by measuring the lens quality through peak-to-valley and RMS values, and gathering process sensor data. With the experimental data and measurements in hand, a machine learning regression model (B) is created to simulate the real manufacturing process. The primary aim of the model is to evaluate the effectiveness of the Lookeahead Bayesian Optimization approach in process parameter optimization. Subsequently, the Lookahead Bayesian Optimization model (C) is configured, selecting the most suitable parameters and functions to optimize its performance. Instead of utilizing the real manufacturing process for experimental evaluations in the Bayesian Optimization step, it uses the predictive regression model.

Figure 15 – Thesis Methodology.



Source: Author

3.1 EXPERIMENTAL SETUP

The experimental setup outlines all the necessary steps for data collection and preparation, which will later be utilized in the Machine Learning model. This setup includes designing experiments with Latin Hypercube sampling, performing the experiments, conducting measurements, and gathering sensor data.

3.1.1 Latin Hypercube Sampling

Latin Hypercube Sampling for the NGM process was implemented using the *pyDOE2* Python library, which is specifically designed for creating experimental designs, alongside other fundamental ones such as *numpy* and *smt*. The objective is to develop a design of experiments that relatively covers the entire parameter space, as a result, the system will be accurately represented. Achieving representative coverage is crucial for ensuring optimal performance in the machine learning model that will act as the surrogate for validation. This is one of the challenges in machine learning, as discussed in Section 2.3.3.

The sampling had to meet specific parameter constraints and requirements that were not inherently supported by the library and were custom developed. The essential constraints included adhering to the predefined bounds and resolution for each parameter, critical due to machine specifications. These parameters and their specifications are detailed in Table 1.

Parameter	**Range/Bounds**	**Resolution**
Pressing Force	40-47 %	1 %
Press Time (Speed)	1-3s	0,5 s
Hold Time After Pressing	1,5-4s	0,5 s
Upper Tool Temperature	450-480 °C	10 °C
Lower Tool Temperature	460-510 °C	10 °C
Oven Temperature	900-1000 °C	25 or 50 °C
Heating Time at 900	190-255s	5 or 10 s
Heating Time at 950	175-215s	5 or 10 s
Heating Time at 1000	155-175s	5 or 10 s

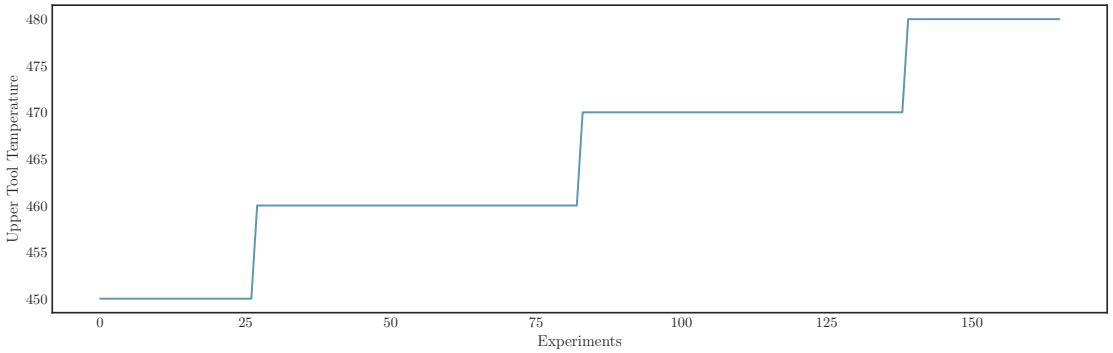
Table 1 – Parameter Specifications

The sampling involved generating 500 samples across six dimensions for each of the three heating times (900 °C, 950 °C, and 1000 °C) to maintain time efficiency and machine constraints. Four optimization methods from the library were utilized to enhance the sampling process, with selection based on best coverage performance.

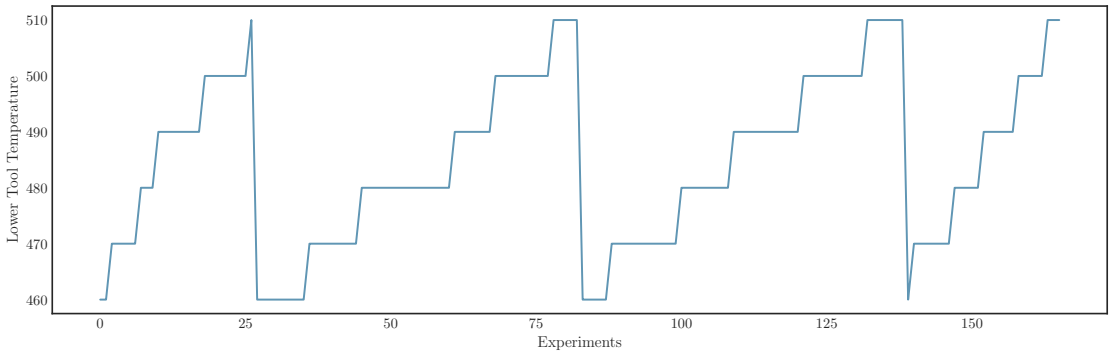
The post-sampling experimental preparation took into account the parameter's importance, since pressing forces are easier to change, while others, like furnace temperature, are more challenging, consuming significant time in the production process, and varying greatly between the highest and lowest temperatures throughout the experiments. This

Figure 16 – Progress over time of temperature within experiences samples

(a) Upper Tool Temperature



(b) Lower Tool Temperature



Source: Author

is demonstrated in Figure 16, indicating minimal fluctuations in temperature throughout the trials.

3.1.2 Lens Machine

The press machine utilized for the lens manufacturing process is manufactured by *Füller Glastechnologie*, but it has been extensively customized by Fraunhofer IPT to include additional sensors and actuators. These customizations enhance the machine's capabilities, allowing extensive research, understanding, and optimization of the lens manufacturing process.

3.1.3 Lens Quality

The lens's quality, as depicted in Figure 17(a), is evaluated using a form measurement machine displayed in Figure 17(b). The machine utilized by IPT is "Form Talysurf", manufactured by Taylor Hobson.

To accurately assess the lens, it undergoes two measurements across its diameters at right angles, forming an orthogonal configuration. This dual orientation ensures a

Figure 17 – Lens Quality

(a) Lens. (b) Lens Measurement.

Source: Author.

comprehensive evaluation, verifying the lens's uniform curvature in all directions and confirming its spherical symmetry.

3.1.3.1 Peak-to-Valley

The Peak-to-Valley (PV) measurement, provided by the machine, represents the disparity between the highest and lowest points on the surface by comparing the actual optic surface to an ideal standard (SPINA et al., 2012). This metric considers only the extreme values, disregarding the intermediary points. Consequently, a lower PV value indicates that the lens more closely aligns with the ideal lens design, signifying higher quality, whereas a higher PV value suggests poorer quality. This method has a drawback in that it does not consider the optics' curvature, thus failing to adequately monitor their performance (SPINA et al., 2012).

For instance, in a given experiment, each lens yields PV values for the 3D shape and at 0°, 45°, 90°, and 135° angles. The 3D shape report, and consequently the PV values, are illustrated in Figure 18.

Following this, the highest (and thus worst) PV value from each lens is manually selected from a dataset similar to that in Table 2. This selection is part of the optimization process, which aims to minimize the largest PV value for each lens.

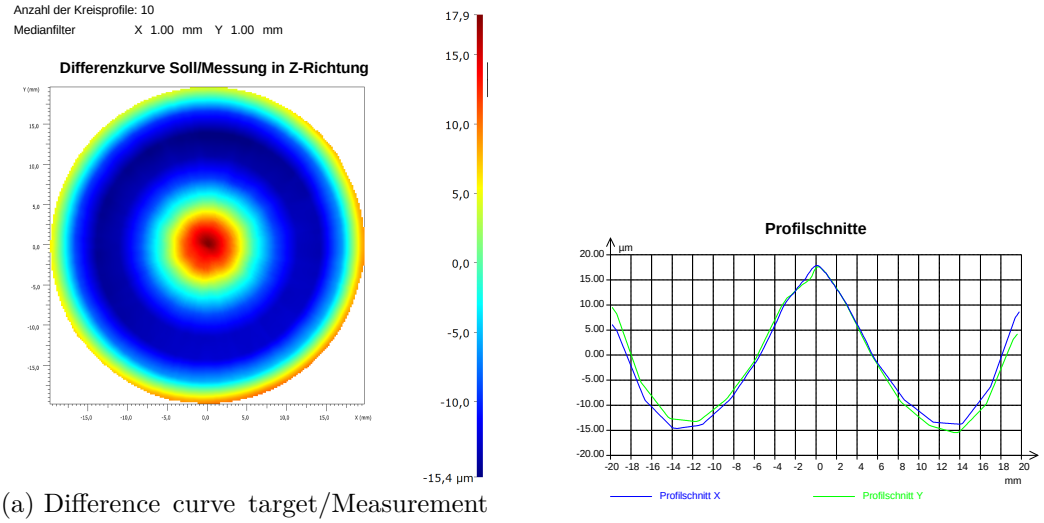
Lens Number	**3d**	**0.0°**	**45.0°**	**90.0°**	**135.0°**	**Max PV**
105114	33.3 µm	32.9 µm	31.8 µm	46.1 µm	32.6 µm	46.1 µm

Table 2 – Lens Measurement Data - PV

3.1.3.2 Root Mean Square

The Root Mean Square (RMS) value, derived from the lens measurements, quantifies the deviation of the lens surface from the ideal form. It measures a large part of the

Figure 18 – Lens' 3D shape report.



(a) Difference curve target/Measurement in Z direction.

(b) Profile Cuts.

Source: Author.

optical surface at many spots and calculates the standard deviation of the surface from its optimal shape. (SPINA et al., 2012)

Ultimately, each lens is assigned a single scalar value that quantifies its RMS and indicates the quality of its curvature. For instance, Table 3 presents the measurement for one lens.

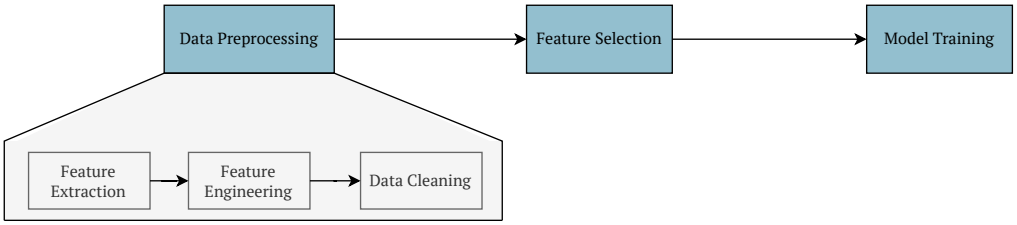
Lens Number	**RMS**
105114	6.9 µm

Table 3 – Lens Measurement Data - RMS

3.2 MACHINE LEARNING PIPELINE

The machine learning pipeline, as depicted in Figure 19, initiates with the data processing step. This phase encompasses feature extraction, engineering, and cleaning, ensuring the data is prepared and refined for subsequent stages. Following data processing, the pipeline proceeds to feature selection and model training, utilizing the entire dataset.

Figure 19 – Machine Learning pipeline.



Source: Author

3.2.1 Objectives and Requirements

The objective of the Machine Learning model is to simulate the real lens manufacturing process. For a given set of input parameters, the model should predict two key quality metrics of the manufactured lenses: peak-to-valley (PV) and root mean square (RMS) values.

The specific requirements are as follows: the PV values must achieve a mean absolute error of less than 80 microns m (equivalent to $80 \cdot 10^{-6}m$). Similarly, for the RMS values, the goal is to maintain a mean absolute error under 80 microns m (also $80 \cdot 10^{-6}m$). These standards were established by process experts.

3.2.2 Data Preprocessing

Before creating the model and doing all the other steps, both the input and target data must undergo preprocessing. The input data includes recorded sensor information and metadata from the process, while the target data comprises PV and RMS values.

3.2.2.1 Target Data Preprocessing

The PV data is initially stored in a dataframe, sourced from the measurement machine as detailed in Section 3.1.3.1. This dataset has outliers resulting in measurement problems. To remove them, all experiments with PV outliers were dropped, resulting in a total of 12 experiments being excluded from the dataset.

