mgr inż. Tomasz Zema

Doctoral dissertation abstract

Deep Learning-Based Motor Fuel Sales Forecasting in Enterprises

Supervisor:
dr hab. inż. Marcin Hernes, prof. UEW

Co-supervisor:
dr inż. Adam Sulich, prof. UEW

Wrocław 2026

Contents

Scientific problem, gaps and research questions 2
Research Objectives.....5
Research Methodology.....6
Research Results.....8
Practical Implications.....12
Conclusions.....14

Scientific problem, gaps and research questions

The contemporary economy, undergoing dynamic digital transformation and increasing pressure for operational efficiency, presents enterprises in the fuel sector with challenges of a scale that is increasingly difficult to address through intuitive management (Brynjolfsson et al., 2011). Growing competition, volatile market conditions, and long-term processes associated with the energy transition increase the cost of erroneous decisions in areas such as inventory planning, delivery organization, and sales management (Syntetos et al., 2009). Under these circumstances, classical forecasting approaches that are based on simple statistics or limited classes of econometric models. Traditional forecasting models often prove inadequate in the face of the complexity of seasonality and nonlinear dependencies that determine demand for motor fuels (Hyndman & Athanasopoulos, 2018).

Fuel industry enterprises operate within complex business ecosystems characterized by intensive network relationships among value chain participants (Adner, 2017). A fuel station plays the role of an intermediary in this ecosystem, connecting midstream and downstream segments, which corresponds to the concept of organizational boundaries in the transaction cost theory of Coase and Williamson (Coase, 1937; Williamson, 1985). Consequently, uncertainty regarding future demand for motor fuels generates significant transaction costs associated with negotiating deliveries, maintaining excess inventory, and responding to unforeseen fluctuations in consumption (Ketokivi & Mahoney, 2020).

The concept of a business ecosystem (Moore, 1993) assumes that enterprises function as elements of complex inter-organizational networks (Stańczyk, 2019; Zema et al., 2024),

comprising upstream (refineries), midstream (distributors), and downstream (fuel stations) segments (Dong & Li, 2021), where information quality determines value chain efficiency (Dekker, 2003). Transaction cost theory indicates that organizations strive to minimize coordination costs (North, 1991), which in fuel distribution translates into reducing information uncertainty through precise sales forecasting. According to Williamson's (1985) framework, these costs include search and information costs (market monitoring), negotiation costs (delivery contracting), and monitoring costs (quality control and auditing).

Research Problem and gaps

The research problem of this dissertation is the improvement of the motor fuel sales volume forecasting process in enterprises using deep learning. Contemporary fuel industry enterprises operate under conditions of increasing market uncertainty, energy transition, and pressure for sustainable development, which makes precise sales forecasting a foundation for effective inventory management and supply chain optimization. A systematic literature review enabled the identification of four significant research gaps that justify the undertaking of this dissertation topic:

- **The taxonomic gap** is expressed by the lack of systematization of concepts and relationships concerning the functioning of fuel industry business ecosystems in the context of forecasting processes. The literature lacks a systematic approach to the relationships between the structure of fuel distribution micro-networks and the possibilities of applying advanced predictive methods. Existing studies concentrate on individual aspects (e.g., oil price modeling, delivery route optimization), not offering an integrated view of forecasting as an element of the business ecosystem.
- **The explanatory gap** consists of insufficient knowledge about forecasting mechanisms using deep learning in the context of data specifics from automated fuel dispensers and fuel sales systems in enterprises. Previous research has focused mainly on crude oil price forecasting at the macroeconomic level, neglecting the issue of sales volume forecasting at the distribution point level. There is a lack of explanation regarding which neural network architectures handle the multi-seasonality of sales data best and what mechanisms (dilation, attention mechanism, basis decomposition) determine a model's generalization ability.
- **The empirical gap** results from the lack of research using real operational data from fuel distribution enterprises. Most existing studies are based on synthetic data or commodity

market indicators (e.g., Brent, WTI stock quotations), which limits the possibility of drawing practically relevant conclusions. Data from automated fuel dispensers are characterized by specific patterns: strong weekly seasonality, the influence of holidays and local events, and the potential impact of the COVID-19 pandemic on consumption patterns.

