Theses of the doctoral (PhD) dissertation

Model–Based Analysis and Optimization of Industrial MDI Manufacturing

Gergely Horváth

University of Pannonia

Doctoral School of
Chemical Engineering and Material Sciences

Supervisor:

Tamás Varga†, PhD

Alex Kummer, PhD

Department of Process Engineering

Veszprém

2026

1 Introduction and the aim of the work

The production of methylene diphenyl diisocyanate (MDI) is a cornerstone of polyurethane chemistry, with growing industrial importance due to its broad application range and increasing demand for high-performance polymer materials. MDI synthesis, however, remains a complex, multi-step process, characterized by intricate reaction networks, sensitive operational parameters, and environmental challenges. This dissertation presents a comprehensive, model-based and data-driven investigation of the industrial production of MDI, with particular focus on the synthesis and behavior of its primary precursor, methylenedianiline (MDA). The research encompasses both kinetic modeling and machine learning techniques to understand, predict, and optimize process behaviors. A detailed kinetic model was developed to describe the formation of MDA through the reaction of aniline and formaldehyde under acidic conditions. The model includes an extended reaction network and supports parameter sensitivity analysis, offering insight into the influence of molar ratios, residence times, and temperature profiles on product distribution, ring formation, and isomer ratios The identification and validation of model parameters was done using laboratory-scale experiments and available literature data demonstrating its reliability and adaptability across varying process conditions. In parallel, multiple soft sensor models were created to estimate complex product quality attributes, such as the color of MDI mixtures. These models employ advanced machine learning algorithms, including linear regression, regression trees, neural networks, support vector machines, and Gaussian process regression. Feature selection methods—such as Minimum Redundancy Maximum Relevance (MRMR), F-test method, ReliefF algorithm, Correlation-based and Aggregated techniques — were applied to identify the most influential process variables, while time-delay analysis and Bayesian optimization were used to enhance model

accuracy and robustness. Among the models, Gaussian Process Regression showed superior predictive performance and the optimal operating parameters were determined by combining the model with Genetic Algorithm. The integration of mechanistic and data-driven modeling approaches enables the identification of optimal production conditions, reduction of undesirable by-products, and the development of explainable, interpretable models aligned with industrial requirements. These outcomes contribute to the advancement of sustainable, efficient, and intelligent MDI production technologies.

2 Theses

Thesis #1. I expanded the reaction system of existing models reported in the literature for the synthesis of Methylenedianiline by incorporating additional components and reaction pathways, I achieved enhanced model accuracy and reduced information loss.

- I created an extended reaction network by incorporating three additional components – N-methylbenzene, aniline, and formaldehyde – enabling the usage of the aniline/formaldehyde molar ratio, an industrial production parameter, through the kinetic model.
- The incorporation of N-methylbenzene and the extended reaction network enabled to accurately represent the formation of molecules containing an odd number of rings, thereby overcoming the limitation of prior models which were constrained to even-ring structures.
- I demonstrated that the developed extended reaction network including new reaction pathways and components more accurately describe the Methylenedianiline synthesis process, as evidenced by improved fit with both concentration trajectories and key performance indicators.

Related publications: 1, 4

Thesis #2. I justified that machine learning models efficiently support the understanding of Methylenedianiline synthesis variables on laboratory-scale, aiming to examine the impact of variables on Methylenedianiline quality and to assist large-scale industrial production.

- I determined the key synthesis parameters and their impact with Shapley values for all Methylenedianiline quality parameters, demonstrating that the increase of HCl / aniline molar ratio or water / aniline molar ratio increases by-product and P-P isomer formation, while a higher aniline / formaldehyde molar ratio considerably decreases them. The 2-ring content and O-P isomer content can be increased with higher aniline / formaldehyde molar ratio, lower condensation reaction time and temperature, however, adjusting these parameters in the opposite way results in an increased amount of polymer products
- I implemented a new, three-level hierarchical method for outlier analysis, designed to pinpoint the outliers of laboratory measurements with optimal efficiency.

Related publications: 2, 5, 6

Thesis #3. Through the proposed intervention strategies, I justified that soft-sensor models can be well applied for Methylenediphenyl diisocyanate color prediction based on real industrial data in an explainable way, illustrated the non-linear behavior of the system and identified the best operational parameters for an industrial Methylenediphenyl diisocyanate production plant.

- I proposed 3 critical intervention strategies within the framework of the industrial production to reduce side reactions: avoid the formation of urea compounds, limit the formation of chloroformamidine-N-carbonyl chloride compounds, and prevent the breakdown of chloroformamidine-N-carbonyl chloride compounds into dichlorides.
- I proved that the accuracy of the machine learning models can be improved by implementing time delays for each industrial operational parameters, which were determined with the use of correlation analysis between input and output data
- I verified the significant non-linearity and complexity of the system with the visualisation of the relationships between each feature on Partial Dependence Plots, which demonstrates that each parameter has local optima depending on the values of other investigated operational parameter and they are highly cross-correlated with each other.
- I identified an explainable and optimal set of operational parameters for the industrial system, achieving an Methylenediphenyl diisocyanate color value close to 1.06 at a normalized level exceeding the initially explored range.