The RMS values, vital for this study, were calculated by Lukas Jäschke[1]. He computed the RMS by doing a comparison between the experimental data points against optimal ones. This involves a lens form correction to address any tilting of the lenses, as well as the removal of outliers due to measurement errors. Outliers were removed from the data when their values exceeded three standard deviations from the mean. This was followed by manual inspection and analysis to ensure that important information was not inadvertently eliminated. The original count of 489 data points was reduced to 455 after outlier removal.

3.2.2.2 Input Data Preprocessing

The input data consists of two dataframes: metadata, representing the actual input parameters used in lens manufacturing, and sensor data. The metadata, derived from the LHS sampling, didn't have any outliers or quality issues. However, the sensor data passed through feature extraction, engineering, and cleaning. These steps aim to ensure data quality and relevant features, which are discussed as challenges in machine learning in Section 2.3.3.

3.2.2.3 Sensor Data Feature Extraction

The sensor data includes 122 time-series datasets for each lens produced. Non-informative sensors, including those with insignificant standard deviation, were removed, leaving 26 out of the original 122 sensors.

Before proceeding to subsequent stages, the sensor data undergoes a pre-cleaning phase. This phase involves interpolating the signal for each of the process steps outlined in Table 4. This step is crucial as it helps to smooth out the time series data, effectively removing numerical noise (WANG; WANG, 2020).

The feature extraction was performed based on the time series of each sensor. This process was elaborated by the author and Lukas Jäschke[2]. The time series has 9 process steps (outlined in Table 4 and visualized in Figure 20), from which can be extracted the actual duration of each process step as the first features. The steps 5 and 8 are not present in the Figure since are not part of the temperature sensor.

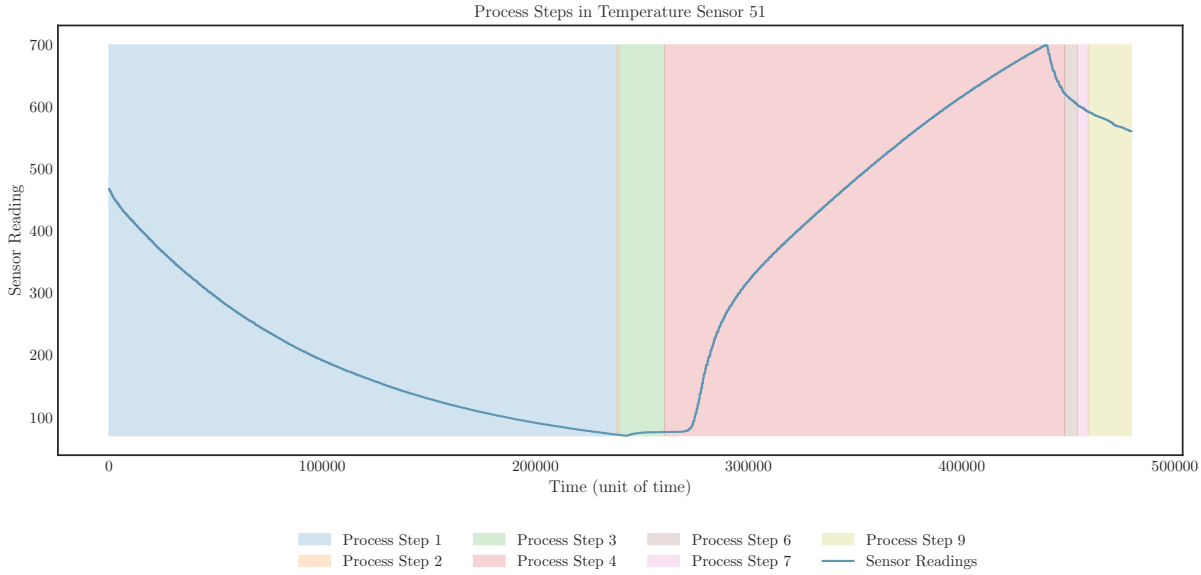
These sensors are placed at strategic positions and stages to monitor the process. In Figure 21, two important steps of this process are highlighted. The first step is the heating phase, where the oven is heated to reach the desired temperature, and the mold containing the glass is inserted. After the heating phase, the glass is then moved by a robotic arm to the form stage, where the lens takes shape.

[1] Lukas Jäschke, Fraunhofer IPT, lukas.jaeschke@ipt.fraunhofer.de.

Process Step	Description
1	Basic setting
2	Preheating
3	Loading
4	Heating
5	Measuring before forming
6	Forming
7	Demolding
8	Temperature measurement after forming
9	Moving the axis away

Table 4 – Process Steps

Figure 20 – Process Steps in Temperature Sensor 51.



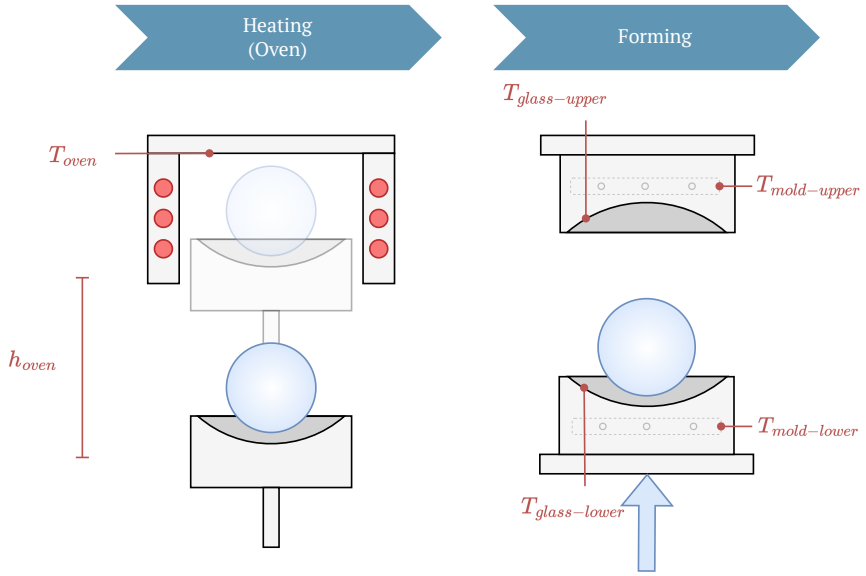
Source: Author.

Temperature Sensors

According to process experts, temperature sensors are crucial for ensuring lens quality. These sensors, strategically positioned at various points within the molds, gather essential data through thermocouples.

For Sensor 51, its role is to monitor the temperature during the heating process step, as depicted in Figure 21, where it is labeled as T_{oven}. Figure 22(a) displays the analyzed time series data of the temperature. In the process of Step 1, an exponential function $f(x) = ab^x + k$ was fitted, and its slope and AUC (Area Under the Curve) were recorded as features, indicating mold setting activities (22(b)). During Process Step 04 (22(c)), a linear function $f(x) = ax + b$ was fitted, and its slope, along with the maximum and minimum temperatures, rate, duration of temperature increase, and AUC, were saved as features.

Figure 21 – Sensor locations.



Source: Author

Sensors 50 and 52 are both responsible for measuring the temperature of the glass during the forming process. They are responsible for monitoring the temperature from the lower and upper parts of the glass, referred to as $T_{glass-upper}$ and $T_{glass-lower}$. The location of these sensors is illustrated in 21. Both sensors exhibit similar time series curves, capturing the temperature dynamics during the forming process.

As shown in Figure 23(a), the full time series data for these sensors reveal a characteristic curve. From this data, several key features were extracted. During Process Step 9, the slope and AUC were determined and recorded as features. Additionally, metrics such as the maximum temperature, the rate of temperature increase, and the duration of this increase were also extracted, as depicted in Figure 23(b).

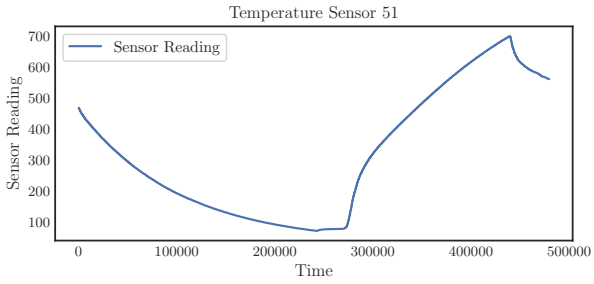
Additional temperature-related sensors, specifically Sensors 38, 39, and 119 for monitoring the upper mold temperature ($T_{mold-upper}$), and sensors 40, 41, 119, and 120 for the lower mold temperature ($T_{lower-upper}$) were also analyzed. These sensors, as indicated in Figure 21, are responsible for regulating the mold temperature. Their time series data, as shown in Figure 24, exhibit temperature fluctuations during the heating process due to their role in temperature control. Several features were extracted, including average temperature, and, through Fast Fourier Transformation, the mean rate of change during temperature rises and falls, mean peak, and cycle duration.

Moreover, for all temperature sensors involved in the glass molding process, both the mean and standard deviation of temperature values were extracted during Process Step 4.

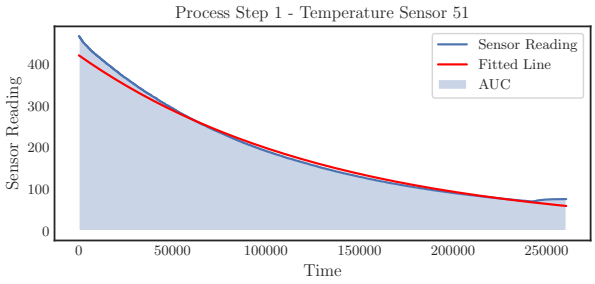
Actuator Sensors

Figure 22 – Temperature Sensor 51.

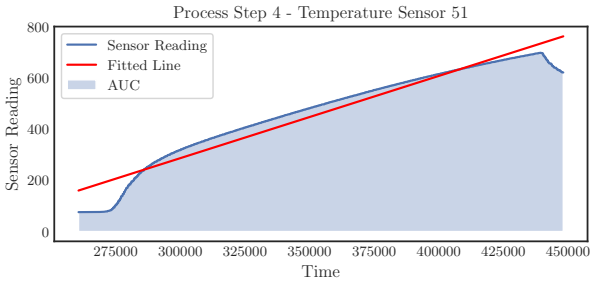
(a) All steps.



(b) Process Step 01.



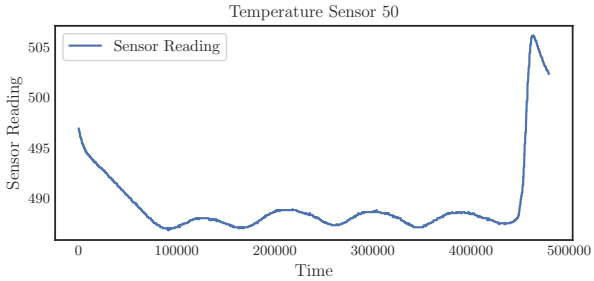
(c) Process Step 04.



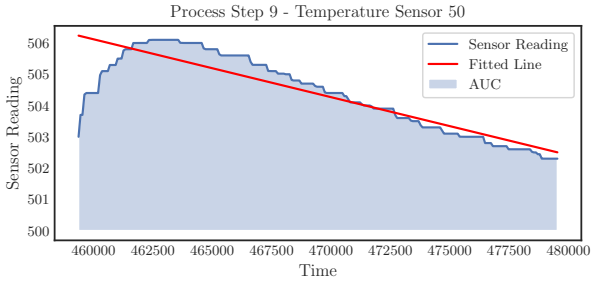
Source: Author.

Figure 23 – Temperature Sensor 50 and 52.

(a) All steps

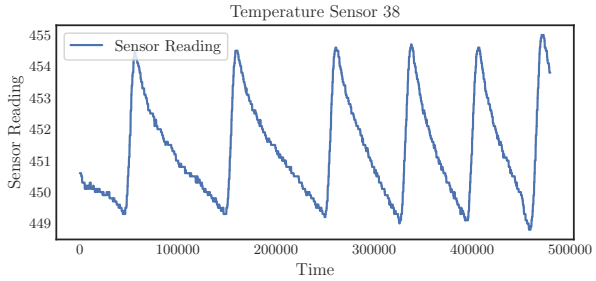


(b) Process Step 9



Source: Author.