- **The methodological gap** manifests as insufficient accuracy of existing motor fuel sales forecasting methods, particularly in the context of modeling complex multi-seasonality (weekly, monthly, annual) and nonlinear dependencies characteristic of high-frequency sales data (Ahaggach et al., 2024). Traditional ARIMA or Holt-Winters models assume linearity of relationships and limited seasonal structure, which may be inadequate given the real complexity of the data (Hewamalage et al., 2021).

The identified research gaps indicate the need to develop a comprehensive method for motor fuel sales forecasting that integrates a theoretical perspective (business ecosystem, transaction costs) with an empirical approach based on deep learning. This is an issue significant both for the development of the management and quality sciences discipline and for the practice of fuel industry enterprise operations. To systematically address the identified research problem, seven research questions were formulated and assigned to individual chapters of the dissertation:

RQ1: How do business ecosystem theory and transaction cost theory explain the specifics of fuel industry enterprise operations and justify the need for motor fuel sales forecasting?

This question refers to the theoretical foundations of the dissertation and requires an analysis of the relationship between the fuel industry ecosystem structure and market uncertainty costs. The answer to this question was formulated in Chapter 1.

RQ2: What motor fuel sales forecasting methods are used in the literature and what are their main limitations in the context of enterprise data utilization?

Addressing this question required conducting a systematic literature review using bibliometric tools. The analysis results are presented in Chapter 2.

RQ3: Which deep learning architectures are most suitable for modeling fuel sales time series and what mechanisms determine their generalization ability?

This question concerns the theoretical foundations of deep learning in the context of time series forecasting. The answer was developed in Chapter 3.

RQ4: What transformation of experimental data will enable the application of deep learning in fuel sales forecasting?

This question focuses on methodological aspects related to data preprocessing and feature engineering. Results are presented in Subsection 4.1.

RQ5: What deep learning models should be applied to increase forecast accuracy compared to existing methods?

Addressing this question required conducting comparative studies of ten deep learning architectures against three traditional methods. Comparison results are provided in subsections 4.2–4.6.

RQ6: What hyperparameter values of a deep neural network will enable achieving the highest level of accuracy?

This question concerns model configuration optimization. Hyperparameter space search results are presented in Subsections 4.3–4.5.

RQ7: How will the developed deep learning models improve motor fuel sales processes?

This question refers to the practical implications of the research findings. The answer was formulated in Subsection 4.7.

Research Objectives

The main objective of the doctoral dissertation was defined as the development and evaluation of motor fuel sales forecasting models in enterprises using deep learning. Achieving the main objective required the realization of specific objectives, which were divided into three groups:

Cognitive Objectives

1. Understanding the functioning of fuel industry enterprises from the perspective of business ecosystem theory and transaction cost theory, with particular emphasis on the importance of sales forecasting for optimizing decision-making processes.
2. Systematizing knowledge regarding motor fuel sales forecasting based on a literature review supported by bibliometric analysis (Wohlin, 2014; Paul & Criado, 2020).

3. Systematizing knowledge regarding deep learning in the context of fuel distribution and sales, with particular emphasis on architectures capable of modeling multi-seasonal time series patterns.

Methodological Objective

Development of methods for data preprocessing, building, and evaluating motor fuel sales forecasting models in enterprises using deep learning. This objective included: (1) developing a procedure for transforming raw transactional data, (2) constructing exogenous variables (calendar and price-based), (3) implementing ten deep learning architectures, (4) developing a validation strategy based on the rolling origin procedure.

Utilitarian Objective

Conducting an effectiveness assessment of the developed motor fuel sales forecasting models using real data and formulating practical implementation recommendations for fuel industry enterprises.

Research Methodology

The research methodology carried out as part of the dissertation consists of the following stages:

1. **Theoretical analysis of fuel industry enterprise operations.** Within the framework of this stage, the functioning of fuel industry organizations was analyzed from the perspective of business ecosystem theory (Moore, 1993; Adner, 2017) and transaction cost theory (Coase, 1937; Williamson, 1985). The research method of literature review was used, including both Polish-language and foreign sources. The results of the work from this stage are presented in Chapter 1 of the dissertation.

2. **Literature review on fuel sales forecasting.** In the next stage, a hybrid literature review supported by bibliometric analysis was conducted using the Scopus database and VOSviewer and Bibliometrix tools (Aria & Cuccurullo, 2017; Donthu et al., 2021). The results of the work from this stage are presented in Chapter 2.