Related publications: 3, 7, 8

3 Summary

The dissertation presents a comprehensive, model-driven and data-informed exploration of the industrial production of methylene diphenyl diisocyanate (MDI), with particular emphasis on the synthesis and mechanistic understanding of methylenedianiline (MDA), the principal precursor to MDI. Through the integration of kinetic modeling, machine learning methodologies, and optimization techniques, a both mechanistic and data-driven approaches were developed to enhance the predictability, efficiency, and controllability of MDI manufacturing processes.

A detailed kinetic model was constructed to describe the complex reaction network of MDA synthesis, accounting for critical intermediates, such as ortho- and para-aminobenzylanilines (ABAs), and their transformation into MDA isomers. The model successfully incorporated extended reaction pathways and was validated using both laboratory-scale experiments and available literature data.

The application of Machine Learning models has shown through sensitivity analyses that synthesis parameters – including molar ratios, residence times, and temperature profiles –significantly impact both isomeric distribution and ring structure. These insights allow for the precise identification of operational conditions that promote the formation of preferred isomers, such as 4,4'-MDA, and control by-product formation, ultimately enhancing the quality of downstream MDI products.

To address limitations in conventional modeling approaches, particularly for parameters such as MDI product color, a suite of soft sensors was developed using advanced machine learning algorithms. Techniques including Linear Regression, Regression Trees, Neural Networks, Support Vector Machines, and Gaussian Process Regression were employed. Feature selection methods such as MRMR, ReliefF, F-test, Correlation-based analysis and an Aggregated approach were combined with time-delay estimation to identify the most impactful process variables. GPR emerged as the most accurate and interpretable modeling technique, demonstrating robust performance even under complex, nonlinear conditions.

An optimization process of industrial operational parameters was accomplished through the integration of the best performing trained and validated Gaussian Process Regression model with Genetic Algorithms. The deployment of Genetic Algorithms, based on the superior predictive capacity of the most effective GPR model, enabled the identification of the optimal operational parameters for a real industrial facility to improve the color quality of the MDI product mixtures.

Altogether, this work exemplifies the synergy between mechanistic and data-driven modeling. It advances the state-of-the-art in chemical process engineering by delivering validated, explainable, and scalable tools for real-time process optimization in industrial MDI production. The methodologies developed herein are adaptable and hold promise for broader application across other complex, multi-step reaction systems within the chemical manufacturing domain.

The outcomes of this research not only contribute to enhanced product quality and process sustainability but also align with the strategic goals of digital transformation and Industry 4.0 within the chemical industry.

4 Publications related to theses

Articles in international journals

1. Horváth, G., Kummer, A., Kozár, Z., & Varga, T. (2023). Exploration and model-based analysis of reaction mechanisms related to the formation of methylenedianiline. *Industrial & Engineering Chemistry Research*, *62*(10), 4297-4311.
   Scimago Journal Ranking: Q1, Impact Factor: 4.009
2. Horváth, G., Trujillo, V. J., Réti, J., Kozár, Z., Kummer, A., & Varga, T. (2024). Exploring the essential features influencing the synthesis of methylenedianiline to support industrial processes. *Chemical Engineering Research and Design*, *208*, 626-647.
   Scimago Journal Ranking: Q2, Impact Factor: 4.345
3. Horváth, G., Trujillo, V. J., Réti, J., Kozár, Z., Varga, T., & Kummer, A. (2025). Soft-sensor development for product quality estimation with time delay and feature selection in industrial MDI production. *Chemical Engineering Journal Advances*, *22*, 100751.
   Scimago Journal Ranking: Q1, Impact Factor: 8.093

Conference abstracts

4. Horváth, G., Varga, T., & Kummer, A. (2022). Exploration and model-based analysis of reaction mechanisms related to the

formation of methylenedianiline, *50. Műszaki Kémiai Napok Jubileumi Konferencia: 50th Engineering Chemistry Jubilee Conference*, Veszprém, Hungary, pp. 30

5. Horváth, G., Trujillo, V. J., Réti, J., Kozár, Z., Varga, T., & Kummer, A. (2023). Development and comparison of machine learning models to analyse the synthesis of methylenedianiline, *PhD hallgatók anyagtudományi napja XXIII: Materials science day XXIII*, Veszprém, Hungary, pp. 5
6. Horváth, G., Varga, T., & Kummer, A. (2023). Data-driven modeling of methylenedianiline (MDA) synthesis, *Műszaki Kémiai Napok 2023: Engineering Chemistry Conference 2023*, Veszprém, Hungary, pp. 43
7. Horváth, G., Varga, T., & Kummer, A. (2024). Identification and model-based analysis of color problems in the synthesis of methylenediphenyl diisocyanate products, *Műszaki Kémiai Napok 2024: Engineering Chemistry Conference 2024*, Veszprém, Hungary, pp. 41
8. Horváth, G., Trujillo, V. J., Réti, J., Kozár, Z., Varga, T., & Kummer, A. (2024). Machine learning-based product quality estimation with time delay-based feature selection in industrial MDI production, *PhD hallgatók anyagtudományi napja XXIV: Materials science day XXIV*, Veszprém, Hungary, pp. 13