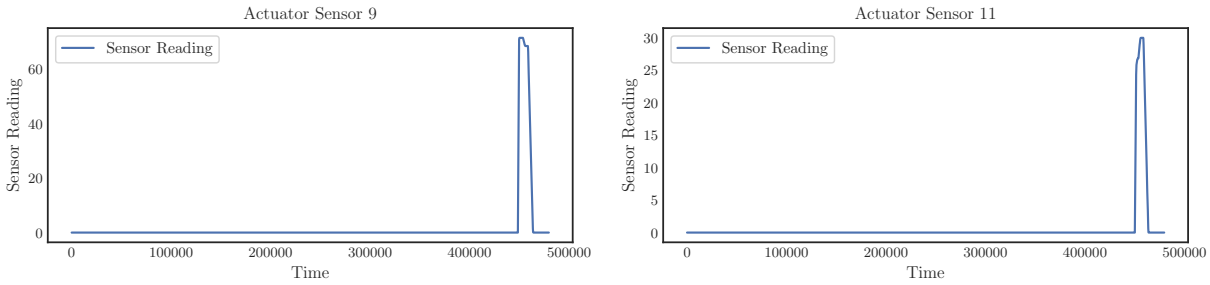
Figure 24 – Temperature Sensor 38.



Source: Author.

Figure 25 – Pressure Sensors.

(a) Sensor 9. (b) Sensor 11.

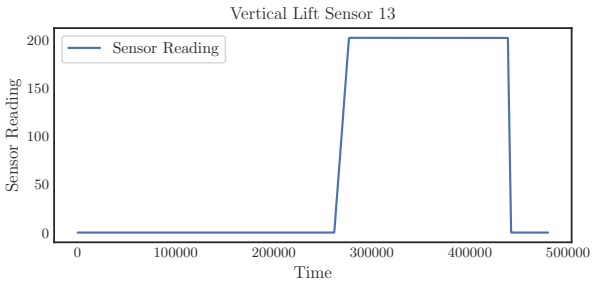


Source: Author.

During the lens forming stage, sensors 9 and 11 capture the movement of the actuator. These sensors enable the extraction of various features including the duration and position of the actuator movement, as well as the pressure applied. As shown in Figure 25, a noticeable increase in the sensor's measurement is observed during Process Step 6, which corresponds to the lens forming stage and the actuator's movement. This increase is followed by a subsequent decrease. Key features derived from these sensors include the position of the actuators, pressure, and the duration of pressing.

Figure 26 – Vertical Lift Sensor.

(a) Sensor 13.



Source: Author.

Vertical Lift Sensors

Finally, Sensor 13 measures the vertical lift of the actuator, as depicted in Figure 21. This sensor monitors the actuators placing the glass in the heating stage, indicated as h_{oven}. Features extracted include the total duration that the workpiece spends in the heating process (Figure 26), as well as the position of the actuator.

In total, 78 features were extracted to be used in the model.

3.2.2.4 Sensor Data Feature Engineering

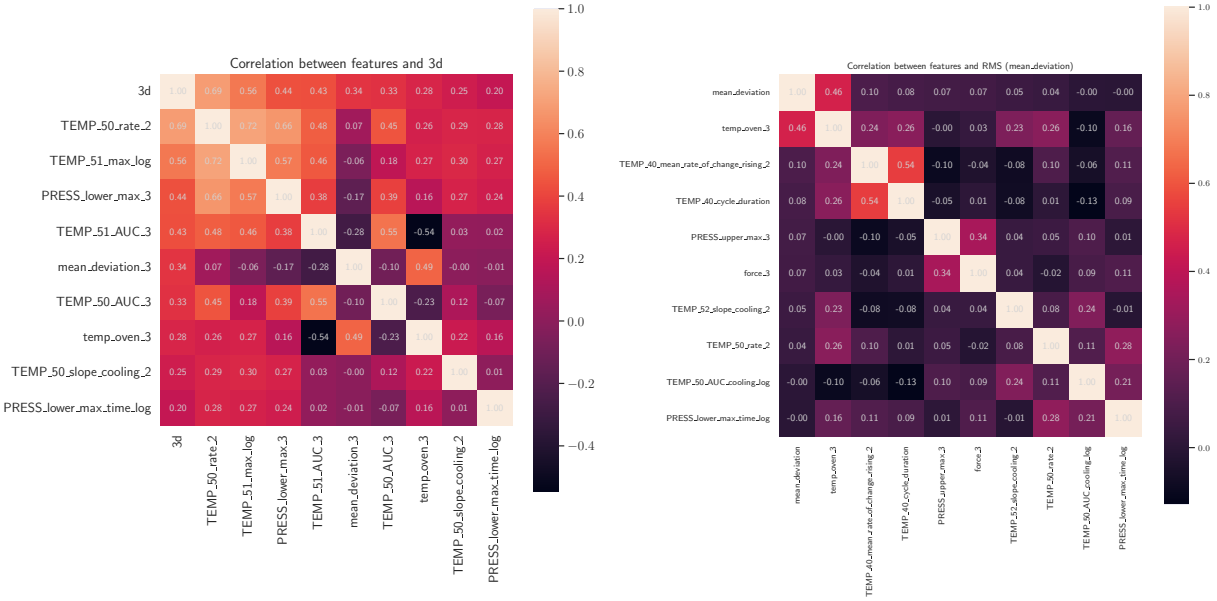
Many of the features have non-linear relationships or non-normal distributions, both of which are important considerations in linear regression. To address these issues, transformations like logarithmic and power functions are applied to the features (BOX; COX, 1964).

Automated feature engineering was employed due to the multitude of important process features. For each feature, transformations including $x^2, x^3, log(x)$, and $\sqrt{x}$ were calculated. Subsequently, supervised feature selection was applied among these transformed variables to select those with the highest correlation to the target feature (QU; HARIRI; YOUSIF, 2005). Given two targets in this use case, the PV value was prioritized for the selection of the transformed features by correlation analysis.

Feature reduction was also performed to remove highly correlated features (correlation greater than 0.8), addressing multicollinearity issues. Multicollinearity happens when two different variables are highly correlated, being more efficient to consider only one to describe the system (ZAMEE; HAN; WON, 2022). If not removed, these extra features can cause inaccurate regression coefficients, inflate the standard errors, and degrade the predictability of the model as mentioned by Hasan, Yusuf, and Faruque (2019). Figure 27 illustrates the most correlated features with both targets post-reduction.

Figure 27 – Correlation between features and target.

(a) PV (b) RMS



Source: Author.

3.2.2.5 Sensor Data Cleaning

The composed features from the sensor data had a lot of outliers and inconsistencies, possibly due to sensor faults or mathematical errors. The outlier removal reduced the dataset count from 455 to 323, just affecting the ones with data exceeding three standard deviations from the mean.

It is noteworthy that data cleaning was performed as the final step. The raw data, initially in time series format, didn't have issues like missing values or temporal misalignments, but rather only contained outliers. This made it more practical to address outlier removal at the end of the preprocessing sequence after having non-time-series data.

3.2.3 Feature Selection

All the available features underwent a selection process using Sequential Feature Selector from Sklearn, with a Lasso model as the base estimator. The selector operates in a backward direction, scores based on negative mean absolute error, and employs 5-fold cross-validation. The primary objective was to identify the 10 most crucial features that ensure optimal model performance without jeopardizing the Bayesian Optimization with Gaussian Process, which struggles with high-dimensional data (YI et al., 2024). The selected features are:

1. Minimum Temperature of sensor 51, with cubic transformation.
2. AUC Temperature of sensor 51, with cubic transformation.

3. Standard deviation of Temperature of sensor 50, with logarithmic transformation.
4. Standard deviation 1 of Temperature of sensor 52, with cubic transformation.
5. Standard deviation 2 of Temperature of sensor 52, with logarithmic transformation.
6. Maximum upper pressure, with cubic transformation.
7. Maximum lower pressure, with cubic transformation.
8. Maximum duration of lower pressure, with logarithmic transformation.
9. Oven temperature, with cubic transformation.
10. Force, with cubic transformation.

3.2.4 Model training

The training and evaluation of the model are conducted using the Sklearn library with a Lasso regression to mitigate overfitting. Given the presence of two targets, separate models are developed for each.

Initially, the dataset for each target is divided into training and testing sets, with the test size comprising 20% of the data. The features are scaled by removing the mean and scaling to unit variance, a crucial step since Lasso Regression is sensitive to the scale of input parameters. The data split helps to generalize the model, ensuring it is well on both training and unseen test data. The model is only trained with the training data, and the test data is utilized to evaluate the model performance using various metrics.

Hyperparameter tuning is applied through grid search to optimize the regression model. Parameters tuned include the alpha coefficient (regularization strength of the Lasso model), maximum number of iterations, and tolerance for optimization. Negative mean absolute error is the chosen metric, evaluated through 5-fold cross-validation to prevent overfitting and guarantee robustness. During the cross-validation step, the model is evaluated in a number of splits (five in this case) using the validation data, and the model that performs best in the k-fold is selected. This hyperparameter tuning process is done for each target, allowing different optimal parameters of each model.

Finally, after finding the best hyperparameters, the selected model is evaluated in the test set.

3.3 LOOKAHEAD BAYESIAN OPTIMIZATION FRAMEWORK

3.3.1 Libraries

This work primarily utilizes Python, which became popular because of its versatility and simplicity. Along with it, some other python libraries were employed, such as Pandas, Numpy, SkLearn, and notably, Pytorch and Botorch.

3.3.1.1 PyTorch

As highlighted by Paszke et al. (2019), Pytorch is a Python library that performs immediate execution of dynamic tensor computations with automatic differentiation and GPU acceleration, and does so while maintaining performance comparable to the fastest current libraries for deep learning. Its significance is amplified when employing the Lookahead Acquisition function, an intensive task due to the exponential increase in problem dimensionality with Lookeahead horizon k steps (JIANG, S. et al., 2020). The library's high performance is attributed to its C++ core, good integration with the Python ecosystem, and its ability to run either on CPU or GPU.

3.3.1.2 BoTorch

Balandat et al. (2020) describe BoTorch as a modular, scalable Monte Carlo framework for Bayesian Optimization, grounded in modern computation paradigms and recent theoretical convergence results.

It is implemented on top of PyTorch, guaranteeing high performance and developer efficiency with a modular interface for optimization components. The library aims to bridge research and production, providing a reliable framework that is very flexible and easily integrated into other platforms.

3.3.2 Lookahead Bayesian Optimization Pipeline

To establish the pipeline for the Lookahead Bayesian Optimization in a multi-objective context, it is essential to define both the surrogate model and acquisition function that are capable of handling multi-objective scenarios. These components are crucial for the iterative steps of BO, which are shown in Figure 6 and explained in Section 2.5.

3.3.2.1 Surrogate Model

In this study, the Gaussian Process (GP) regressor serves as the surrogate model, a common choice in Bayesian Optimization. The selected kernel to be used is Matérn, configured with standard parameters of ν (smoothness parameters) and ω (length scale). When dealing with multiobjective scenarios, one can either employ a separate GP for each target or a single GP for all targets. In this thesis, due to the lack of correlation between

the targets and to enhance the GP's performance, the approach of using a distinct GP model for each target was adopted. Consequently, for each target—PV and RMS values — a separate GP was employed for each.

3.3.2.2 Acquisition Function

For this multiobjective Bayesian Optimization, five different acquisition functions were used, being two of them provided by *BoTorch*, and the rest tailored to improve the optimization of this use case.

3.3.2.2.1 BoTorch Acquisition Functions

In this present work, two base acquisition functions will be used from BoTorch, which are "parallel Expected Hypervolume Improvement (qEHVI)" and "parallel Noisy Expected Hypervolume Improvement (qNEHVI)". Both of them are based on the principle of the Expected Hypervolume Improvement explained in Section 2.5.2.2, with the difference that the former also considers noisy settings.

3.3.2.2.2 Custom Acquisition Functions

One latent problem in Bayesian Optimization is the difficulty in incorporating prior knowledge, such as process specifics and expert insights, into standard Bayesian Optimization (WANG; MING, et al., 2019). This limitation is particularly relevant in scenarios where there is initial domain knowledge available.

Furthermore, high-dimensional use cases, like the one addressed here, create a big search space. As a result, as noted by Ramachandran et al. (2020), the algorithm visits low function values regions more often before locating the optimum. This happens because the algorithm always assumes equal probability to all search space to be the optimum.

To address this limitation in BO, incorporating expert knowledge early in the optimization process can significantly enhance efficiency. Process experts often possess valuable insights into the characteristics of "good" and "bad" regions within the process. Leveraging these insights allows for the development of an informed acquisition function that guides the optimization strategy, relying on the specialized knowledge of process experts.