3. **Characterization of deep learning in the context of time series forecasting.** In this phase, a systematic review of deep neural network architectures was conducted in terms of

their suitability for modeling fuel sales time series. Four classes of models were characterized: (a) recurrent networks—RNN, LSTM, GRU, DilatedRNN (Goodfellow et al., 2016; Cho et al., 2014; Chang et al., 2017); (b) temporal convolutional networks—TCN, BiTCN (van den Oord et al., 2016); (c) basis expansion models—N-BEATS, N-HiTS (Oreshkin et al., 2020; Challu et al., 2022); (d) attention-based architectures—TFT, PatchTST (Lim et al., 2021; Nie et al., 2023). The results of the work from this stage are presented in Chapter 3.

4. **Data preprocessing.** An anonymized dataset from an enterprise manufacturing automated fuel dispensers was used, comprising 492,178 transactional records from the 2017–2020 period, aggregated to 1,177 daily observations. Preprocessing included: data cleaning and modal imputation, stationarity diagnostics using ADF and KPSS tests, multi-seasonal MSTL decomposition (periods: 7, 14, 30, 60/90, 365 days), calendar variable construction, and exogenous variable preparation (WTI oil price changes). Data were split chronologically: 90% training, 10% test (forecast horizon H=30 days).

5. **Deep learning model development and training.** Ten deep learning architectures were implemented (DilatedRNN, N-BEATSx, BiTCN, LSTM, GRU, RNN, TCN, N-HiTS, TFT, PatchTST) using the NeuralForecast library (Nixtla, 2025) with PyTorch 2.0 as the computational engine. The regularization strategy included: dropout (Srivastava et al., 2014), early stopping, and L2 normalization. Hyperparameter optimization was performed via Ray/Tune framework with HyperOpt algorithm (Bayesian TPE). Three traditional methods served as baseline models: SeasonalNaive, AutoARIMA, and Holt-Winters.

6. **Validation and statistical analysis.** A multi-level validation strategy was applied: (a) out-of-time validation on a 30-day test window; (b) rolling origin cross-validation with 10 windows and step of 30 days (Bergmeir & Benítez, 2012; Tashman, 2000); (c) paired tests including Wilcoxon Signed-Rank and Diebold-Mariano (Diebold & Mariano, 2002); (d) Friedman test with Nemenyi post-hoc procedure (Demšar, 2006); (e) Holm correction for multiple comparisons; (f) effect size analysis using Cohen's d (Cohen, 1988); (g) residual analysis with Shapiro-Wilk and Ljung-Box tests.

Research Results

The conducted experiments demonstrated that all ten deep learning architectures outperform traditional statistical methods. Table 1 presents the results of deep learning architectures and baseline models, ordered by increasing MAE error.

Based on the results presented in Table 1, the following observations can be formulated. All ten deep learning architectures outperform the SeasonalNaive reference model, reducing MAE error by 49–61% (194–242 liters). The best results were achieved by DilatedRNN (MAE = 155.5 L, $R^2 = 0.814$) and BiTCN (MAE = 161.0 L, $R^2 = 0.809$) models, which also exhibit high stability (SI ≥ 3). Analysis of traditional model results reveals that AutoARIMA achieves the lowest MAE error (320.45 liters) among classical methods, demonstrating the effectiveness of automatic ARIMA parameter selection considering weekly seasonality (s=7). The DilatedRNN model's advantage over AutoARIMA is 36% in terms of explained variance ($R^2 = 0.814$ vs $R^2 = 0.60$). The DilatedRNN_Ctx30_NoOil model achieved the highest forecast accuracy among all tested models. The DilatedRNN architecture extends classical recurrent networks with a dilation mechanism that enables processing sequences at different time scales without significantly increasing the number of parameters (Chang et al., 2017).

Table 1. Fuel sales forecasting model ranking.