Incorporating custom acquisition functions aims to utilize process expert knowledge in lens production to make more assertive decisions in the optimization process. Research, such as that presented in Khatamsaz et al. (2023), demonstrates that this approach can significantly accelerate convergence to the solution and potentially outperform traditional methods.

Employing an informed Bayesian Optimization with a non-myopic approach, as discussed by Astudillo and Frazier (2022), shows great potential. This approach also

benefits significantly from considering multiple lookahead steps, rather than just a single step.

At Fraunhofer IPT, process experts have an intuition of three crucial parameters for lens quality: peak-to-valley, RMS values, and input parameters. After careful selection, the chosen benchmarks for peak-to-valley and RMS are both 0.02. Consequently, the optimal solution is defined as $pareto\_front\_y = [0.02, 0.02]$. Additionally, the best input process parameters ($pareto\_front\_x$) for producing high-quality lenses were also determined based on process expert knowledge and past experiments.

3.3.2.2.3 Custom Biased Expected Hypervolume Improvement (qBiasEHVI)

Developed by the author, this method aims to enhance the sampling process within an acquisition function by introducing a reward mechanism for samples near the process expert optimal solution beliefs. This strategy involves identifying promising areas in the parameter space and creating bias in sampling new candidates toward these regions in the early stages of optimization. The bias is implemented by assigning weights for each sample generated by the posterior, based on its proximity to the ideal Pareto Front.

Formally, being S as the set of posterior samples, where each sample is a vector $S_i \in R^n$, l as the $pareto\_front\_y \in R^n$, and β as the bias strength. The custom acquisition functions samples from the posterior, and calculate the Euclidean distance between S_i and l by $d_i = ||\mathbf{s}_i - \mathbf{l}||_2$. It calculates the weight for each sample using an exponential function based on the distance and bias strength $w_i = \exp(-\beta \cdot d_i)$. And in the end, apply the weights to the original sample $\mathbf{s}'_i = w_i \cdot \mathbf{s}_i$.

The equation for this process is presented in Equation 17, and the methodology is outlined in Algorithm 2.

$$\mathbf{s}'_i = \exp(-\beta \cdot ||\mathbf{s}_i - \mathbf{l}||_2) \cdot \mathbf{s}_i \tag{17}$$

3.3.2.2.4 Custom Prior-Weighted Acquisition Function (qWeightedEHVI)

Based on Hvarfner et al. (2022), this approach seeks to improve an existing acquisition function heuristics $\alpha(x, \mathcal{D}_n)$ (in this case, qEHVI) by integrating it with a weighting scheme, forming a prior-weighted version. The weighting scheme denoted as $\pi(x)$, highlights the favorable points of the prior based on their similarity, such as proximity, to the $pareto\_front\_x \in R^n$, as identified by the process expert. The weighted version is calculated as shown in Equation 18 and described in Algorithm 3.

$$x_n = \arg\max_{x \in \mathcal{X}} \alpha(x, \mathcal{D}_n)\pi(x). \tag{18}$$

Algorithm 2 qBiasEHVI

Require: $model$, ref_point, $partitioning$, $sampler$, $objective$, $constraints$, $X\_pending$, eta, fat, $iteration$, $pareto\_front\_y$
Ensure: qBias Expected Hypervolume Improvement
function GetBiasedPosteriorSample($posterior, pareto\_front\_y, bias\_strength$)
    Sample from posterior
    Calculate weights based on distance from $pareto\_front\_y$
    return Weighted samples
end function
function Forward(X)
    Compute posterior using the model
    Obtain biased samples from posterior using GetBiasedPosteriorSample($posterior, pareto\_front\_y, bias\_strength$)
    return Biased qEHVI values
end function

Algorithm 3 qWeightedEHVI

Require: $model$, ref_point, $partitioning$, $sampler$, $objective$, $constraints$, $X\_pending$, eta, fat, $iteration$, $pareto\_front\_x$
Ensure: qWeighted Expected Hypervolume Improvement
function ComputeSimilarity(X)
    Compute Euclidean distance from each point in X to $pareto\_front\_x$
    Invert the distances to get similarity measure
    Normalize similarities to have a maximum of 1
    return normalized similarities
end function
function Forward(X)
    Compute posterior using the model
    Obtain samples from posterior
    Compute qEHVI values using ComputeQEHVI
    Compute expert values using ComputeSimilarity(X)
    Compute Weighted qEHVI
    return Weighted qEHVI values
end function

3.3.2.2.5 Custom Decaying Prior-Weighted Acquisition Function (qDecayWeightedEHVI)

Expanding on the concept introduced by Hvarfner et al. (2022), this acquisition function improves the custom Prior-Weighted Acquisition Function by incorporating a decaying factor. This factor raises the prior to a power of $\gamma_n \in^+$, similar to the approach in Souza et al. (2021). The decaying factor γ_n is defined as $\gamma_n = \beta/n$, where $\beta \in R^+$ is the decaying factor, and n the optimization iteration number.

It is a valuable method as it allows for increasing trust in the surrogate model over the prior as optimization progresses. Being $\pi(x)$ the weighting scheme, the weighted version is presented in Equation 19 and Algorithm 4. In line with the findings of Hvarfner et al. (2022), a decaying factor of $\beta = 10$ is recommended for optimal performance.

$$\alpha_{\pi,n}(x, \mathcal{D}_n) = \alpha(x, \mathcal{D}_n)\pi(x)^{\beta/n}. \tag{19}$$

Algorithm 4 qDecayWeightedEHVI

Require: $model$, ref_point, $partitioning$, $sampler$, $objective$, $constraints$, $X\_pending$, eta, fat, $iteration$, $pareto\_front\_x$
Ensure: qDecayWeighted Expected Hypervolume Improvement
function ComputeSimilarity(X)
    Compute Euclidean distance from each point in X to $pareto\_front\_x$
    Invert the distances to get similarity measure
    Normalize similarities to have a maximum of 1
    return normalized similarities
end function
function Forward(X)
    Compute posterior using the model
    Obtain samples from posterior
    Compute qEHVI values using ComputeQEHVI
    Compute expert values using ComputeSimilarity(X)
    Compute Weighted qEHVI
    return Weighted qEHVI values
end function

3.3.2.2.6 Lookahead Bayesian Optimization Framework

This thesis does not delve into the specifics of the code implementation for the Lookahead Bayesian Optimization framework, as it involves extensive detail. The implementation was carried out using an object-oriented approach, to create a user-friendly and efficient package applicable to various projects.

To utilize the framework, users can configure initial settings in a file named *config.json*. This configuration includes specifying the initial data for the algorithm, process experts' opinions on the Pareto Front, and the bounds of the input parameters. The

data and optimizer are then initialized using the command *generate_initial_data()* and *BayesianOptimization()*, as shown in Listing 3.1. Following this initialization, the optimization process can start to determine the next candidates for testing in the actual manufacturing process. The *optimizer.optimize* method allows customization, including the selection of the number of candidates, batch size (which influences the parallelization of the algorithm), number of fantasies (indicating the number of lookahead steps and their complexity), and the choice of acquisition function. Available acquisition functions include qEHVI, qNEHVI, and custom ones.

```
# Initialize initial data
train_x, train_obj, train_obj_true, n, bounds, dim =
    generate_initial_data()

# Initialize Bayesian Optimization
optimizer = BayesianOptimization(
    train_x, train_obj, train_obj_true, n, bounds, dim, config
)

# Optimize and get the next candidates
optimizer.optimize(
    N_BATCH=1, q=10, q_batch_sizes=[5, 2, 2, 2], num_fantasies
   =[5, 5, 5, 5], aqf="qehvi_lookahead_d", iteration=1,
)
```

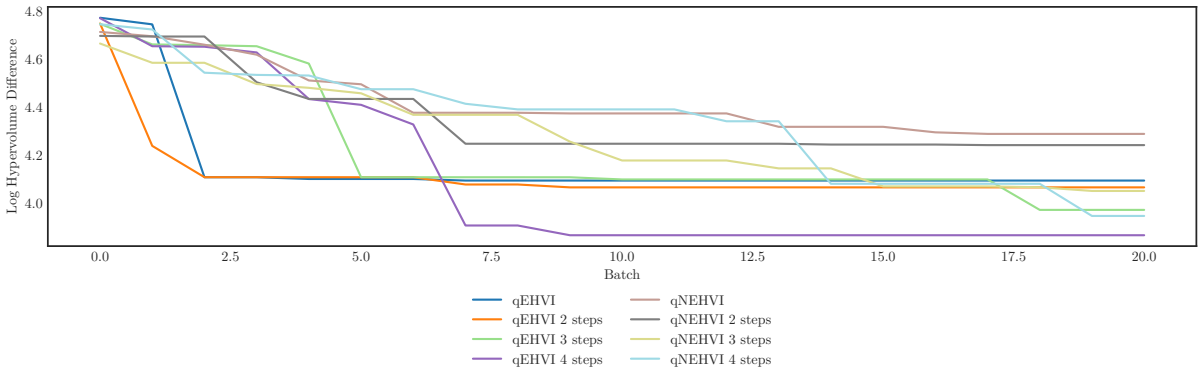
Listing 3.1 – Usage of Lookahead Bayesian Optimization Framework

3.3.2.2.7 Benchmark Acquisition Functions

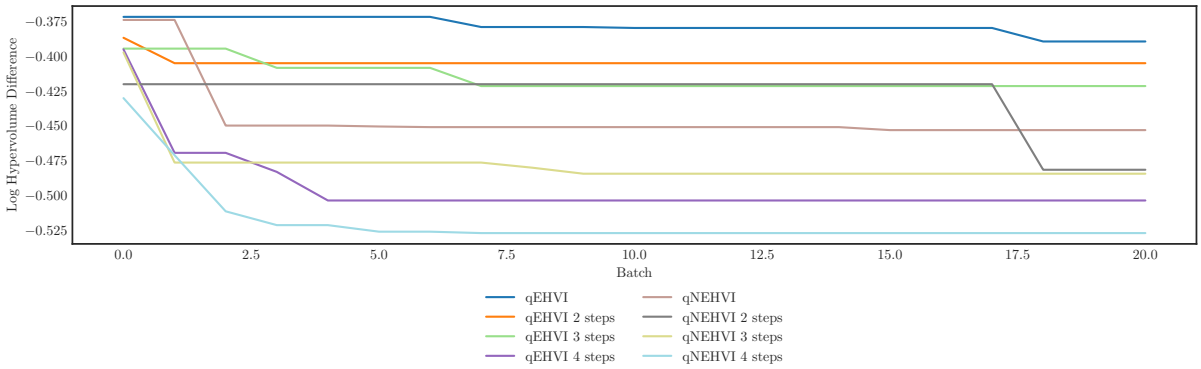
This section evaluates various acquisition functions in test problems, aiming to select the most effective one for real process optimization. Direct testing in real processes is often impractical, so test functions are used to assess the quality in advance of these acquisition functions.

Initially, a baseline acquisition for the acquisition functions was established. Two benchmark tests involving "qEHVI" and "qNEHVI" were conducted to determine their performance, and later choose the best one.

The setup of the tests was aligned with conditions similar to the Optics use case which is the goal of the present thesis. It involves a 10-dimensional parameter space with two objectives to optimize, specifically minimizing them. Each iteration includes 10 experimental runs, referred to as candidates. The tests were limited to 20 batches to simulate scenarios where real experiments are costly. Additionally, different lookahead steps for each function were examined.



(c) DTLZ1



(d) DTLZ2

Figure 28 – Performance of Acquisition Functions over Test Functions.

Source: Author.

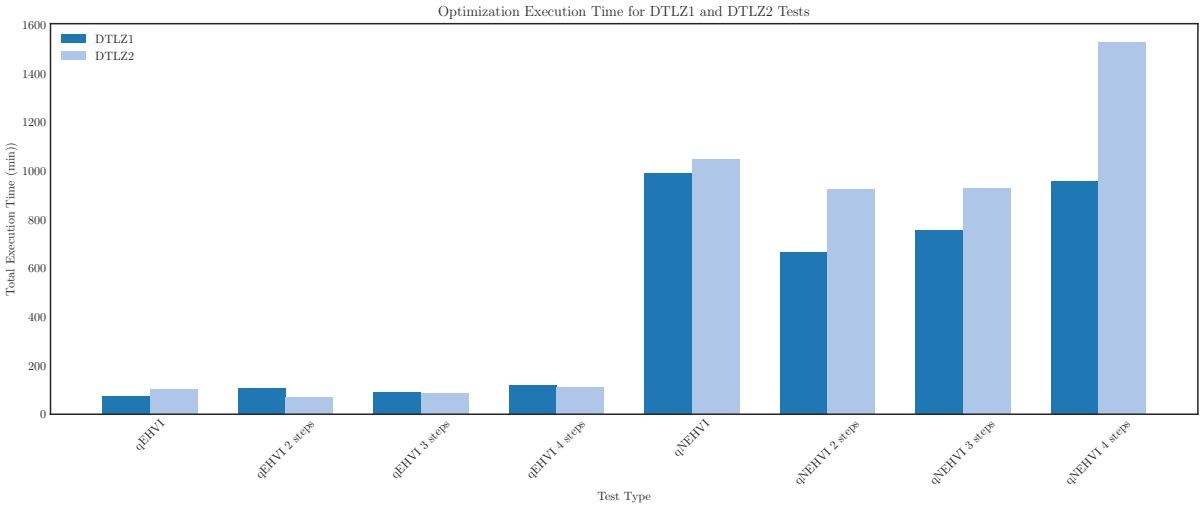
Figure 28 illustrates the performance of the acquisition function on both benchmark functions. It is noticed that "qEHVI" outperforms "qNEHVI" in this setup, and its execution time is radically faster, as shown in Figure 29.

Based on the insights from previous experiments, qEHVI has been chosen to be the base acquisition function. Subsequent benchmark tests were conducted to compare its performance with custom acquisition functions and various lookahead steps. These comparative analyes are depicted in Figures 30, 31, and 32. The tests included the DTLZ1 through DTLZ7 functions, all configured similarly with 10-dimensional parameter spaces, two objectives, and 10 candidate solutions per iteration.

In the DTLZ1 problem, qWeightedEHVI exhibited rapid convergence without the need for lookahead steps. In contrast, qEHVI showed good performance by the end of 20 batches, effective in scenarios both with and without lookahead steps. For DTLZ2 and DTLZ4, qDecayWeightedEHVI indicated promising results, demonstrating adaptability both with and without lookahead steps. In DTLZ3, qEHVI achieved the best outcomes, maintaining its effectiveness regardless of the use of lookahead steps. Finally, in both DTLZ5 and DTLZ7, qEHVI and qDecayWeightedEHVI performed well.

The results lead to two conclusions:

Figure 29 – Execution time.



Source: Author

1. The most effective acquisition functions are qEHVI and qDecayWeightedEHVI qEHVI.

2. The lookahead feature in Bayesian Optimization improves convergence and leads to a better Pareto front.

Other acquisition functions, such as Biased and Prior-Weighted, did not consistently have satisfactory performance. The Biased function, while optimizing effectively over time, sometimes overly concentrated on specific areas, failing to identify an optimal Pareto front, especially in DTLZ3 and DTLZ7. Similarly, Prior-Weighted, though adequate, was outperformed by the aforementioned functions.

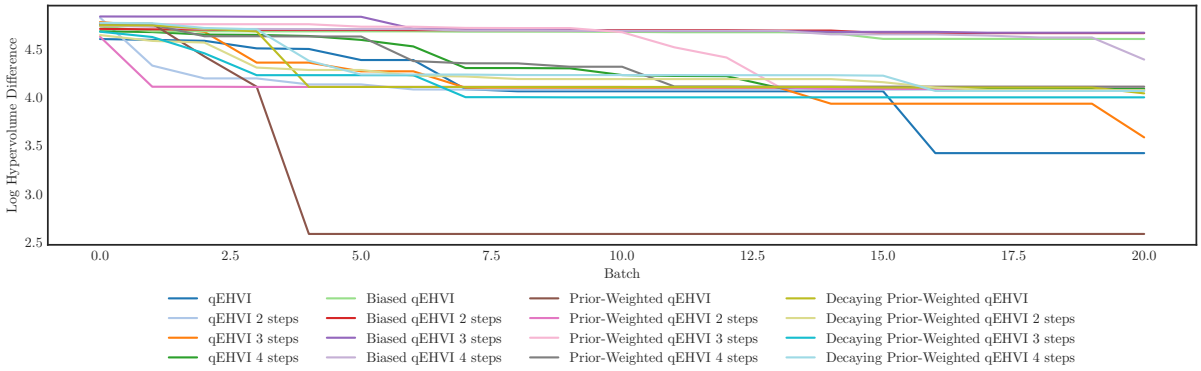
Finally, tests on a real Vehicle Safety optimization problem were conducted, as depicted in Figure 32. This problem, differing from the Lens use case, had a 5-dimensional parameters space with three objectives. The performance of qEHVI, with and without lookahead steps, reached the best optimization in the end.

3.3.3 Lookahead Bayesian Optimization in real process

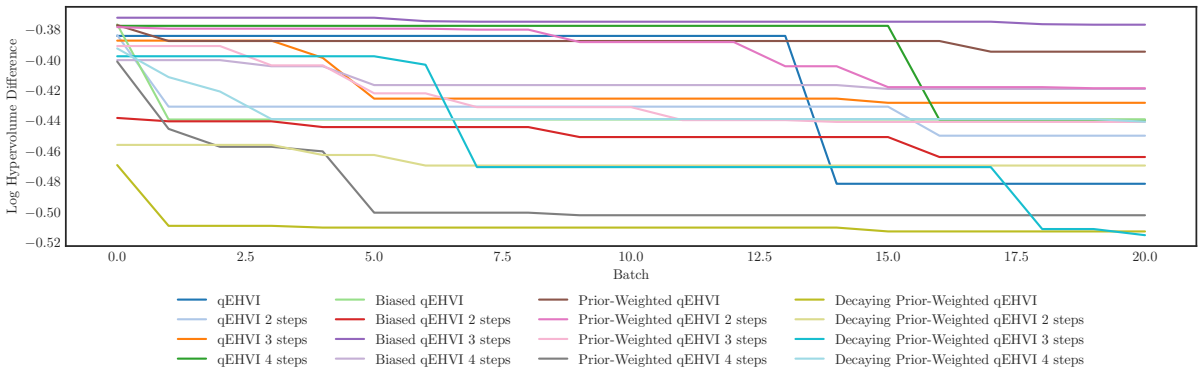
The Lookahead Bayesian Optimization framework is set to be implemented in the real lens manufacturing process. This process, typically conducted in batch operations, requires a cooling period and manual quality assessment for each batch, making it time-intensive. Given the long duration required to assess quality and prepare for subsequent batches in a one-batch process, an optimization framework capable of handling batch operations and incorporating lookahead strategies is essential.

Figure 30 – Performance of Acquisition Functions over Test Functions.

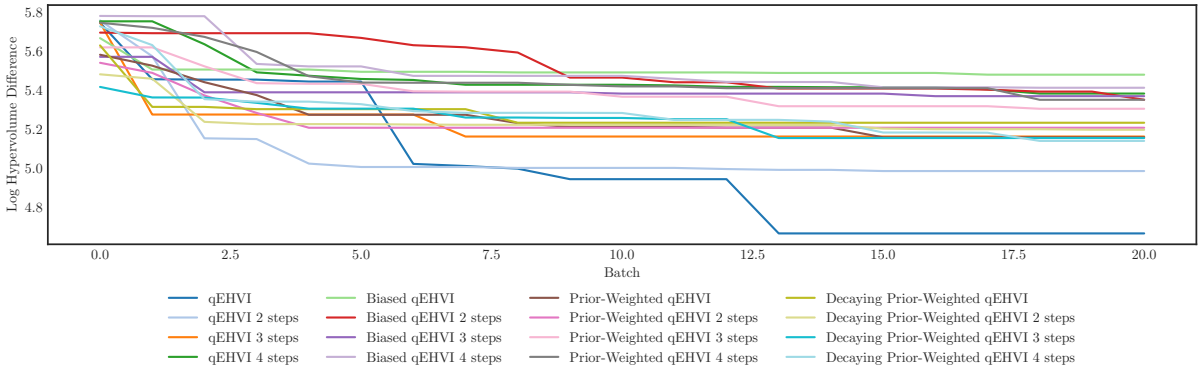
(a) DTLZ1



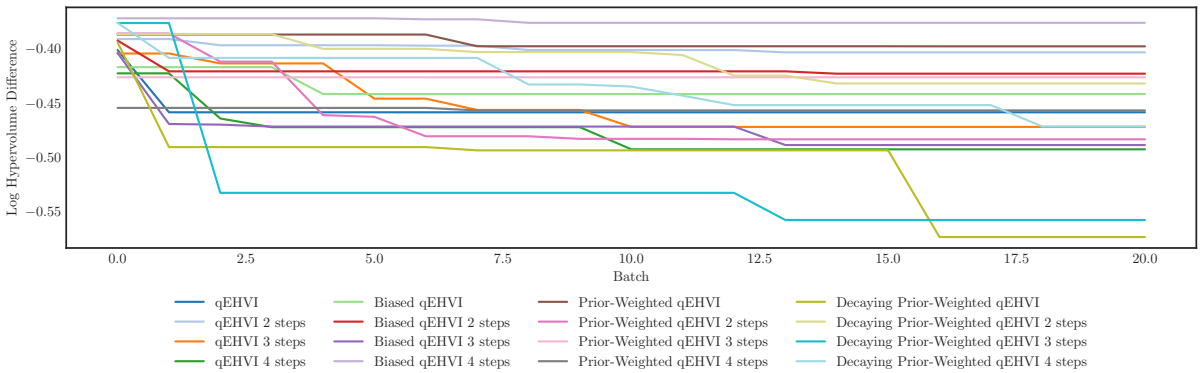
(b) DTLZ2



(c) DTLZ3



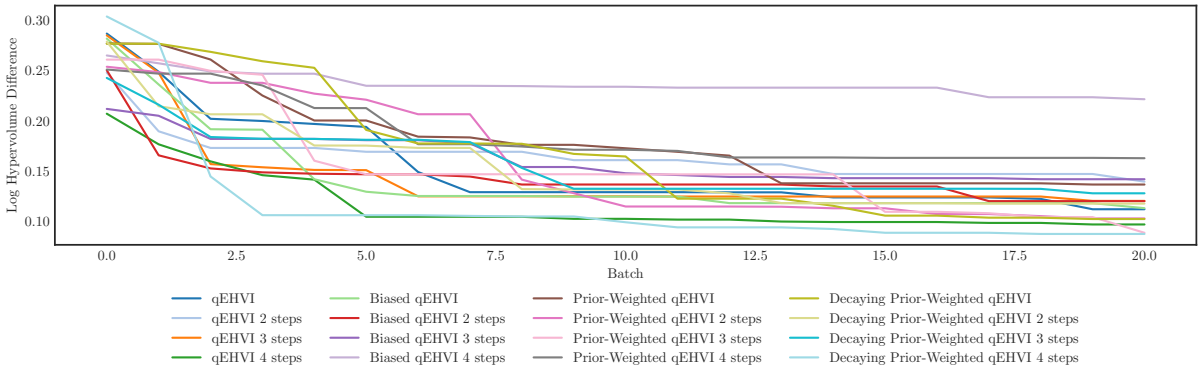
(d) DTLZ4



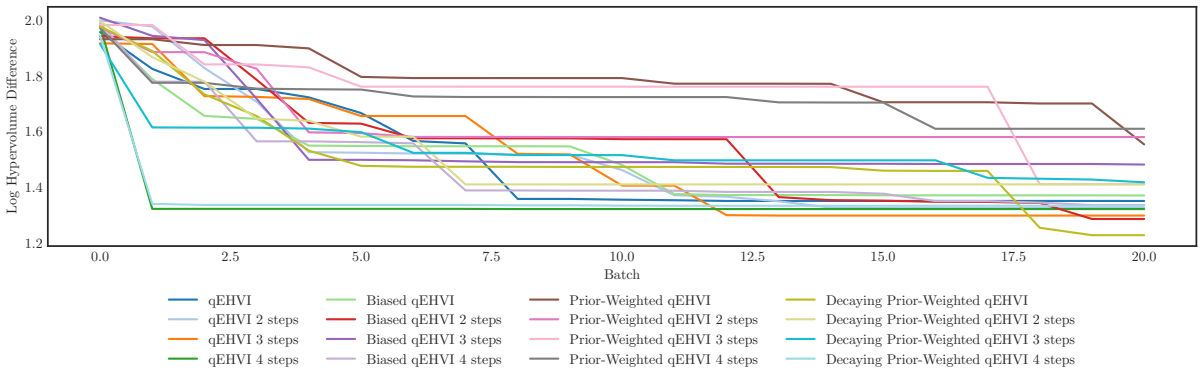
Source: Author.