Rank	Model	Type	RMSE [L]	MAE [L]	R^2	MAE_rel
1	DilatedRNN	DL	196.47	155.49	0.814	-61%
2	BiTCN	DL	199.01	160.96	0.809	-59%
3	N-BEATSx	DL	222.69	169.26	0.778	-57%
4	N-HiTS	DL	234.72	172.05	0.770	-57%
5	TFT	DL	220.59	178.17	0.766	-55%
6	LSTM	DL	241.67	178.45	0.752	-55%
7	GRU	DL	228.03	183.87	0.754	-54%
8	PatchTST	DL	232.85	186.55	0.739	-53%
9	TCN	DL	243.71	196.87	0.711	-50%
10	RNN	DL	233.82	203.25	0.683	-49%
11	AutoARIMA	Trad.	440.53	320.45	0.60	-19%

12	Holt-Winters	Trad.	453.96	357.27	0.58	-10%
13	SeasonalNaive	Trad.	531.72	397.37	0.42	0%

Source: Author's elaboration. MAE_rel = percentage MAE change vs SeasonalNaive.

The main element of the DilatedRNN architecture is the dilation mechanism (dilations = [[1, 2], [4, 8]]), which enables processing sequences at different time scales. The first dilation level ([1, 2]) handles short-term dependencies (1–2 days), while the second level ([4, 8]) models weekly patterns without excessively increasing the number of parameters. To explain the best results, a detailed ablation study was conducted, testing 10 combinations of exogenous variables. Results are presented in Table 2.

Table 2. Ablation study of exogenous variable impact on DilatedRNN model.

Configuration	**Exogenous variables**	**MAE [L]**	**RMSE [L]**	**R^2**	**ΔR^2**
NoOil	Calendar only	155.49	196.47	0.814	(REF)
Cyclic	Calendar + harmonic encoding	186.46	237.65	0.728	−0.086
Lag7Only	Calendar + y_lag7	170.31	240.26	0.722	−0.092
FutrWithOil	Calendar + oil price	175.66	214.03	0.780	−0.034
RollOnly	Calendar + rolling stats	171.84	214.79	0.778	−0.036
Both	Calendar + historical lags	189.91	264.66	0.663	−0.151
CyclicAdv	Calendar + harmonic + adv_weekend	224.17	291.18	0.592	−0.222
AdvWeekend	Calendar + adv_weekend	256.17	335.51	0.458	−0.356
HistOnly	Historical lags only	204.90	247.81	0.704	−0.110
All	All variables	244.80	303.11	0.558	−0.25

Source: Author's elaboration. ΔR^2 = change relative to NoOil configuration.

The ablation study results presents that configuration using only calendar variables (8 features: day of week, month, year, quarter, weekend, holidays) achieves the best results (R^2 = 0.814, MAE = 155.49 L). Adding any additional variables worsens forecast accuracy, suggesting that the DilatedRNN architecture effectively extracts seasonal patterns directly from raw calendar

encodings. Advanced weekend variables (adv_weekend) drastically worsen results, lowering R^2 from 0.814 to 0.458 (a drop of 35.6 percentage points). The probable cause is the introduction of noise through redundant differentiation of Saturdays and Sundays and seasonal interactions. Adding the oil price lowers R^2 from 0.814 to 0.780 (a drop of 3.4 percentage points), which may result from the lack of direct correlation between global commodity prices and local fuel demand at the individual station level.

Statistical Analysis

To ensure rigorous verification of observed differences between models, multi-level statistical analysis was conducted (Table 3). All tests were performed at a significance level of α=0.05 with Holm correction for multiple comparisons.

Table 3. Statistical test results: DL models vs SeasonalNaive.

Model	**ΔMAE [L]**	**p (Wilcoxon)**	**p (D-M)**	**Cohen's d**	**Holm p**
DilatedRNN	-242.1	0.0002	0.0003	-0.64	0.0021
BiTCN	-236.6	0.0002	0.0003	-0.63	0.0021
N-BEATSx	-228.4	0.0004	0.0005	-0.58	0.0036
N-HiTS	-225.5	0.0005	0.0006	-0.56	0.0040
TFT	-219.4	0.0008	0.0009	-0.54	0.0056
LSTM	-219.1	0.0008	0.0009	-0.53	0.0056
GRU	-213.7	0.0012	0.0014	-0.51	0.0072
PatchTST	-211.0	0.0015	0.0018	-0.50	0.0090
TCN	-200.7	0.0025	0.0030	-0.49	0.0100
RNN	-194.3	0.0035	0.0042	-0.48	0.0105

Source: Author's elaboration. ΔMAE = difference vs SeasonalNaive (397.37 L). Holm p = p-value after multiple comparison correction.