Figure 31 – Performance of Acquisition Functions over Test Functions.

(a) DTLZ5



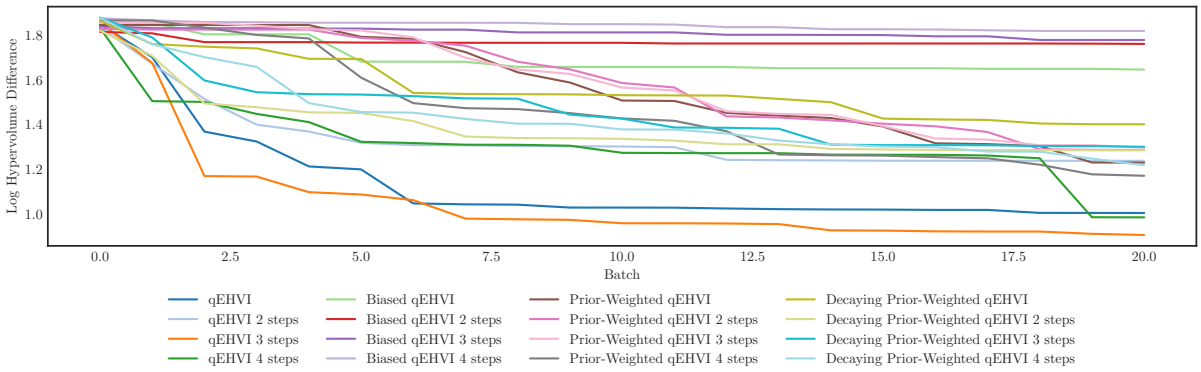
(b) DTLZ7



Source: Author.

Figure 32 – Performance of Acquisition Functions over real Vehicle Safety problem.

(a) Vehicle Safety



Source: Author.

The importance of this framework lies in its ability to manage processes that are both costly and time-consuming. With a limited number of possible experiments, it's important to select and conduct only highly informative experiments that yield maximum insights about the unknown function, which in this case is the real manufacturing process. Lookahead strategies are employed to address this challenge, allowing for the careful selection of an acquisition function, determination of the number of lookahead steps, and choice of the number of experimental points for each batch.

For this particular implementation, it will be conducted 10 experiments in each iteration. This number is reflective of a realistic manufacturing scenario, where time is allocated for both the production and quality measurement of lenses. Drawing from benchmark results presented in Section 3.3.2.2.7, qDecayWeightedEHVI have been selected as the acquisition function for the real optimization of lenses. A 4-step lookahead has been chosen for both functions, as they generally demonstrated the best performance.

To summarize, the optimization process involves the following steps: In each sequential iteration, 10 experiments will be selected for execution. The quality of each lens will be inspected, and sensor data will be collected. This information will be used to update the GP model for the next iteration. Additionally, each optimization will include a 4-step lookahead to evaluate the expected hypervolume improvement of each candidate point over the next four steps.

4 RESULTS AND DISCUSSION

In this chapter, the results from the experiments will be presented, and a discussion will follow, examining both the performance of the Machine Learning model and the Lookeahead Bayesian Optimization. The focus will be on determining if the results satisfy the previously established requirements.

4.1 MACHINE LEARNING MODEL

The Machine Learning model selected is the Lasso Regression, as detailed in Section 3.2.4. Table 5 displays the results for both targets. The Mean Absolute Errors values for both targets meet the requirements, demonstrating excellent performance with values significantly lower than 80 µm. The good performance can be attributed to the availability of a lot of experimental data and thorough data preparation prior to training, surpassing the previous requirements and models developed at Fraunhofer IPT. The Root Mean Squared Error is notably low, indicating the algorithm's effectiveness and the Mean Squared Error is also low, suggesting that the model's predictions closely align with the true values.

Additionally, the error metrics from both the validation (used in model training) and testing sets are similar, which indicates that the model is not overfitting. Notably, for the PV values, the testing set performs better than the validation set, suggesting that the model generalizes well to unseen data. In the case of RMS data, the testing set performs slightly worse than the validation set. However, the small difference implies that the model still generalizes effectively.

Metric	**PV**		**RMS**	
	Validation	**Testing**	**Validation**	**Testing**
Mean Absolute Error	1.7 µm	1.5 µm	1.1 µm	1.5 µm
Max Absolute Error	5.5 µm	3.6 µm	3.6 µm	4.6 µm
Mean Squared Error	4.4×10^{-6}	3.2×10^{-6}	2.2×10^{-6}	3.5×10^{-6}
Root Mean Squared Error	2.1 µm	1.7 µm	1.4 µm	1.8 µm

Table 5 – Comparison of Lasso Regression Model Metrics: Training vs Testing for Targets PV and RMS

4.2 INTEGRATION LOOKAHEAD BAYESIAN OPTIMIZATION WITH MACHINE LEARNING MODEL

The Lookahead Bayesian Optimization, as specified in Section 3.3.2.2.7, was conducted with 10 candidates for each iteration, looking four steps ahead at each iteration. The optimization is integrated with the surrogate (machine learning model), so when the optimization suggests the next points, the model promptly calculates their true values.

Even though a model simulates the actual process, the optimization process aims to mimic the real-world optimization, which is expensive, complex, and needs to achieve effective optimization early on. Consequently, the optimization was structured to include 20 batches. Each batch would assess 10 candidates, beginning with an initial set of 20 random experiments to initiate the optimization. This totals 220 experiments, intending to find the Pareto front of the complex process.

The optimization aimed to minimize both the PV and RMS values within these 220 experiments, continuously fine-tuning the Gaussian Process at each iteration.

Upon completing the 20 batches of optimization, Figure 33 illustrates the resulting Pareto front. The optimization achieved a well-distributed and comprehensive Pareto front, representing a diverse set of solutions. The progression of the Pareto front over the iterations is also evident, with the framework yielding favorable results from the early stages of optimization. Ultimately, the process succeeds in attaining minimum PV and RMS values of less than 0.0325 and 0.0045, respectively.

From the Pareto front, the optimal point was selected where both objectives are minimized effectively. This point had PV and RMS values of 31.4 µm and 4.7 µm, respectively. The evaluation of these results was compared to the best lens produced from 500 experiments using Latin Hypercube Sampling, as shown in Table 6. The Lookahead Bayesian Optimization framework demonstrated significant improvement in the PV values, indicating smoother peaks and valleys. However, the RMS value increased slightly, suggesting a higher overall shape deviation. Notably, these results were also achieved with less than 50% of the number of experiments originally performed using LHS.

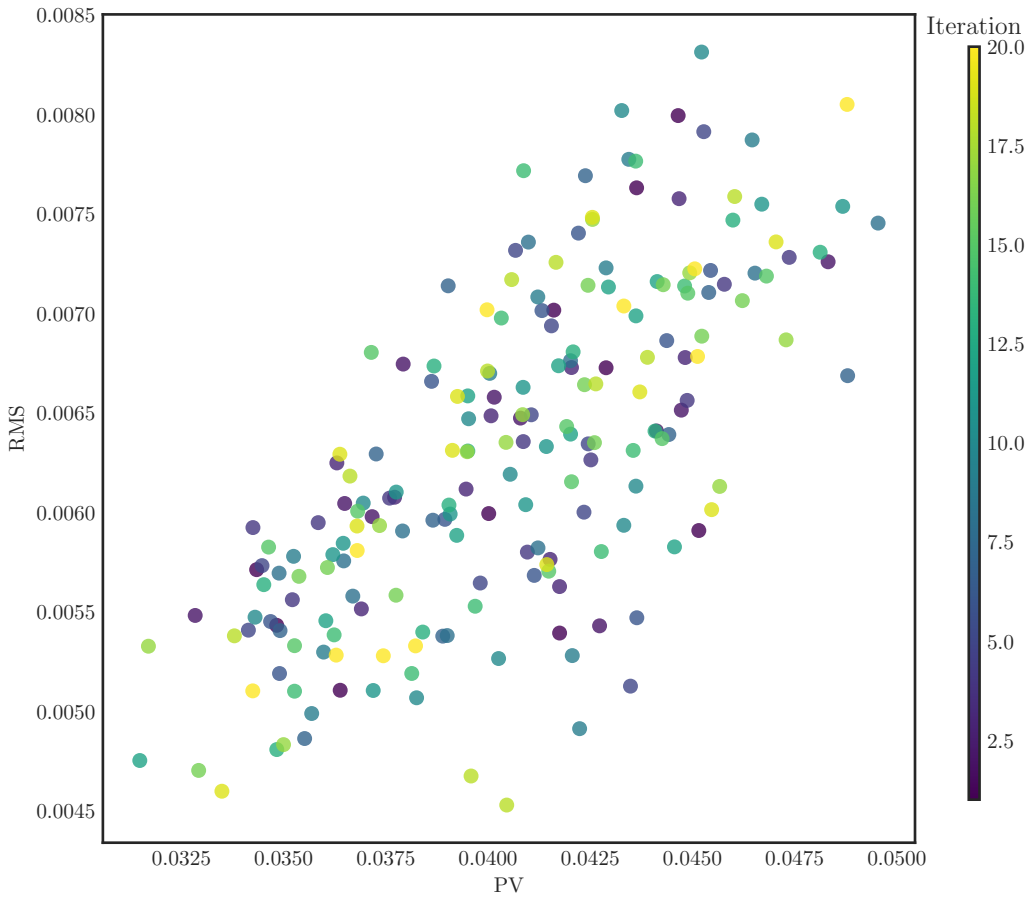
In the end, the optimization of the NGM process at Fraunhofer IPT yielded positive outcomes. The PV value obtained is considered excellent, representing a significant advancement. Meanwhile, the RMS value, though not as remarkable as the PV, is deemed acceptable and contributes positively to the overall improvement efforts. These results underscore the effectiveness of the optimization strategies employed, highlighting the exceptional performance of the PV value in particular.

Method	**PV**	**RMS**
BO	31.4 µm	4.7 µm
LHS	36.2 µm	2.2 µm

Table 6 – Comparison of results from Bayesian Optimization and Latin Hypercube Sampling.

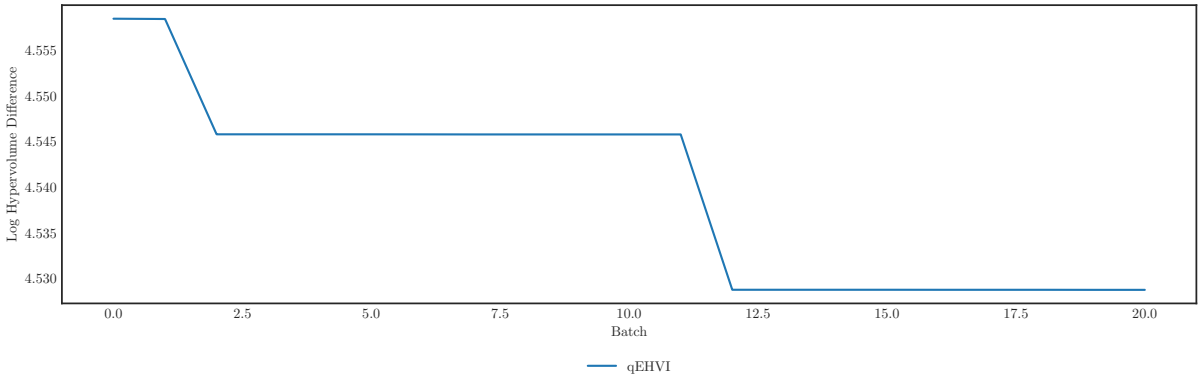
Additionally, the progress of the hypervolume during training can be observed in Figure 34. This figure highlights the improvement and development of the Pareto front throughout the 20 batches. There were several stages where the hypervolume remained constant, likely indicating exploration phases. Overall, there was significant progress across the 20 batches.

Figure 33 – Pareto front post optimization.



Source: Author

Figure 34 – Hypervolume post optimization.



Source: Author

5 CONCLUSION

The majority of industries involve significant trial and error to identify optimal parameters for product manufacturing. This challenge is present in the production of glass lenses using the non-isothermal process at Fraunhofer IPT, a method renowned for its cost efficiency due to the ability to produce lenses with different temperatures in the mold and glass. This thesis proposes a solution through the use of a Lookahead Bayesian Optimization framework, designed to sequentially conduct experiments and optimize the process with fewer experiments.

The effectiveness of this solution is assessed using a surrogate model that accurately mimics the real process. This approach provides flexibility for evaluating the framework and for future research. For the creation of the surrogate, a Latin Hypercube Sampling is performed to design the experiments to be done in the machine which carefully will represent how the process works. Following the completion of these experiments, lens measurements were taken, followed by extensive data preprocessing. Due to the use of time series data from sensors, it was crucial to extract important features, organize, and finally clean the data by removing outliers. A machine learning model employing Lasso regression was trained as the surrogate, surpassing performance requirements and accurately representing the process.

The Lookahead Bayesian Optimization Framework aims to reduce the number of experiments and accelerate production optimization. This is achieved by carefully selecting experimental points, balancing exploration and exploitation, and calculating the probable outcomes of each test. The framework requires the definition of two key components: the surrogate model and the acquisition function. A Gaussian Process is chosen as the surrogate model, and custom acquisition functions incorporating process expert knowledge are developed and tested against standard functions in benchmark tests. The number of lookahead steps for the optimization is also evaluated.

The optimization, conducted using the surrogate model for validation, involved a total of 200 tests, initiated with an initial set of 20 tests to establish the model. The chosen acquisition function for this test was qDecayWeightedEHVI, a custom function detailed in the thesis. This strategy led to significant improvements in peak-to-valley (PV) values, enhancing them from 36.2 µm to 31.4 µm. However, there was a slight increase in shape deviation (RMS), from 2.2 µm to 4.7 µm, when compared to the existing dataset. Notably, these results were achieved while reducing the number of experimental tests by more than 50%.

In conclusion, this project successfully achieved its objectives by developing an efficient Lookahead Bayesian Optimization framework, validated through benchmark testing and functional code. Additionally, a Machine Learning model acting as a surrogate of the process was created, demonstrating accurate performance. Ultimately, the real process

was optimized, achieving better shape accuracy with significantly fewer experiments.

Future Work

With the framework now validated through benchmark tests and using a surrogate model of the process, the next step could be to apply the framework to the actual process and evaluate its performance. The framework can also be applied to various processes, such as lens manufacturing using the PGM method, demonstrating its flexibility across different applications. This versatility means the framework could significantly benefit various industry sectors aiming to optimize products or processes. Potential areas of application include material science, where it can enhance design and reduce the number of required experiments; natural language processing, particularly in improving text extraction; and recommendation systems.

This work focused on creating innovative acquisition functions that incorporate process expert knowledge and using the Gaussian process as a surrogate model. Future research could explore other types of acquisition functions and other types of surrogate models that are not yet supported by BoTorch.

Additionally, while the current focus is on multiobjective optimization, the framework and custom acquisition functions could be expanded and tested for single-objective optimizations, opening new opportunities for research and development.

REFERENCES

ASTUDILLO, Raul; FRAZIER, Peter I. **Thinking inside the box: A tutorial on grey-box Bayesian optimization**. [S.l.: s.n.], 2022. arXiv: 2201.00272 [cs.LG].

BALANDAT, Maximilian et al. **BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization**. [S.l.: s.n.], 2020. arXiv: 1910.06403 [cs.LG].

BANKO, Michele; BRILL, Eric. Scaling to Very Very Large Corpora for Natural Language Disambiguation. In: PROCEEDINGS of the 39th Annual Meeting on Association for Computational Linguistics. Toulouse, France: Association for Computational Linguistics, 2001. (ACL '01), p. 26–33. DOI: 10.3115/1073012.1073017. Available from: https://doi.org/10.3115/1073012.1073017.

BLIEDTNER, Jens; GRÄFE, Günter. **Optiktechnologie: Grundlagen – Verfahren – Anwendungen – Beispiele**. 2., aktualisierte Auflage. [S.l.]: Carl Hanser Verlag GmbH & Co. KG, 2010. P. 419. ISBN 978-3-446-42215-5. DOI: 10.3139/9783446424661.

BOROVITSKIY, Viacheslav et al. **Matérn Gaussian processes on Riemannian manifolds**. [S.l.: s.n.], 2023. arXiv: 2006.10160 [stat.ML].

BOX, G. E. P.; COX, D. R. An Analysis of Transformations. **Journal of the Royal Statistical Society. Series B (Methodological)**, [Royal Statistical Society, Wiley], v. 26, n. 2, p. 211–252, 1964. ISSN 00359246. Available from: http://www.jstor.org/stable/2984418. Visited on: 17 Jan. 2024.

CHAPMAN, Peter. CRISP-DM 1.0: Step-by-step data mining guide. In. Available from: https://api.semanticscholar.org/CorpusID:59777418.

DEB, K. et al. Scalable multi-objective optimization test problems. In: PROCEEDINGS of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600). [S.l.: s.n.], 2002. 825–830 vol.1. DOI: 10.1109/CEC.2002.1007032.

DEUTSCH, Jared L.; DEUTSCH, Clayton V. Latin hypercube sampling with multidimensional uniformity. **Journal of Statistical Planning and Inference**, v. 142, n. 3, p. 763–772, 2012. ISSN 0378-3758. DOI: https://doi.org/10.1016/j.jspi.2011.09.016. Available from: https://www.sciencedirect.com/science/article/pii/S0378375811003776.

EPPS, R.W. et al. Artificial Chemist: An Autonomous Quantum Dot Synthesis Bot. **Adv. Mater.**, v. 32, n. 30, e2001626, July 2020. DOI: 10.1002/adma.202001626. eprint: 2020Jun. Available from: https://doi.org/10.1002/adma.202001626.

ERIKSSON, David et al. **Scalable Global Optimization via Local Bayesian Optimization**. [S.l.: s.n.], 2020. arXiv: 1910.01739 [cs.LG].

FIORE, Francesco Di; MAININI, Laura. **Non-Myopic Multifidelity Bayesian Optimization**. [S.l.: s.n.], 2022. arXiv: 2207.06325 [cs.LG].

FIORINI, Laura Battistella. **Creation and configuration of hybrid machine learning models for process optimization in the series production of complex optics**. July 2023. TCC (graduação) – Universidade Federal de Santa Catarina, Centro Tecnológico, Engenharia de Controle e Automação. TCC (graduação). Available from: https://repositorio.ufsc.br/handle/123456789/249030.

FRAUNHOFER. **Joseph von Fraunhofer**. [S.l.: s.n.], 2023. Accessed on 2024-01-28. Available from: https://www.fraunhofer.de/en/about-fraunhofer/profile-structure/chronicles/joseph-von-fraunhofer.html#1.

_____. **MP3 History**. [S.l.: s.n.], 2023. Accessed on 2024-01-28. Available from: https://www.mp3-history.com/.

_____. **Profile / structure**. [S.l.]: Fraunhofer, May 2023. Available from: https://www.fraunhofer.de/en/about-fraunhofer/profile-structure.html.

FRAZIER, Peter I. **A Tutorial on Bayesian Optimization**. [S.l.: s.n.], 2018. arXiv: 1807.02811 [stat.ML].

GARNETT, Roman. **Bayesian Optimization**. [S.l.]: Cambridge University Press, 2023.

GELMAN, Andrew; VEHTARI, Aki. **What are the most important statistical ideas of the past 50 years?** [S.l.: s.n.], 2021. arXiv: 2012.00174 [stat.ME].

GERON, Aurelien. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 2nd. [S.l.]: O'Reilly Media, Inc., 2019. ISBN 1492032646.

GRAND VIEW RESEARCH. **LED Lighting Market Size, Share And Growth Report, 2030**. [S.l.: s.n.], 2023. https://www.grandviewresearch.com/industry-analysis/led-lighting-market. Accessed: 11/2023.

GREENHILL, Stewart et al. Bayesian Optimization for Adaptive Experimental Design: A Review. **IEEE Access**, v. 8, p. 13937–13948, 2020. DOI: 10.1109/ACCESS.2020.2966228.

GRUVER, Nate et al. Effective Surrogate Models for Protein Design with Bayesian Optimization. In. Available from: https://api.semanticscholar.org/CorpusID:237361487.

HASAN, A S M Jahid; YUSUF, Jubair; FARUQUE, Rumana Binte. Performance Comparison of Machine Learning Methods with Distinct Features to Estimate Battery

SOC. In: 2019 IEEE Green Energy and Smart Systems Conference (IGESSC). [S.l.: s.n.], 2019. P. 1–5. DOI: 10.1109/IGESSC47875.2019.9042399.

HELTON, J.C.; DAVIS, F.J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. **Reliability Engineering System Safety**, v. 81, n. 1, p. 23–69, 2003. ISSN 0951-8320. DOI: https://doi.org/10.1016/S0951-8320(03)00058-9. Available from: https://www.sciencedirect.com/science/article/pii/S0951832003000589.

HOPE, Thomas M.H. Chapter 4 - Linear regression. In: MECHELLI, Andrea; VIEIRA, Sandra (Eds.). **Machine Learning**. [S.l.]: Academic Press, 2020. P. 67–81. ISBN 978-0-12-815739-8. DOI: https://doi.org/10.1016/B978-0-12-815739-8.00004-3. Available from: https://www.sciencedirect.com/science/article/pii/B9780128157398000043.

HUTTER, Frank; HOOS, Holger H.; LEYTON-BROWN, Kevin. Sequential Model-Based Optimization for General Algorithm Configuration. In_____. **Learning and Intelligent Optimization**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. P. 507–523.

HVARFNER, Carl et al. **πBO: Augmenting Acquisition Functions with User Beliefs for Bayesian Optimization**. [S.l.: s.n.], 2022. arXiv: 2204.11051 [cs.LG].

IPT, Fraunhofer. **Fraunhofer IPT**. [S.l.: s.n.], 2024. Accessed on 2024-01-28. Available from: https://www.ipt.fraunhofer.de/.

JIANG, Cheng et al. Simulation of the Refractive Index Variation and Validation of the Form Deviation in Precisely Molded Chalcogenide Glass Lenses (IRG 26) Considering the Stress and Structure Relaxation. **Materials**, v. 15, n. 19, 2022. ISSN 1996-1944. DOI: 10.3390/ma15196756. Available from: https://www.mdpi.com/1996-1944/15/19/6756.

JIANG, Shali et al. **Efficient Nonmyopic Bayesian Optimization via One-Shot Multi-Step Trees**. [S.l.: s.n.], 2020. arXiv: 2006.15779 [cs.LG].

JONES, Donald R.; SCHONLAU, Matthias; WELCH, William J. Efficient Global Optimization of Expensive Black-Box Functions. **Journal of Global Optimization**, v. 13, n. 4, p. 455–492, Dec. 1998. ISSN 1573-2916. DOI: 10.1023/A:1008306431147. Available from: https://doi.org/10.1023/A:1008306431147.

KHATAMSAZ, Danial et al. A physics informed bayesian optimization approach for material design: application to NiTi shape memory alloys. **npj Computational Materials**, v. 9, n. 1, p. 221, 2023. ISSN 2057-3960. DOI: 10.1038/s41524-023-01173-7. Available from: https://doi.org/10.1038/s41524-023-01173-7.

KUSHNER, Harold J. A versatile stochastic model of a function of unknown and time varying form. **Journal of Mathematical Analysis and Applications**, v. 5, n. 1, p. 150–167, 1962. ISSN 0022-247X. DOI: https://doi.org/10.1016/0022-247X(62)90011-2. Available from: https://www.sciencedirect.com/science/article/pii/0022247X62900112.

LEVY, Sigal; ADAMS, Ryan P.; STEINBERG, David M. Computer experiments: a review. **AStA Advances in Statistical Analysis**, 2010. DOI: https://doi.org/10.1007/s10182-010-0147-9.

LIAO, Xingtao et al. Multiobjective optimization for crash safety design of vehicles using stepwise regression model. **Structural and Multidisciplinary Optimization**, v. 35, n. 6, p. 561–569, 2008. ISSN 1615-1488. DOI: 10.1007/s00158-007-0163-x. Available from: https://doi.org/10.1007/s00158-007-0163-x.

LIM, Hyun-Il. A Linear Regression Approach to Modeling Software Characteristics for Classifying Similar Software. In: 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC). [S.l.: s.n.], 2019. P. 942–943. DOI: 10.1109/COMPSAC.2019.00152.