The results analysis leads to the following conclusions. All ten deep learning models demonstrate statistically significant superiority over SeasonalNaive in both the Wilcoxon test ($p < 0.008$) and the Diebold-Mariano test ($p < 0.009$). After applying the Holm correction, all models maintain statistical significance ($p_{Holm} < 0.01$). The effect size (Cohen's d) for all models falls within the medium range ($|d| = 0.48–0.64$), indicating practically significant differences in the business application context. For global comparison of all thirteen models, the Friedman test was

applied (Friedman, 1937). The Friedman test statistic was $\chi^2 = 39.99$ with $p = 0.0001$, indicating a significant difference between models at the $\alpha = 0.001$ level.

Cross-Validation

To verify generalization ability, rolling origin cross-validation was conducted on the full dataset (10 windows, step 30 days, expanding window), following best practices for time series model evaluation (Bergmeir & Benítez, 2012; Tashman, 2000). Table 4 presents results for DL models.

Table 4. Cross-validation results for deep learning models.

Rank	Model	MAE [L]	SD [L]	R^2
1	TCN	269.54	44.65	0.714
2	BiTCN	287.23	56.82	0.689
3	GRU	301.45	52.89	0.661
4	LSTM	315.67	53.41	0.642
5	TFT	326.47	67.23	0.618
6	N-BEATSx	334.89	78.56	0.598
7	RNN	345.12	89.34	0.571
8	DilatedRNN	353.08	95.67	0.555
9	PatchTST	374.25	102.45	0.512
10	N-HiTS	419.73	122.93	0.428

Source: Author's elaboration. SD = standard deviation of MAE across windows.

The Table 5 shows model ranking by stability based on Cross-validation results. Cross-validation presented significant differences compared to single-window tests. The TCN model, which ranked 9th in the manual test (MAE = 197 L), achieved the lowest average MAE (269.54 L) in cross-validation with the lowest standard deviation ($\sigma = 44.65$ L). Conversely, DilatedRNN, despite the best results on a single window (MAE = 155.5 L), ranked 8th in cross-validation with MAE = 353.08 L.

Table 5. Model ranking by stability (MAE standard deviation).

Rank	Model	SD [L]	Coefficient of variation
1	TCN	44.65	16.6%
2	GRU	52.89	17.5%
3	LSTM	53.41	16.9%
4	BiTCN	56.82	19.8%
5	TFT	67.23	20.6%
6	N-BEATSx	78.56	23.5%
7	RNN	89.34	25.9%
8	DilatedRNN	95.67	27.1%
9	PatchTST	102.45	27.4%
10	N-HiTS	122.93	29.3%

Source: Author's elaboration.

The discrepancy between rankings demonstrates that model effectiveness varies across validation strategies, suggesting the need for comprehensive hyperparameter tuning and architecture selection based on deployment conditions (Belkin et al., 2019; Nakkiran et al., 2021). DilatedRNN achieved very good results on the final test window (a period of relative stability before the pandemic), but its effectiveness drops significantly under conditions of variability. TCN exhibits the opposite characteristic—moderate results on a stable window, but the best average in the cross-validation procedure, suggesting high resilience of the convolutional architecture to structural changes in data.

Practical Implications

The developed forecasting models provide information that can be transformed into concrete management actions in fuel distribution enterprises. Based on the conducted experiments, four areas of direct impact were identified:

- **Fuel order planning.** Forecast accuracy ($R^2 = 0.814$ for DilatedRNN, $R^2 = 0.714$ for TCN in cross-validation) enables precise fuel order planning with 30-day advance notice, eliminating ad-hoc decisions.
- **Inventory management.** MAE error reduction by 242 liters daily (vs SeasonalNaive) allows for reducing safety stock, releasing capital tied up in inventory (Babai et al., 2022). At an average fuel price of 6 PLN/L and a capital cost of 5% annually, reducing safety stock by 500 liters generates savings of approximately 150 PLN annually per station. At the scale of a 50-station network, this means releasing capital of approximately 150,000 PLN.
- **Logistics cost reduction.** Better forecasting reduces the frequency of stockouts requiring costly emergency deliveries. A single emergency delivery generates costs of approximately 750 PLN (tank truck travel cost, driver bonus, downtime). Assuming a 50% reduction in emergency deliveries (from 12 to 6 annually per station), savings amount to 4,500 PLN annually per station. The TCN model, due to its highest stability, minimizes the risk of unexpected forecast errors.
- **Strategic decision support.** The DilatedRNN model enables identification of long-term sales trends that can form the basis for decisions regarding the expansion or reduction of distribution networks in individual locations. Trend analysis from MSTL decomposition revealed the presence of sales volume growth in the first three years of observation, which can support investment decisions.