LIM, Yee-Fun et al. Extrapolative Bayesian Optimization with Gaussian Process and Neural Network Ensemble Surrogate Models. **Advanced Intelligent Systems**, v. 3, n. 11, p. 2100101, 2021. DOI: https://doi.org/10.1002/aisy.202100101. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aisy.202100101. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202100101.

LIU, Weidong. Precision glass molding: Toward an optimal fabrication of optical lenses. **Frontiers of Mechanical Engineering**, n. 5, p. 2017, 2017. DOI: 10.1007/s11465-017-0408-3. Available from: https://doi.org/10.1007/s11465-017-0408-3.

LOH, Wei-Liem. On Latin hypercube sampling. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 24, n. 5, p. 2058–2080, 1996. DOI: 10.1214/aos/1069362310. Available from: https://doi.org/10.1214/aos/1069362310.

MARTINEZ-CANTIN, Ruben; TEE, Kevin; MCCOURT, Michael. **Practical Bayesian optimization in the presence of outliers**. [S.l.: s.n.], 2017. arXiv: 1712.04567 [cs.LG].

MARTÍNEZ-PLUMED, Fernando et al. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. **IEEE Transactions on Knowledge and Data Engineering**, v. 33, n. 8, p. 3048–3061, 2021. DOI: 10.1109/TKDE.2019.2962680.

MCKAY, M. D.; BECKMAN, R. J.; CONOVER, W. J. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. **Technometrics**, [Taylor Francis, Ltd., American Statistical Association, American Society for Quality], v. 21, n. 2, p. 239–245, 1979. ISSN 00401706. Available from: http://www.jstor.org/stable/1268522. Visited on: 19 Dec. 2023.

MENDE, Hendrik et al. On the importance of domain expertise in feature engineering for predictive product quality in production. **Procedia CIRP**, v. 118, p. 1096–1101, 2023. 16th CIRP Conference on Intelligent Computation in Manufacturing Engineering. ISSN 2212-8271. DOI: https://doi.org/10.1016/j.procir.2023.06.188. Available from: https://www.sciencedirect.com/science/article/pii/S2212827123004158.

NASER, M. Z.; ALAVI, Amir H. Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences. **Architecture, Structures and Construction**, v. 3, n. 4, p. 499–517, Dec. 2023. ISSN 2730-9894. DOI: 10.1007/s44150-021-00015-8. Available from: https://doi.org/10.1007/s44150-021-00015-8.

NONISOTHERMAL glass molding for the cost-efficient production of precision freeform optics. [S.l.: s.n.], 2016. Accessed on 2024-01-28. DOI: 10.1117/1.OE.55.7.071207. Available from: https://publica.fraunhofer.de/handle/publica/246935.

PASZKE, Adam et al. **PyTorch: An Imperative Style, High-Performance Deep Learning Library**. [S.l.: s.n.], 2019. arXiv: 1912.01703 [cs.LG].

QU, G.; HARIRI, S.; YOUSIF, M. A new dependency and correlation analysis for features. **IEEE Transactions on Knowledge and Data Engineering**, v. 17, n. 9, p. 1199–1207, 2005. DOI: 10.1109/TKDE.2005.136.

RAMACHANDRAN, Anil et al. Incorporating expert prior in Bayesian optimisation via space warping. **Knowledge-Based Systems**, Elsevier BV, v. 195, p. 105663, May 2020. ISSN 0950-7051. DOI: 10.1016/j.knosys.2020.105663. Available from: http://dx.doi.org/10.1016/j.knosys.2020.105663.

RASMUSSEN, Carl Edward; WILLIAMS, Christopher K. I. **Gaussian Processes for Machine Learning**. [S.l.]: The MIT Press, Nov. 2005. ISBN 9780262256834. DOI: 10.7551/mitpress/3206.001.0001. Available from: https://doi.org/10.7551/mitpress/3206.001.0001.

RAY, Susmita. A Quick Review of Machine Learning Algorithms. In: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). [S.l.: s.n.], 2019. P. 35–39. DOI: 10.1109/COMITCon.2019.8862451.

SANTNER, Thomas J.; WILLIAMS, Brian J.; NOTZ, William I. **The Design and Analysis of Computer Experiments**. 1. ed. [S.l.]: Springer New York, NY, 2010. P. xii, 284. (Springer Series in Statistics). Springer Book Archive. ISBN

978-1-4419-2992-1. DOI: 10.1007/978-1-4757-3799-8. Available from: https://doi.org/10.1007/978-1-4757-3799-8.

SCHEWINSKI, Gustavo. **Bayesian Optimization visualizations code (GitHub).** [S.l.: s.n.], 2024. https://github.com/gustavoschewinski/bo-visualizations.

SCHULZ, Eric; SPEEKENBRINK, Maarten; KRAUSE, Andreas. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. **Journal of Mathematical Psychology**, v. 85, p. 1–16, 2018. ISSN 0022-2496. DOI: https://doi.org/10.1016/j.jmp.2018.03.001. Available from: https://www.sciencedirect.com/science/article/pii/S0022249617302158.

SHAHRIARI, Bobak et al. Taking the Human Out of the Loop: A Review of Bayesian Optimization. **Proceedings of the IEEE**, v. 104, n. 1, p. 148–175, 2016. DOI: 10.1109/JPROC.2015.2494218.

SHAN, Songqing; WANG, G. Gary. Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. **Structural and Multidisciplinary Optimization**, v. 41, n. 2, p. 219–241, Mar. 2010. ISSN 1615-1488. DOI: 10.1007/s00158-009-0420-2. Available from: https://doi.org/10.1007/s00158-009-0420-2.

SHIELDS, Michael D.; ZHANG, Jiaxin. The generalization of Latin hypercube sampling. **Reliability Engineering System Safety**, v. 148, p. 96–108, 2016. ISSN 0951-8320. DOI: https://doi.org/10.1016/j.ress.2015.12.002. Available from: https://www.sciencedirect.com/science/article/pii/S0951832015003543.

SMITH, Kirstine. On the Standard Deviations of Adjusted and Interpolated Values of an Observed Polynomial Function and its Constants and the Guidance they give Towards a Proper Choice of the Distribution of Observations. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 12, n. 1/2, p. 1–85, 1918. ISSN 00063444. Available from: http://www.jstor.org/stable/2331929. Visited on: 3 Jan. 2024.

SOLOMOU, Alexandros et al. Multi-objective Bayesian materials discovery: Application on the discovery of precipitation strengthened NiTi shape memory alloys through micromechanical modeling. **Materials Design**, v. 160, p. 810–827, 2018. ISSN 0264-1275. DOI: https://doi.org/10.1016/j.matdes.2018.10.014. Available from: https://www.sciencedirect.com/science/article/pii/S026412751830769X.

SOUZA, Artur et al. **Bayesian Optimization with a Prior for the Optimum**. [S.l.: s.n.], 2021. arXiv: 2006.14608 [cs.LG].

SPINA, Roberto et al. Analysis of lens manufacturing with injection molding. **International Journal of Precision Engineering and Manufacturing**, v. 13, n. 11, p. 2087–2095, Nov. 2012. ISSN 2005-4602. DOI: 10.1007/s12541-012-0276-z. Available from: https://doi.org/10.1007/s12541-012-0276-z.

TANABE, Ryoji; ISHIBUCHI, Hisao. An easy-to-use real-world multi-objective optimization problem suite. **Applied Soft Computing**, Elsevier BV, v. 89, p. 106078, Apr. 2020. ISSN 1568-4946. DOI: 10.1016/j.asoc.2020.106078. Available from: http://dx.doi.org/10.1016/j.asoc.2020.106078.

VIANA, Felipe A. C. A Tutorial on Latin Hypercube Design of Experiments. **Quality and Reliability Engineering International**, v. 32, n. 5, p. 1975–1985, 2016. DOI: https://doi.org/10.1002/qre.1924. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qre.1924. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.1924.

VU, Anh Tuan; GRUNWALD, Tim; BERGS, Thomas. Thermo-viscoelastic Modeling of Nonequilibrium Material Behavior of Glass in Nonisothermal Glass Molding. **Procedia Manufacturing**, v. 47, p. 1561–1568, 2020. 23rd International Conference on Material Forming. ISSN 2351-9789. DOI: https://doi.org/10.1016/j.promfg.2020.04.350. Available from: https://www.sciencedirect.com/science/article/pii/S2351978920314177.

VU, Anh Tuan; HELMIG, Thorsten, et al. Numerical and experimental determinations of contact heat transfer coefficients in nonisothermal glass molding. **Journal of the American Ceramic Society**, v. 103, n. 2, p. 1258–1269, 2020. DOI: https://doi.org/10.1111/jace.16756. eprint: https://ceramics.onlinelibrary.wiley.com/doi/pdf/10.1111/jace.16756. Available from: https://ceramics.onlinelibrary.wiley.com/doi/abs/10.1111/jace.16756.

VU, Anh-Tuan; KREILKAMP, Holger; DAMBON, Olaf, et al. Nonisothermal glass molding for the cost-efficient production of precision freeform optics. **Optical Engineering**, SPIE, v. 55, n. 7, p. 071207, 2016. DOI: 10.1117/1.OE.55.7.071207. Available from: https://doi.org/10.1117/1.OE.55.7.071207.

VU, Anh-Tuan; KREILKAMP, Holger; KRISHNAMOORTHI, Bharathwaj Janaki, et al. A hybrid optimization approach in non-isothermal glass molding. **AIP Conference Proceedings**, v. 1769, n. 1, p. 040006, Oct. 2016. ISSN 0094-243X. DOI: 10.1063/1.4963428. eprint: https://pubs.aip.org/aip/acp/article-pdf/doi/10.1063/1.4963428/13726112/040006_1_online.pdf. Available from: https://doi.org/10.1063/1.4963428.

WANG, Qianwen; MING, Yao, et al. ATMSeer: Increasing Transparency and Controllability in Automated Machine Learning. In: PROCEEDINGS of the 2019 CHI Conference on Human Factors in Computing Systems. [S.l.]: ACM, May 2019. (CHI '19). DOI: 10.1145/3290605.3300911. Available from: http://dx.doi.org/10.1145/3290605.3300911.

WANG, Xi; WANG, Chen. **Time Series Data Cleaning with Regular and Irregular Time Intervals**. [S.l.: s.n.], 2020. arXiv: 2004.08284 [cs.DB].

WILSON, James; HUTTER, Frank; DEISENROTH, Marc. Maximizing acquisition functions for Bayesian optimization. In: BENGIO, S. et al. (Eds.). **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2018. Available from: https://proceedings.neurips.cc/paper_files/paper/2018/file/498f2c21688f6451d9f5fd09d53edda7-Paper.pdf.

YANG, Kaifeng et al. **Efficient Computation of Expected Hypervolume Improvement Using Box Decomposition Algorithms**. [S.l.: s.n.], 2019. arXiv: 1904.12672 [cs.LG].

YI, Zeji et al. **Improving sample efficiency of high dimensional Bayesian optimization with MCMC**. [S.l.: s.n.], 2024. arXiv: 2401.02650 [cs.LG].

YUE, Xubo; KONTAR, Raed Al. **Why Non-myopic Bayesian Optimization is Promising and How Far Should We Look-ahead? A Study via Rollout**. [S.l.: s.n.], 2022. arXiv: 1911.01004 [cs.LG].

ZAMEE, Muhammad Ahsan; HAN, Dongjun; WON, Dongjun. Online Hour-Ahead Load Forecasting Using Appropriate Time-Delay Neural Network Based on Multiple Correlation–Multicollinearity Analysis in IoT Energy Network. **IEEE Internet of Things Journal**, v. 9, n. 14, p. 12041–12055, 2022. DOI: 10.1109/JIOT.2021.3133002.

ZHANG, Bin et al. Immune System Multiobjective Optimization Algorithm for DTLZ Problems. In: 2009 Fifth International Conference on Natural Computation. [S.l.: s.n.], 2009. P. 603–607. DOI: 10.1109/ICNC.2009.135.

ZHU, Ke Jun et al. Finite Element Analysis on Non-Isothermal Glass Molding. In: ULTRA-PRECISION Machining Technologies, CJUMP2011. [S.l.]: Trans Tech Publications Ltd, Apr. 2012. (Advanced Materials Research), p. 240–244. DOI: 10.4028/www.scientific.net/AMR.497.240.