From the perspective of transaction cost theory, the developed DilatedRNN model, by reducing forecast error by 61% compared to the reference method, lowers risk associated with supply chain transactions. More accurate forecasting enables more precise delivery contracting, reducing renegotiation and emergency delivery costs. The TCN model, exhibiting the highest stability in cross-validation ($\sigma = 44.65$ L), additionally reduces business partners' uncertainty regarding forecast reliability, building trust in network relationships.

Improved forecast quality strengthens the fuel station's ability to serve as a flow stabilizer between supply chain segments, limiting decision variability and lowering coordination costs in network relationships. The choice of model architecture has organizational implications: a high-accuracy model (DilatedRNN) favors centralization at the distributor level, while a highly stable model (TCN) enables decentralization to individual stations. The estimated economic benefits of

implementing the developed method amount to approximately 656,000 PLN annually for a 50-station network, comprising safety stock reduction, emergency delivery reduction, and margin optimization.

In the traditional model (staffed station with multi-category store), accurate forecasting enables optimization of personnel work schedules and resource planning in the retail-service section (e.g., cashier service, workstation preparation) during peak and trough demand periods. Reducing emergency delivery frequency translates into fewer duty hours requiring the presence of an employee responsible for receiving goods. In the autonomous model (unstaffed automated fuel dispensers), forecasting becomes significant for planning service rounds and tank replenishment. The absence of permanent staff means inventory management must be fully automated, and forecasting systems constitute a critical element of IT infrastructure. The TCN model, due to forecast stability, is predisposed for use in automatic order generation systems without human intervention.

The developed forecasting models align with the sustainable development paradigm of the fuel industry (Sołoducho-Pelc & Sulich, 2020). More accurate sales forecasting leads to emission reductions associated with fuel transport through: delivery route optimization (fewer trips due to better planning), emergency delivery reduction (elimination of inefficient single-tank truck deliveries), and reduced fuel consumption by the transport fleet (optimal vehicle loading). An estimated 50% reduction in emergency deliveries translates into CO2 emission reduction of approximately 0.5 tons annually per station, assuming an average delivery distance of 100 km and emissions of 0.8 kg CO2/km. For a 50-station network, this means a carbon footprint reduction of 25 tons of CO2 annually, which can be an element of the enterprise's ESG reporting.

Conclusions

Achievement of the main objective: "Development and evaluation of motor fuel sales forecasting models in enterprises using deep learning," was realized through multi-stage research procedure. Necessary stages for achieving this objective were the considerations conducted in Chapters 1, 2, and 3, while the objective was fully achieved in Chapter 4.

Cognitive objectives were achieved in Chapters 1, 2, and 3 through literature and industry data analysis, identifying: (1) the specifics of fuel industry enterprise operations from a business

ecosystem perspective, (2) the state of knowledge regarding motor fuel sales forecasting, (3) deep learning architectures capable of modeling multi-seasonal patterns. In implementing the methodological objective, ten deep learning models forecasting motor fuel sales volume were developed and evaluated, creating a ranking of these models and formulating precise specifications for the recommended DilatedRNN architecture. The utilitarian objective was achieved through verification of model effectiveness using real data and formulation of practical implementation recommendations that directly translate into improved management processes and financial savings generation.

Research Gap Fulfillment

The taxonomic gap was filled by systematizing concepts related to fuel sales micro-networks and the specifics of fuel industry enterprise operations. In the explanatory dimension, the work explains mechanisms affecting prediction effectiveness, pointing to the "simplicity paradox" and the role of the dilation mechanism in modeling weekly seasonality. The empirical gap was bridged by conducting a systematic comparative study on a unique, real dataset, confronting ten deep learning architectures with three traditional statistical methods. In the methodological aspect, the work's contribution is the development of original methods for data preprocessing, building, and evaluating motor fuel sales forecasting models. The conducted study has the following limitations:

1. **30-day forecast horizon** results in low statistical power of tests. The DilatedRNN model, despite significance before correction ($p_Wilcoxon = 0.018$), maintains significance after Holm correction ($p = 0.0021$). Effect size (Cohen's $d = -0.64$) confirms practical significance of differences.

2. **Single data source.** Data come from one enterprise from the Lower Silesia region, operating several fuel stations. Generalization of results to other station types (highways, city centers) or regions requires external validation.

3. **Research period 2017–2020** includes the beginning of the COVID-19 pandemic, which may have affected fuel consumption patterns in the final part of the test set. Data may not fully reflect demand patterns under macroeconomic stability conditions.

4. **Residual autocorrelation.** The Ljung-Box test ($p = 0.0001$) indicates unexploited temporal patterns (e.g., monthly seasonality, weather effects), suggesting potential for further model improvement through architecture extension or introduction of additional variables.

Based on the conducted research and identified limitations, the following directions for future work can be indicated:

1. **Dataset expansion.** Include data from various fuel types and a longer time horizons.
2. **Residual autocorrelation elimination**. Introduction of additional explanatory variables (weather, local events, competitor promotions) or error correction mechanisms.
3. **Probabilistic forecasting.** Extension to interval forecasts (quantile regression) enabling uncertainty quantification.
4. **Ensemble methods.** Combining forecasts from multiple models (e.g., DilatedRNN, N-BEATSx, ARIMA) to reduce variance and improve forecast stability.
5. **Model explainability (XAI).** Application of SHAP methods or attention mechanism analysis for model decision interpretation, which will increase acceptability in the business environment. Interpretable architectures include N-BEATSx and N-HiTS (built-in trend/seasonality decomposition) and TFT (attention weights).
6. **Introduce LLM models.** Explore the potential of using large language models (LLM) in long-term forecasting.

The conducted empirical research confirms that deep learning models, particularly the DilatedRNN architecture, outperform traditional statistical methods by 36% in terms of explained variance (R^2) and reduce MAE error by 61% compared to the SeasonalNaive reference method. In cross-validation, the TCN model exhibited the highest stability ($\sigma = 44.6$ L). An inverse relationship between model complexity and accuracy was also identified: the optimal configuration uses a minimal set of calendar variables and a short context window.

The dissertation contributes to the management and quality sciences discipline, providing tools that improve decision-making processes, enhance operational efficiency, and support building competitive advantage under conditions of increasing environmental variability (Teece, 2007).

References

1. Adner, R. (2017). Ecosystem as structure: An actionable construct for strategy. Journal of Management, 43(1), 39–58. https://doi.org/10.1177/0149206316678451
2. Ahaggach, H., Abrouk, L., & Lebon, E. (2024). Systematic mapping study of sales forecasting: Methods, trends, and future directions. Forecasting, 6(3), 502–532. https://doi.org/10.3390/forecast6030028
3. Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007
4. Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proceedings of the National Academy of Sciences, 116(32), 15849–15854. https://doi.org/10.1073/pnas.1903070116
5. Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. Information Sciences, 191, 192–213. https://doi.org/10.1016/j.ins.2011.12.028
6. Babai, M. Z., Boylan, J. E., & Syntetos, A. A. (2022). Demand forecasting in supply chains: A review of aggregation and hierarchical approaches. International Journal of Production Research, 60(1), 324–348. https://doi.org/10.1080/00207543.2021.2005268
7. Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decision making affect firm performance? SSRN Electronic Journal.
8. Challu, C., Olivares, K. G., Oreshkin, B. N., Garza, F., Mergenthaler-Canseco, M., & Dubrawski, A. (2022). N-HiTS: Neural hierarchical interpolation for time series forecasting. arXiv preprint arXiv:2201.12886.
9. Chang, S., Zhang, Y., Han, W., Yu, M., Guo, X., Tan, W., Cui, X., Witbrock, M., Hasegawa-Johnson, M., & Huang, T. S. (2017). Dilated recurrent neural networks. In Advances in Neural Information Processing Systems (Vol. 30, pp. 77–87). Curran Associates, Inc.
10. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1724–1734). Association for Computational Linguistics.
11. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.
12. Coase, R. H. (1937). The nature of the firm. Economica, 4(16), 386–405. https://doi.org/10.1111/j.1468-0335.1937.tb00002.x
13. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

14. Dekker, H. C. (2003). Value chain analysis in interfirm relationships: A field study. Management Accounting Research, 14(1), 1–23. https://doi.org/10.1016/S1044-5005(02)00067-7
15. Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 7, 1–30.
16. Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. Journal of Business & Economic Statistics, 20(1), 134–144.
17. Dong, F., & Li, W. (2021). Research on the coupling coordination degree of "upstream-midstream-downstream" of China's wind power industry chain. Journal of Cleaner Production, 283, 124633. https://doi.org/10.1016/j.jclepro.2020.124633
18. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070
19. Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 32(200), 675–701.
20. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
21. Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008
22. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice. OTexts.
23. Ketokivi, M., & Mahoney, J. T. (2020). Transaction cost economics as a theory of supply chain efficiency. Production and Operations Management, 29(4), 1011–1031. https://doi.org/10.1111/poms.13148
24. Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012
25. Moore, J. F. (1993). Predators and prey: A new ecology of competition. Harvard Business Review, 71(3), 75–86.
26. Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., & Sutskever, I. (2021). Deep double descent: Where bigger models and more data can hurt. Journal of Statistical Mechanics: Theory and Experiment, 2021(12), 124003. https://doi.org/10.1088/1742-5468/ac3a74
27. Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with Transformers. In International Conference on Learning Representations (ICLR).
28. Nixtla. (2025). NeuralForecast documentation. https://nixtlaverse.nixtla.io/neuralforecast/
29. Olivares, K. G., Challu, C., Marcjasz, G., Weron, R., & Dubrawski, A. (2023). Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx. International Journal of Forecasting, 39(2), 884–900. https://doi.org/10.1016/j.ijforecast.2022.03.001

30. Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2020). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In International Conference on Learning Representations (ICLR).
31. Pascanu, R., Mikolov, T., & Bengio, Y. (2012). On the difficulty of training recurrent neural networks. arXiv preprint arXiv:1211.5063.
32. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958.
33. Stańczyk, S. (2019). Business ecosystem identity construct. Transformations in Business & Economics, 18(2B), 674–693.
34. Syntetos, A. A., Boylan, J. E., & Disney, S. M. (2009). Forecasting for inventory planning: A 50-year review. Journal of the Operational Research Society, 60(S1), S149–S160. https://doi.org/10.1057/jors.2008.173
35. Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. International Journal of Forecasting, 16(4), 437–450. https://doi.org/10.1016/S0169-2070(00)00065-0
36. van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. In 9th ISCA Speech Synthesis Workshop (p. 125).
37. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (NeurIPS) (Vol. 30, pp. 5998–6008). Curran Associates, Inc.
38. Williamson, O. (1985). The economic institutions of capitalism. The Free Press.
39. Zema, T., Sulich, A., & Hernes, M. (2024). Identification of research areas in fuel sales forecasting within the business ecosystem context: A review, theoretical synthesis, and extension. Journal of Economics and Management, 46(1), 79–110. https://doi.org/10.22367/jem.2024.46.04
40. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001
41. North, D. C. (1991). Institutions. Journal of Economic Perspectives, 5(1), 97–112. https://doi.org/10.1257/jep.5.1.97
42. Paul, J., & Criado, A. R. (2020). The art of writing literature review: What do we know and what do we need to know? International Business Review, 29(4), 101717. https://doi.org/10.1016/j.ibusrev.2020.101717
43. Sołoducho-Pelc, L., & Sulich, A. (2020). Between sustainable and temporary competitive advantages in the unstable business environment. Sustainability, 12(21), 8832. https://doi.org/10.3390/su12218832

44. Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal, 28(13), 1319–1350. https://doi.org/10.1002/smj.640
45. Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. https://doi.org/10.1145/2601248.2601268

List of tables

Table 1. Fuel sales forecasting model ranking. 8
Table 2. Ablation study of exogenous variable impact on DilatedRNN model. 9
Table 3. Statistical test results: DL models vs SeasonalNaive. 10
Table 4. Cross-validation results for deep learning models. 11
Table 5. Model ranking by stability (MAE standard deviation). 12